Application of Deep Learning Algorithm for Breast Cancer Classification Using Fully Convolutional Neural Networks

Shwetha.K[1], Varshaa.V[2], Shivani.R[3], G.Brenie Sekar[4]
[1,2,3] UG Students / Department of ECE, R.M.D Engineering College, Tamilnadu 601206
[4]Assistant Professor / Department of ECE, R.M.D Engineering College, Tamilnadu 601206

***Abstract*- Nowadays, the standard of classification systems depends on the presentation of the dataset, a method that takes time to use in-depth data to provide specific characteristics. Meanwhile, deep learning will extract options from a dataset, while not having to style feature extractors. Convolutional Neural Network (CNN) has been come upon as Associate in nursing intense category of models for image acknowledgment problems. CNN may be a deep learning model that extracts the feature of a images and use these feature to classify a image. Different classification algorithmic rule must extract the feature of a image exploitation feature extraction algorithmic rule like grey Level Cooccurrence Matrix. We tend to show that a convolutional neural network is an efficient tool in classifying carcinoma histopathological images and appraise its performance as a binary classifier within the field of carcinoma designation exploitation whole slide imaging. Performance of CNN is far higher when put next to different rumored results on Break His dataset exploitation different classification algorithmic rule for classifying a images.**

***Index terms*- Breast Cancer, Convolutional Neural Network, Image Classification, Deep Learning.**

I.INTRODUCTION

Cancer is one of the top leading diseases cause of death worldwide in the world today. According to WHO (World Health Organization), there were 8.2 million deaths from cancer in 2012 and it is expected that there will be 27 million cases before 2030 [1]. In particular, breast cancer death rate is very high when compared to other types of cancer [2]. Detection and diagnosis of BC can be achieved by imaging procedures such as diagnostic mammograms (x-rays), ultrasound (solography), magnetic resonance imaging, etc. Diagnosis from a histopathology image is standard in diagnosing almost all types of cancer including BC [3].

Most authors have focused to the field of automated two classes classification of BC histopathology image analysis: benign or malignant, with computer-supported diagnosis. Kowal et al. [4] perform various algorithms to segment nuclei over a set of 500 images, with accuracy greater than 96%. Filipczuk et al. [5] launch the BC diagnostic system using four different classifications trained and achieved a 98% performance efficiency over 737 images. Other authors use one-class classifiers on the same database with a recognition rate of 92%. In general, they exploit various methods to classify two main classes and release own results. But we see that the above classification works are not enough effective to determine detailed breast cancer for giving treatment solution. Therefore, we propose to classify and recognize specific eight breast cancer subclasses in two benign and malignant classes with the same BreakHis dataset.

An example of deep neural network is a multilayer perceptron in which each layer is fully connected to the other layer. In our paper, we have used deep Convolutional Neural Network to classify a breast cancer Image. In Convolutional Neural Network, few of the nodes in a below layer are connected to the node in the next layer and the last layer is a fully connected network where the actual classification task is performed. Deep learning is widely used in diagnosis of the medical image. Training a deep CNN takes a lot of time and it needs very high computing resources like GPU processor to train a CNN network. Deep architecture [4] of CNN is the basic building block which is used to extract the features of an image. CNN requires a large amount of labeled

data to train the neural network which is difficult to get in case of the medical image. Cancer is basically an abnormal growth of cells. Breast Cancer is increasing rapidly nowadays. Breast cancer is one the main reason for growing death of women in the world. It is the most commonly occurring cancer among the women in the world. Nearly 1.7 million cases were diagnosed in 2012, which represents the 12 % of new cancer cases and 25 % of cancer case among the woman. It ranks a fifth most common case for the death of the woman. Predicting a breast tumor is a challenging task for many Doctors and physicians. With the advancement of new technology and a large amount of patient data available has given a motivation for the development of new techniques to predict and detect the breast tumor.
In spite of significant advancement in diagnostic image technology, diagnostic of breast cancer image, including grading and staging, continues being done by pathologist applying visual inspection of histological samples under the microscope. With the recent advancement in image processing and machine learning technique allows building computer aided Detection (CAD) systems that can help pathologist to be more productive and accurate in diagnosis.

