Breast Cancer Classification: CNN based

Shinde Yash Vikas
Computer Science and Engineering Department,
Lovely Professional University
Phagwara, Punjab, India.
yashshinde227722@gmail.com

***Abstract*---** **Breast cancer affects the females in the ratio of 1:8 across the globe. It is discovered by detecting the malicious cells of breast tissue. Modern medical image processing techniques work on histopathology images captured by a microscope, and then analyze them by using different algorithms and techniques [1]. At present, ML algorithms are now being used to process medical imagery and pathological tools. Manual detection of a cancer cell is a tedious job and involves human error, and therefore computer-based mechanisms are put in to execute for better score considered with manual pathological detection systems. Usually, this is finished with the help of extraction features through a convolutional neural network (CNN) and further categorized using a fully connected network. Deep learning is broadly used in the medical imaging field, as it does not require prior expertise in a related field. In this report, we have trained a convolutional neural network and determined a prediction accuracy up to 98.3%.**

***Keywords*---** **Classification, Histopathology image, Medical image processing, Convolutional neural network, Deep learning, Breast cancer.**

I. INTRODUCTION

Among all sorts of cancer in women, breast cancer is being experienced mostly. It has the second highest death ratio after Lung & Bronchial cancer, and nearly 1/3 of newly occurred cases. In order to fight against cancer needs early detection which can only be achieved with an efficient detection system. Hence, Techniques have been developed to detect breast cancer, including medical image processing and digital pathology defined in [1]. These images are taken by histopathology, which generally includes biopsy of the infected tissue.

Tissues affected by the tumor are retrieved by the pathologist and stained by H& E, which is a union of histological stains called hematoxylin and eosin, after which it is observed under a microscope to detect cancerous cells by finding malignant features in cellular structures like nuclei. These microscopic images got collected then used for developing computer-based detection systems. Manual detection is a tedious, tiring task and most likely to comprise human error, as most parts of the cell are frequently part of irregular random and arbitrary visual angles. The motive is to identify whether a tumor is benign or of a malignant in nature, as malignant tumors are cancerous and should be treated as early as possible to reduce and prevent further complications. In short, it is a binary classification problem and can be resolved by various ML methods as it has been proven in the past that ML algo performance is better than the human of cancer pathologist. Most of scholars have also found more accurate solutions using image preprocessing with respect to the objective diagnosis given by a pathologist. A study in Europe has been conducted by Phillips named analyst in which a list of algorithms along with breast images provided more accurate classification. This inspects that if we use high-resolution images with better algorithms will increase the performance and accuracy of cancer classification.

II. LITERATURE SURVEY

There are various ways to detect breast cancer including, Magnetic Resonance Imaging (MRI) Scans, Computed Tomography (CT) Scans, Ultrasound, and Nuclear Imaging. Although, none can give a completely correct prediction of cancer. Tissue-based diagnosis is mainly done with a

staining methodology. In this procedure elements of tissues are colored by some staining element, usually hematoxylin and eosin. Cell structure, types are painted and then put under the microscope or captured images by camera. Now, to detect the tumors we need to do a histopathology test. It is an ancient concept to predict the cancer cells and has several disadvantages or we can say faults like involves intra-observer variations, cancer cells can have multiple appearances, many figures can have the same hyperchromatic features which eventually makes the identification difficult and inaccurate. The choice of area is also a factor as the process is going to happen on a small area, so the area we choose should be lies in the tumor boundary.

The above-mentioned problems can be achieved by using deep learning strategies. It is a popular subset of machine learning technology which is inspired by the functioning of the human brain to analyze unstructured patterns. Deep learning models have a high success rate because they train on hierarchical representations. Additionally, they can extract and organize different features and hence do not require any prior domain knowledge. On the contrary, unimportant methods need rigorous feature engineering to fetch features, which involves domain expertise. Many DL methods have been implemented to predict the class of tumor. Most of them are binary classification but some of them have used multivariable classification from [1]. Proper formatted data with some network attributes suitable to the problem only required by DL. One can also use already designed networks such as AlexNet, Inception, and many more.

There are many methods and manual networks derived by scholars to classify breast cancer other than the predefined networks given above. For example, Artificial Neural Networks depend upon MLE (Maximum Likelihood Estimation) RBF Neural Networks, the GRU-SVM model which is the ML algorithm combined with a type of recurrent neural network (RNN) and gated recurrent unit (GRU) with the support vector machine (SVM). Along with these techniques, other scholars have developed methodology to obtain better results with less computational complexity. For reducing the input feature size, Karabatak et al. proposed the AR þ NN method, in which the number of features are reduced by applying association rules [2]. A combination of NN and multivariate adaptive regression splines (MARS) are also utilized to detect cancer. There is another system that consists of the fuzzy-artificial immune system and the KNN algorithm. Descriptors such as CLBP, GLCM, LBP, LPQ, ORB, PFTAS with classification of breast cancer up to a maximum accuracy of 85.1%. As for the Break-His dataset published in 2015, only some scholars have used this. For example, Fabio Spanhol at el. describes parameters and the network setup which has obtained an accuracy ranging from 80 to 85% [1]. This is further enhanced by the proposed method described herein. In deep learning algorithms, a sequence of steps is performed. First step is image preprocessing, here the required data is converted into the format which is directly put as an input to the network. This step involves multiple channeling of images, then segmentation is done (only if required, e.g. if there is a need to separate regions of interest from the background or exclude parts that are not needed for training). On this stage, data is ready for training, either in a supervised or an unsupervised manner. The next step is feature extraction. Features represent the visual content of the histopathology image. In the case of supervised feature extraction, features are already recognized and different strategies are applied to find them, but in case of unsupervised feature extraction methods, features are not known and acquired implicitly in proposed solutions through the Convolutional Neural Network (CNN). The last step is classification, which places an image into the respective class (benign or malignant) and can be done with a fully connected layer using an activation function such as softmax that we have implemented in this report.

III. DATASET USED

Dataset used in this project, has images of tissue of breast tumor captured in .png format under the microscope collected from 82 patients with the help of distinguish resolution factors (40X, 100X, 200X, and 400X). Additionally, it is separated into 2 divisions named benign and malignant in histopathology [3]. Benign is considered as a term that pertaining to a wound in simpler, which has dissimilarity compared with the features of malignancy like carcinoma. In general, tumors that grows slowly and restricted are said “innocents” treated as benign tumors. Malicious tumors are considered as a symptom for

the cancer referred as malignant.

Magnification	Benign	Malignant	Total
40x	652	1370	1995
100x	644	1437	2081
200x	623	1390	2013
400x	588	1232	1820
Total	2480	5429	7909

Table 1. Distribution of images in dataset.

IV. IMAGE CLASSIFICATION

The entire classification of images is done in 3 steps as:

1) Input includes the set of training data that contains N images and they are categorised into 2 classes i.e. malignant and benign.

2) After that, the training set guides a binary classifier in order to figure out about the prediction of the images.

3) Lastly, unlabelled data which is not seen before is tested by classifier using the model and then differentiation is made between the actual labels and the predicted labels.

V. CONVOLUTIONAL NEURAL NETWORK

CNN is a modified variety of deep neural net which depends upon the correlation of neighbouring pixels. It uses randomly defined patches for input at the start and modifies them in the training process. Once training is done, the network uses these modified patches to predict and validate the result in the testing and validation process. Convolutional neural networks have achieved success in the image classification problem, as the defined nature of CNN matches the data point distribution in the image. As a result, many image processing tasks adapt CNN [4] for automatic feature extraction. CNN is frequently used for image segmentation and medical image processing as well.

The CNN architecture has two main types of transformation. The first is convolution, in which pixels are convolved with a filter or kernel. This step provides the dot product between image patch and kernel. The width and height of filters can be set according to the network, and the depth of the filter is the same as the depth of the input. A second important transformation is subsampling, which can be of many types (max_pooling, min_pooling and average_pooling) and used as per requirement. The size of the pooling filter can be set by the user and is generally taken in odd numbers. The pooling layer is responsible to lower the dimensionality of the data and is quite useful to reduce overfitting. After using a combination of convolution and pooling layers, the output can be fed to a fully connected layer for efficient classification.