II. PROBLEM STATEMENT

Some previously completed research was used to understand the types of breast cancer, types of medical image used for classification, types of neural network for image classification. Tensor flow tutorial on Convolutional Neural Network is official method that uses CNN. There is an example on CIFAR-10 classification with various over fitting, good training, and performance boosting techniques [5]. Image Net Classification with Deep Convolutional Networks is designed to category machine learning and image classification [6]. Many of techniques that use on model have references from [6]. They are the preprocessing techniques idea.\
Consequently, some works focus on Classification of breast cancer histology. The latest version of histology image classification by using CNN, but histology image has its limitation which it is takes long time for the lab usage [7]. Reference [8] is the most cited paper that use MRI image for classification by using ANN, but MRI has a lot of radiation exposures that may harm to human health. However, mammogram image is the most suitable target for this work, since it is cheap than other treatment, and has very tiny radiation exposure from mammogram [8]. Therefore, many patients can have early checking if a faster mammograms image has been diagnosed. But mammogram also has its limitations for example some not clearly cancer cannot be scanned.
Some previously completed paper had used different types of neural network for classification mammogram image in breast cancer. Reference [9] is for the prediction of breast cancer by using artificial neural networks for classification and prediction [9]. The wavelet neural network is employed for breast cancer diagnosis [10]. Both neural networks are designed for general decision making purpose, so both of them need to setup many parameters than CNN which is designed for purpose recognize visual imagery. Same number of hidden layer, standard neural networks working on processing visual imagery need have 3x106 parameters. But for CNN, it only needs around 600 parameters for processing visual imagery.
Reference [11] suggested a mammogram image detection using CNN. But the accuracy percentage is too low for a medical side solution which is around 60% for all classes detection, 75% for only masses class, and 100% for only calcification. Except only calcification argument, all argument and mass only argument can further be improved the accuracy to get a better performance [11].

III. METHODOLOGY

Training was done on 70% of the dataset from Break His from a total of 322 images. Fig. 1 shows the methodology followed for the proposed system.

Fig. 1. Conceptual level block diagram of the training and testing procedures in the proposed system

In this paper we trained the CNN from scratch for our application. Network layers in a CNN works as a detection filter for the presence of specific patterns or features present in an image. First layers of a CNN detect large features which can be interpreted easily. Next layers detect smaller features that are more abstract. Last layer is able to make very detailed classification by merging all the features detected by the previous layers. CNN contains 7 layers with weights as shown in Fig. 3, the first four layers are convolutional l layers and the remaining 3 are fully connected layers. The inputs of the DCNN are gray scale images. Each neuron computes a dot product of weights to the local region which is connected to the input volume. We have used 4, 16 and 80 number of filters of size (2, 3, 5), padding of size (3, 2, 1) along all edges of the input layer. Filter size [3 3] specifies filters of height 3 and width 3. Each filter is slided across the width and height of the input. Two pooling layers are used which performs down sampling to minimize the computation and enhance the robustness. Pooling layers with filter size of 2 by 2 pixels which outputs the maximum value of 4 inputs in each local region.

IV. SYSTEM IMPLEMENTATION

A. Dataset

The BreakHis database[5] consists of minuscule biopsy pictures of malignant and benign breast tumors. Pictures were gathered through a clinical report from January 2014 to December 2104[1].Patients with a clinical sign of breast cancer, alluded to the P & D lab. During the period January 2014 to December 2014, patients suffering from breast tumors at P & D lab Brazil were invited to participate in the study of breast tumor. Board individual from the establishment affirmed the investigation and all patients gave a composed assent. BreakHis dataset consists of 9,109 pictures of breast tumor tissue. It is gathered from 82 patients available on site http://web.inf.ufpr.br/vri/breast-cancer-database with different magnification factor (40 X, 100 X, 200 X, and 400 X). It consists of 5,429 malignant and 2,480 benign examples (700 X 460 pixels). This dataset has been made in Brazil with a collaboration of P&D lab - Pathological Anatomy and Cytopathology, Parana, Brazil. It is publically available for researchers to explore the breast tumor tissue.

BreakHis Dataset is of two categories: benign tumor and malignant tumor. In general, the cause of benign tumor is not known. It develops as the cells in the body grow and divide. A benign tumor is basically a noncancerous tumor but it is found in most of the women. It does not metastasize to other parts of the body. Generally, malignant tumor refers to cancer. It metastasizes to other parts of the body and may lead to the death of the person. Samples were collected by different biopsy method.