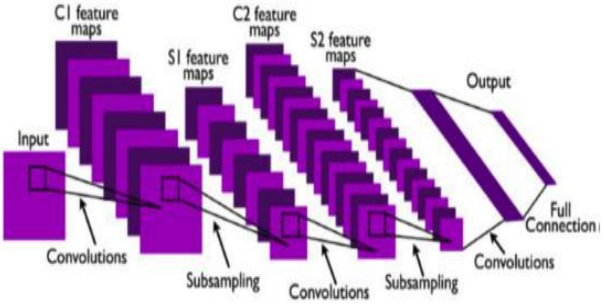

Figure 1. CNN architecture

Apart from the architecture of CNN, there is an additional key point, i.e., that simplicity to the user is helpful on the development side, as CNN requires a tremendous amount of data for training. It also requires more training time as compared to other supervised and unsupervised training approaches.

VI. PROPOSED METHODOLOGY

This paper introduces and assesses a deep learning architecture for automized breast cancer detection that incorporates concepts of ma- chine learning and image classification. We have described different Deep Neural Network architectures, especially those adapted to image data such as Convolutional Neural Networks. This used the labeled (benign/malignant) input image from the raw pixels and highlighted the visual patterns, and then utilize those patterns to distinguish between non-cancerous and cancer containing tissue, working akin to digital staining, which spotlights image segments crucial for diagnostic decisions, with the help of a classifier network. The CNN was trained using 2480 benign and 5429 malignant samples belonging to the RGB color model. Therefore, provides an effective

classification model for classifying breast tissue as being either benign or malignant.

A. Image preprocessing and Loading

Import all libraries and dependencies that are required. Next, images should be kept in their respective folder. Now, numpy arrays are coded for naming benign and malignant images as arrays of 0s and 1s respectively [5]. The dataset is also need to re-organized and developed the labels into categorical format. The dataset is divided into 2 sets, one is of training sets with 80% of images and the other is of testing sets with 20% images respectively. For example, some sample images are shown in *figure 2*.

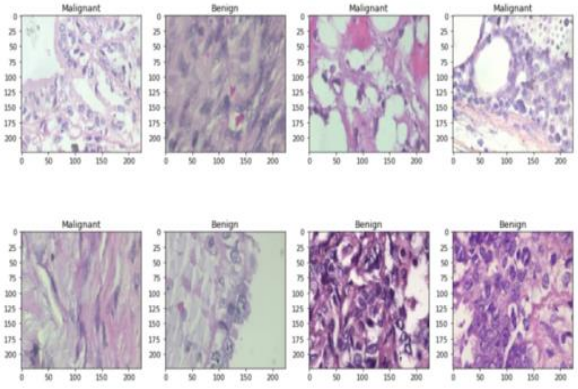

Figure2. Malignant and benign sample images

B. Feature Extraction – Labelling, Data generation

Batch size (e.g. 16 is used in our method) is among the key factors that can be used to regulate the deep learning. It is evident that if the batch size is larger then, it will be more beneficiary to maintain the model due to fact that it improves computational speed as allows to work parallel with GPUs. But, by using overlarge batch it will disturb or not provides the appropriate generalization. On the other hand, engaging a batch which is sufficient to the entire set of data surely converge to the global optimal of the targeted function. In contrast, if batch size is taken smaller then, that leads to faster convergence to produce good results. It is also demonstrated on the basis of the fact that the smaller batch allows the model to start irrespective of having ascertain all information. The worst part of using the small batch size is there is no guarantee to converge to the global optimal by the model. Hence it is suggested that we should begin at a little batch size taking the advantage of fast training dynamics and then, can slowly increase the batch size through the training.

Data Augmentation is also carried out in order to enhance the dimensions of the training dataset. It allows the network to find out more diversified data but can also considered them as representative data points during the training. After this, knowledge generation takes place in which a generator is implemented to fetch the information from the image folders and load them into Keras in an automatic way because, Keras supplies required functions for the same.

C. Model Building

Steps to attend sequentially:

1) Apply DenseNet201as each layer having the additional inputs from the previous layers in the form of knowledge and passes on its own feature maps within Image Net.

2) Apply average pooling layer in order to deduce the overfitting of the dataset with the dropout regularization of 50percent.

3) Apply batch normalization to standardize the inputs to a layer which accelerates the training process.

4) Apply Adam optimizer in order to fulfill the loss function that is used to calculate the quantity which the model needed to minimize during the training.

D. Training and Evaluation

Without callback the model is like driving a car without having breaks. Callback is the set of fuctions that to be used at given stages in the training process to automate the tasks that carried out at every epoch that helps to have control over training. The two callbacks that applied in this work:

1) Model Checkpoint: This is the callback that saves the model after every epoch at the time when the validation

accuracy is minimum between the training in order to tackle the problem of overfitting. Here, mode is been set as "min'' if the observed value is val_loss and minimization is needed to optimize.

2) Reduce LR On Plateau: When the metrics of data analysis get restricted and optimization is not consume then reduced learning rate callback is applied. Generally, the training rate deduce by the models is 2 to 10 once learning stagnates. It observes a quantity and if there no development has taken place for a sufficient count of epoch then it results in the reduction of training rate. Here, this model is trained for 20 epochs.

3) Performance matrices: Matrices which is primarily used to examine the power and the performance of the model. Accuracy is obtained from the same matrices eventually shows the % of correctness and errors. In this project, the model which is applied is predicting with the accuracy of approximately 98%. Still it could not able to detect malignant images of having cancer.

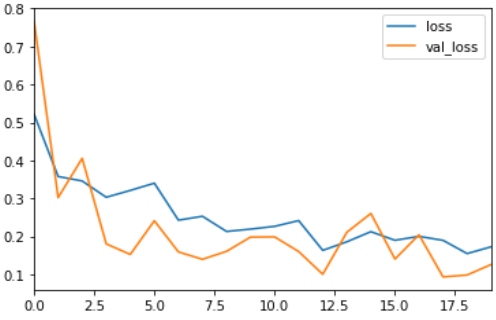

Figure 3. *Loss vs Epoch*

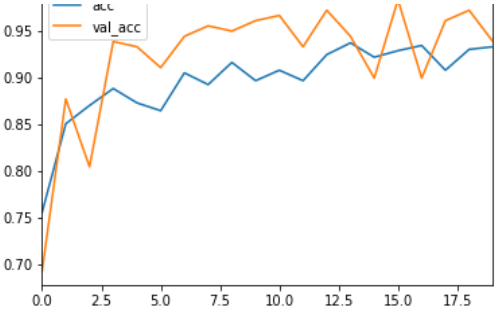

Figure 4. Accuracy vs Epoch

E. Prediction

The improved approach to determine the misleading of the model is using the subsequent matric. With the help of this matric, true positive, true negative, false positive, false negative is retrieved. The ratio of true positive divide by true positive plus false positive is precision whereas Recall is that the ratio of true positive divided by true positive plus false negative that are monitored in actual class and if want to calculate the average weight then F-1 score is considered.

$$F1 = \frac{2 * (Recall * Precision)}{(Recall + Precision)}$$

It is directly proportional to the performance of model. For the matrix, 1 is treated as an optimal whereas 0 is treated as the opposite (worst case).

1) Confusion Matrix

In the matrix, actual label is taken as column and predicted label is taken as row. Diagonal depicts the samples that are present in actual class and classified truly. Here, as shown in *figure 5*. Matrix not only guides to find the misclassified class but also why or where it went wrong. Hence, concluded that it plays a vital role in identification.

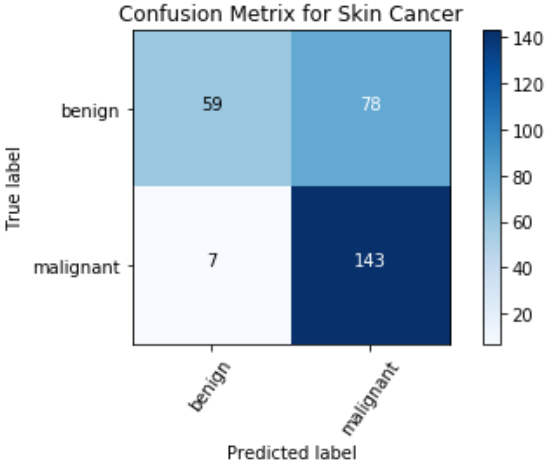

Figure 5. Confusion matrix for breast cancer

2) *ROC Curves*

The random line which inclined 45 degree to the X-axis as shown in the *figure 6* has the area under its curve i.e. 0.5. The curve to that mentioned above line has an importance. In fact, if the area under the curve is high then assumed that more effective will be the model. Right-angled triangle is made then the model is treated as the best with an AUC of 1. The ROC curve also can help debug a model. for instance, if random line near to rock bottom left corner then the model is misclassifying at Y=0.