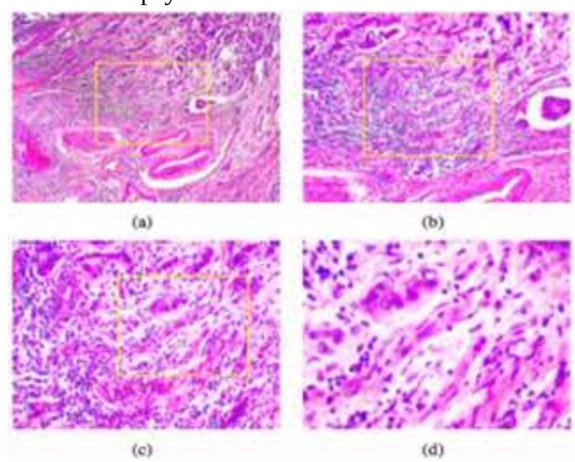

Fig 2: Breast malignant tumor slides observed with different magnification ratio 40 X , 100 X , 200 X and 400 X.

B. Preprocessing

Image enhancement is processing the mammogram images to increase contrast and suppress noise in order to aid radiologists in detecting the abnormalities. There are many image enhancement techniques as in among which is the adaptive contrast enhancement (AHE). The AHE is capable of improving local contrast and bringing out more details in the image. It is an excellent contrast enhancement method for both natural and medical images However, it can also produce significant noise. In this manuscript, contrast-limited adaptive histogram equalization (CLAHE) which is a type of AHE will be used to improve the contrast in images. One of the disadvantages of AHE is that it may over enhance the noise in the images due to the integration operation. Therefore, the CLAHE is employed as it uses a clip level to limit the local histogram in order to restrict the amount of contrast enhancement for each pixel.

The CLAHE algorithm can be summarized as follows:

1. Divide the original image into contextual regions of equal size,
2. Apply the histogram equalization on each region,
3. Limit this histogram by the clip level,
4. Redistribute the clipped amount among the histogram, and
5. Obtain the enhanced pixel value by the histogram integration.

C. Segmentation

Image segmentation is used to divide an image into parts having similar features and properties. The main aim of segmentation is to simplify the image by presenting in an easily analyzable way. Thresholding methods are the simplest methods for image segmentation. The image pixels are divided with respect to their intensity level. The most common type of thresholding method is the global threshold. This is done by setting an appropriate threshold value (T). This value of (T) will be constant for the whole image. On the basis of (T) the output image p(x,y) can be obtained from the original image q(x,y) as given in Eq. (1),

$$p(x,y) = \begin{cases} 1, if q(x,y) > T \\ 0, if q(x,y) < T \end{cases} \quad (1)$$

V. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are effective for image classification problems because the convolution operation produces information on spatially correlated features of the image. Convolution is performed by initializing a square matrix with specific values. This matrix, or kernel, is then applied to each pixel in an image. For each pixel in an image, the kernel multiplies the pixel and its adjacent pixels that the kernel covers by their corresponding kernel values. The products are then summed and this value is set as the pixel value in the convolved image at the initial pixel's location.

As a result of convolving, the image is filtered for specific features and those patterns are enhanced to produce a new overall effect. For example, convolving may result in edges becoming more prominent or the entire image becoming more blurred. This can be valuable in extracting specific features unique to certain images that indicate a particular class. After convolution, an image's specific identifying features may be more readily learned by a fully connected neural network than they would be without the convolutional step. Our CNN takes an image and convolves various types of kernels over the image, creating different output nodes that later get fed into more convolutional or fully connected layers. More informative kernels that help with classification become the more active nodes.

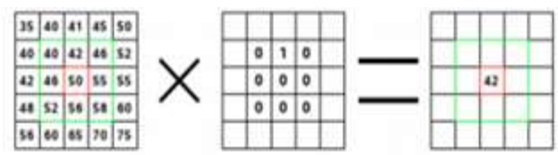

Figure 3: Convolution of a matrix with a 3 x 3 kernel [15]

Though one convolutional layer can only detect elementary features, due to the nature of convolution, feeding the output of one convolutional layer to another allows for high order feature extraction. For example, an initial kernel may be optimized to extract edges within the initial convolutional layer, and a second kernel may soften more organic shapes in the next layer, and so on. Our CNN also uses a technique called max pooling. Max pooling takes the output of the convolution and splits it up into tiles. We chose our tiles to be 2 × 2 pixels each. Only the largest value from each tile is used in the next layer of the network. In the past many researchers using CNNs used average (or mean) pooling. This can be seen in the work of Lo et al. The reasoning behind average pooling makes sense in that taking an average of the pixels will assure that no information is completely lost in the pooling step. However, increasingly often max pooling is used to extract the most prominent groups of features from each convolution. This means that the output into later layers is filtered for the most informative patterns relating to the problem domain. Another consequence of pooling is that the input is reduced in size, which reduces computation time by reducing the number of inputs to the fully connected neural network.

As with most CNNs, ours uses back propagation in or der to update weights to be closer to their optimal values. This means that every time input is passed through the weighted layers of the network, an error value is calculated for the expected output. That error is propagated back through the network to update the weights that contribute most to the error. After multiple iterations or steps, the weights learn by being updated to make more and more accurate predictions based on the training data expectations.

There are several parameters of CNNs that can help optimize training time, such as learning rate and stride. Learning rate determines the rate at which neuron weights are updated during back propagation. A low learning rate will likely yield a higher accuracy (though it could get stuck in local optima) as it will be able to optimize weights to a higher significant figure, but the convergence time will be non-trivial. Conversely, a learning rate that is too high may not get close enough to the global optima and diverge. Stride dictates how many pixels of the input image the kernel slides over and skips between individual convolutions. This effectively reduces both the number of convolutions in the current layer and the dimensions of the image outputted by the convolution, meaning the reduction in convolutions and processing time in future layers will be of an even larger factor.

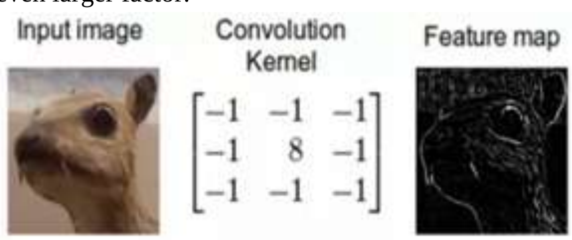

Figure 4: Example of edge extraction effect on an image due to convolution [14]

VI. SYSTEM IMPLEMENTATION

In our paper, we have used CNN architecture with some variation in parameter that has given the best result which contains the following layers.

1. Input layer: In this layer, an image is given as an input and produces output which is used to feed the Convolutional layers. In our example, input is of (32 X 32 or 64 X 64 pixels) is taken into consideration and the number of channel of image is 3 for RGB.
2. Convolutional layers: In this layer, input image is convolved with a set of learnable filters, which produces a feature map corresponding to each image in the output image. In our model, there are six Convolutional layers. For first three layers size of the kernels is of 5 X 5 and padding is set to zero and the stride is set to two and for next three layers, we have kept the size of kernel to be 3 X 3.
3. ReLU layers: This layer is also known as Rectified Linear Unit. This is an activation function which activates the neurons above a certain threshold value. Let the given input value is y, the ReLU layers computes the neurons output as y if f(y) > 0 and 0 if f(y) < = 0.

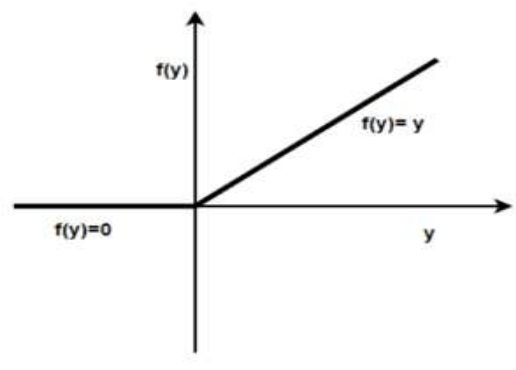

Fig 5: Shows the ReLU activation function

4. Pooling Layers: Pooling layer is basically used to reduce the size of an image keeping the high level features of an image. This layer is responsible for down sampling the spatial dimension of the input image. We can have one pooling layer after each Convolutional layer. Each of the pooling layer is set to use 3 X 3 receptive field spatial extent) with a stride of 2.In the first three pooling layer, max pooling function is applied to the image to get the maximum pixel value in a window. For other three pooling layer, we have used average pooling.
5. Fully Connected Layer or inner product layer: In this layer neurons are fully connected to each other and producing the result. Here input is simply treated as a vector and produce an output in the form of a single vector. In our model, we have used two inner product layers. The last layer is a fully connected layer where a softmax layer is used to classify the input image. Total number of classes used in our case is two, one for benign and other for malignant.

A. Training of the Convolutional Neural Network

Convolutional Neural Network deals with high resolution images which are basically used for breast cancer Histopathological image classification. Using a deep neural network [10] with the larger set of image data will lead to over fitting and a lot of parameter is to be updated in the hidden layers of neurons, which can increase the complexity of the model. As a result, time taken to train the model and update the parameter can be very large. CNN is basically used for the extraction of filter or small patches of image which is basically used to train the network and the combination of these filters forms a feature map which helps to recognize image. To learn the features and parameters of CNN, only small patches of images are used for training. The main idea is to extract the features of a high resolution image.