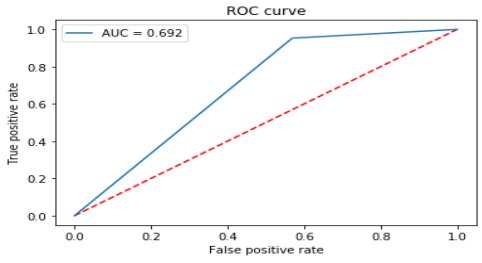

Figure 6. Area Under the Curve

VII. RESULTS

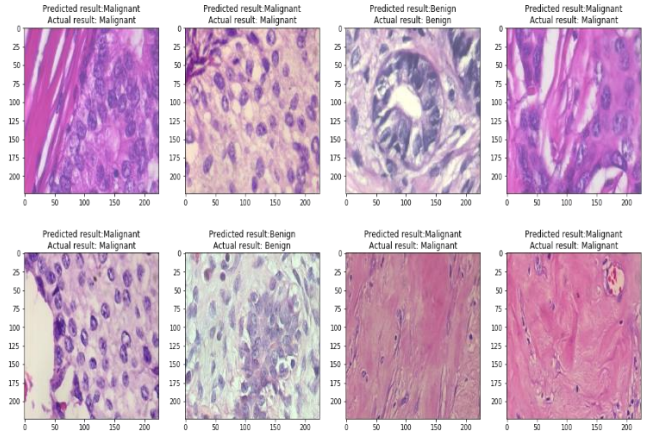

Figure 7. Correct/Incorrect classification samples

Accuracy	Precision	Recall	F1 score	ROC-AUC
98.3%	0.65	0.95	0.77	0.692

Table 2. Accuracy

Above *Table 2.* shown that the accuracy is 98.3%, precision is 0.65, recall is 0.95, f1 score is 0.77 and ROC-AUC as 0.692. By using this automated process there is a possibility of inexpensive detection of cancer in the early stages, which can ultimately increase survival rate.

VIII. CONCLUSION AND FUTURE SCOPE

In this work, we design a CNN which may not fully work in the real-life conditions, but it works great binary classification like here in this project, where breast cancer histopathology images are classified as benign or malignant. It has also opened a door to new opportunities for research as there are many undiscovered areas that can be revealed by techniques and tools of machine learning and deep learning. We may obtain improved results by altering the network design and parameters. As an improvement to the proposed method, one can implement an autoencoder instead of manually reducing image size. Additionally, we may combine various imaging technologies such as MRI, CT Scan, ultrasound, and mammographic images, and determine their collective results. This technique is known as multi-model fusion. Problems stated above can again readily be solved by deep learning and can be used to perform high quality research that might provide even better result.

IX. REFERENCES

[1] F.A. Spanhol, L.S. Oliveira, P.R. Cavalin, C. Petitjean, Deep features for breast cancer histopathological image classification 2017 IEEE international conference on systems, man, and cybernetics, SMC 2017, banff, AB, Canada, October 5-8, 2017 (2017), pp. 1868-1873.

[2] M. Karabatak, M.C. InceAn expert system for detection of breast cancer based on association rules and neural network Expert Syst Appl, 36 (2) (2009), pp. 3465-3469.

[3] W. H. Wolberg, W. N. Street, O. L. Mangasarian, Breast cancer Wisconsin (diagnostic) data set, UCI ML Repository [http://archive.ics.uci.edu/ml/].

[4] Y. LeCun, Y. Bengio, G. Hinton Deeplearning, nature, 521 (7553) (2015), p. 436, 10.1038/nature14539 https://doi.org/10.1038/nature14539.

[5] R.C. González, R.E. Woods, S.L. EddinsDigital image processing using MATLAB Pearson (2004).