Stochastic Gradient Descent (SGD) algorithm is used to optimize the network by minimizing the objective or loss function .Back propagation algorithm is used to compute the gradient with respect to the parameter of the network. For our training purpose, we have used supervised technique which is frequently in practical for image recognition. As usual in supervised model, the Stochastic Gradient decent algorithm with back propagation technique to compute gradient and mini batch size of 1 was used to update the neural network parameter with a learning rate of 0.0001.

We have used 10 folded for cross validation as one of the samples is used for validating the data and remaining 9 samples is used for training the CNN. In fully connected CNN, 1024 filter with a dropout of 0.6 is used to train the network. Finally, in last layer we have used two filters with softmax layer to classify the image into two classes: benign and malignant. In our experiment, we have used Categorical Cross Entropy as the objective function. We have trained the CNN using a tensor flow (which is a deep learning framework) on CPU processor.

B. Classification

For the image recognition, mask or patch are put for the whole image. Since the models are trained on small patches of image, we need a method to divide the original test image into patches of image. As we extract more number of patches or filters of the image, better result we would get but getting more number of features or patches tales lot of time and the computation time increases.

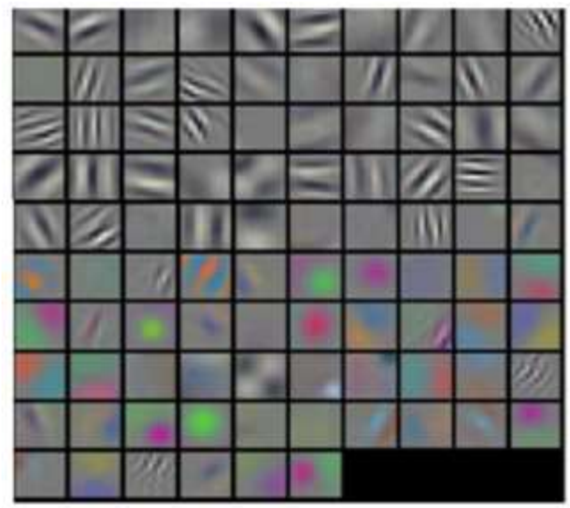

Fig 6: Feature map learned by the first Convolutional layer

The proposed Convolutional Neural Network algorithm extracts the features of BreakHis dataset image and classifies the image between benign and malignant tumor. In our model, we have used CNN network for both the feature extraction as well as the classification task. In final layer (Softmax layer) which is used to classify the image into two classes: benign and malignant.

VII. RESULTS AND DISCUSSION

BreakHis dataset is divided into two group training (70%) and testing (30%) set as in accordance with the experimental rules. Dataset is split in such a manner that the training set is not used for testing set so that the algorithm would be able to classify the unseen patient image correctly.

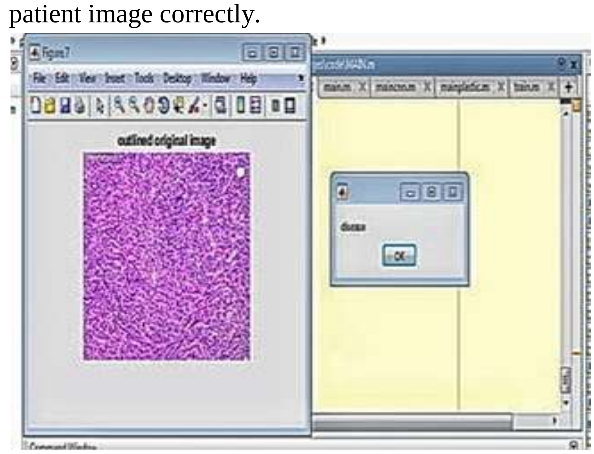

Fig 7:Experimental Results

Figure 8 presents accuracy chart in training and testing process. Figure 8 presents loss chart of process respectively.. This process takes about a few hours. As we can see that the optimal number of epochs is 32 for the highest validation accuracy (73.68%) corresponding to the lowest validation loss.

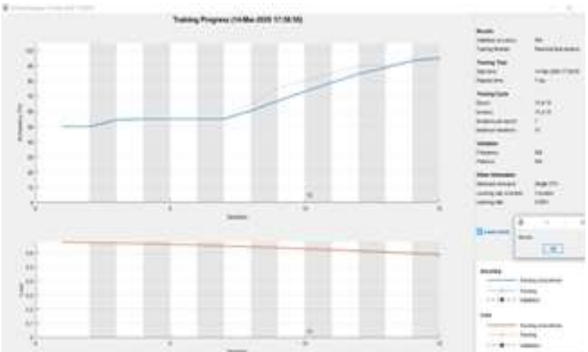

Fig 8.Accuracy obtained by the CNNs with different convolutional filter sizes on the raw images

VIII. COCLUSION

This paper describes a new approach for histopathology image related breast cancer diagnosis. The task is to automatically distinguish healthy and diseased cases. The approach is based on the analysis of cytological (histopathology) images. Due to the fact that most of the current segmentation methods do not work properly on the new very high resolution images used in this study, we decided to dispense with accurate segmentation in favor of estimating cell nuclei by circles. For this purpose, the circular Hough transform was used together with subsequent removal of incorrect or less reliable detections using a CNN-based procedure. The presented solution allows for removal of unreadable areas from consideration and allows for the determination of features based on only certain high-quality isolated nuclei. This, together with the proposed features and classifiers gave very good results. The best obtained effectiveness reached 98.51% indicating that the presented method is effective and capable of providing valuable diagnostic information.

REFERENCES

[1] J. Ferlay, I. Soerjomataram I, M. Ervik, R. Dikshit, A. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman and F. Bray. GLOBOCAN 2012 v1.1, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 .

[2] “U.S. Breast Cancer Statistics” Breastcancer.org 120 East Lancaster Avenue, Suite 201 Ardmore, PA 19003 web. Available: http://www. breastcancer.org/. Last accessed : 29 Sept 2017.

[3] Breast Cancer Care. “Three quarters of NHS Trusts and Health Boards say "not enough" care for incurable breast cancer patients” breast cancer care pink ribbon 25. Sophie Softley Pierce, PR Manager, Breast Cancer Care. Available: https://www.breastcancercare. org. uk/aboutus/media/pressreleases/three-quarters-nhs-trusts-health-boards-say-not enough-careincurable Last Accessed: 02 October 2017.

[4] Breast Cancer. “Mammography: Benefits, Risks, What You Need to Know” Breastcancer.org Susan Greenstein Orel, M.D. Available: http://www. breastcancer.org/symptoms/testing/ types/mammograms/bene fits_risks Last Accessed: 02 Octorber 2017.

[5] Tensor Flow. Available: https://www.tensorflow .org/ Last Accessed: 2 Octorber 2017.

[6] A. Krizhevsky, L. Sutskever, & G. E. Hinton, “Imagenet classification with deep convolutional neural networks.” In Advances in neural information processing systems , pp.1097-1105, 2012.

[7] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, & A. Campilho. “Classification of breast cancer histology images using Convolutional Neural Networks,” PloS one, vol.12(6), e0177544, 2017.

[8] J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, and P.S. Meltzer, “ Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks,” Nature medicine,vol. 7(6), pp. 673, 2001.

[9] I. Saratas, "Prediction of breast cancer using artificial neural networks," Journal of Medical Systems, vol. 36(5),pp. 2901-2907, 2012.

[10] V. Dheeba, N.A. Singh, and J. A. P. Singh, "Breast Cancer Diagnosis: An Intelligent Detection System Using Wavelet Neural Network," Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013. Springer, Cham, 2014.

[11] H. Zhou, Y. Zaninovich and C. Gregory, “Mammogram Classification Using Convolutional Neural Networks”, Varialable: http://ehnree.github.io/documents/papers/mammogram_conv_net.pdf Last Accessed: 02 October 2017.
[12] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, and P. Taylor, “The mammographic image analysis society digital mammogram database,” In Exerpta Medica. International Congress Series, vol. 1069, pp. 375-378, July 1994.
[13] H. H. Wang, and B. Raj, “A Survey: Time Travel in Deep Learning Space: An Introduction to Deep Learning Models and How Deep Learning Models Evolved from the Initial Ideas,” CoRR, arXiv: 1510.04781, 2015.
[14] A. Karpathy. CS231n: Convolutional Neural Networks for Visual Recognition. Stanford University. Avaialable: http://cs231n.stanford.edu/ Last Accessed: 2 October 2017.
[15] Sentdex (2016) Full Classification Example with ConvNetRetrievedfrom:https://www.kaggle.com/sentdex/full-classificationexample-with-convnet. Last Accessed: 2 October2017.