**CHAPTER 1**

**INTRODUCTION**

* 1. **Background**

Cancer has become a deadly disease and more people are suffering from Cancer and a survey says one in every 30 women suffer from this disease in their lifetime and so basically the project was first thought of because of the increase in cases of breast cancer and one thing which is very important is that if we can detect the Cancer at an early stage then there is an increased chances of it getting cured. So this project lays a foundation in making the detection of the cancer automated so that more and more people can get it diagnosed early so as get cured.

Breast cancer is the most common form of cancer in women, and Invasive Ductal Carcinoma (IDC) is the most common form of invasive breast cancer. It accounts for 55% of breast cancer incidence upon diagnosis. Breast carcinomas arise from the same segment of TDLU. In general, breast carcinoma is divided into DCIS and IDC. Accurately identifying and categorizing breast cancer subtypes is an important clinical task, and automated methods can be used to save time and reduce error. Breast cancer is a major public health problem in developing countries. It is one of the leading causes of death for women globally [1].

Breast cancer is uncontrolled growth of breast cells. It is not only found in breast cells but also in many parts of the body. It forms lumps in the ducts which carry milk. A small number of cancers start in other tissues in the breast. There are almost 6 stages of breast cancer. It is always found that the detection of cancer at the first stage can cure it. A simple image is taken as an input and compared with the image already stored in the database detection with cancer. Pre-processing is done on that image. If the detection is found successful then corresponding Treatment is suggest. The stage of cancer is been demonstrated and respectively treatment is been advised to the patient. A simple image is taken and using machine learning, system is given instructions to perform like human so that it can compare and detect cancer.

The World Health Organization's International agency for Research on Cancer in Lyon, France, estimates that more than 150 000 women worldwide die of breast cancer each year. The breast cancer is one among the top three cancers in American women. In United States, the American Cancer Society estimates that, 215 990 new cases of breast carcinoma has been diagnosed, in 2004. It is the leading cause of death due to cancer in women under the age of 65. However, cervical cancer is still number one in rural India. Although the incidence is lower in the developed countries, the burden of breast cancer is alarming. Organ chlorines are considered a possible cause for hormone-dependent cancers. Detection of early and subtle signs of breast cancer requires high-quality images and skilled mammographic interpretation. In order to detect early onset of cancers in breast screening, it is essential to have high-quality images. Radiologists reading mammograms should be trained in the recognition of the signs of early onset of, which may be subtle and may not show typical malignant features. Mammography screening programs have shown to be effective in decreasing breast cancer mortality through the detection and treatment of early onset of breast cancers. Emotional disturbances are known to occur in patient's suffering from malignant diseases even after treatment. This is mainly because of a fear of death, which modifies QOL. Most imaging studies and biopsies of the breast are conducted using mammography or ultrasound, in some cases, magnetic resonance (MR) imaging. Although by now some progress has been achieved, there are still remaining challenges and directions for future research such as developing better enhancement and segmentation algorithms.

In this project, we propose using an image recognition system that utilizes a convolutional neural network in order to detect and classify abnormalities in mammograms. In general, a mammogram is classified as either normal, benign (non-cancerous abnormality), or malignant (cancerous abnormality). In practice, it can be difficult for patients to obtain a quick diagnosis regarding a breast abnormality solely from a doctor or radiologist’s examination of a mammogram. A breast biopsy is usually required before a medical professional can make a diagnosis. As a surgical procedure, there exist risks and side effects such as chronic pain and infection that may result from receiving a biopsy. Moreover, scheduling a biopsy, finding a doctor, and waiting for the lab results all prolong the time required to make a diagnosis, which may give the cancer enough time to leave irreversible effects on the patient’s health.

In the context of Nepal, breast cancer is a substantial social and economic burden. Prevention has been demonstrated to be among the most effective long-term strategies to lessen the increasing chronic disease burden [3]. However, due to socioeconomic disparities and insufficient financial resources, to date, the prevention of breast cancer has not been well conducted in Nepal. As a developing nation, Nepal is faced with several challenges with regards to the care of patients with breast cancer with inadequate funding; the uneven distribution of resources and services; inadequate numbers, training and distribution of health-care personnel and equipment; and a lack of adequate care for many populations based on socioeconomic and geographic factors. In the present review, the epidemiological characteristics, risk factors and current breast cancer awareness and screening efforts in Nepal are summarized. Additionally, alternative ways to improve the prevention of breast cancer in Nepal are discussed.

* 1. **Motivation**

Breast cancer is the revelatory health problem among the world and becomes the most common malignancy in women. Breast cancer places a substantial burden on the healthcare system, but information regarding the number of women living with breast cancer is not well recorded. Especially in countries with lower levels of resources, breast cancers are commonly diagnosed at late stages and women may receive inadequate treatment, pain relief or palliative care. We consulted many friends, teachers and seniors as well as pursued various similar documents online on the process of determining a satisfying and suitable project. We realized that breast cancer detection project is good for this level. Also, our senior who had previously built similar system recommended on developing a similar project, and all of these motivated us to build this project.

* 1. **Statement of Problems**

After making case study on systems related to breast cancer detection, we noted down some of the problems that are faced in the past projects:

* Inadequate number of training and distribution of health-care personnel and equipment
* Lump of less than 10mm are not considered on medical diagnosis, which may probably be cancerous but is not detected during clinical examination.
  1. **Objectives**

The main objective of our project is to build an expert system that predicts whether a person has breast cancer or not.

* 1. **Scopes**

Proposed system is based on CNN. If the system achieves higher accuracy, it can help doctors and highly skilled radiologist in minimizing the false negative cases, which is common because there are very few highly skilled radiologist, and high volume of cases. It can be used in hospital or in clinics as an expert system assisting doctors and radiologist for detecting early stages of breast cancer which will be more effective.

**CHAPTER 2**

**LITERATURE REVIEW**

**Review of Literature**

Mammography is the only reliable screening test proven in breast imaging [1]. Determining best preprocessing technique on the basis of peak signal to noise ratio for set of mammography images [2].During the past year, Breast cancer is standout amongst the most widely recognized disease among the women of the creating nation on the planets, and it has also converted a foremost source of death. Different commitment has been made in literature regarding to utilization of example of pattern recognition techniques for best tumor conclusion in tissue level [3].

Hala al-Shamlan and his group proposed to extract the feature values for analysis breast cancer mammogram images to classify the breast tumor (hala al-shamlan et al. 2010). They are resolved situated essential and discriminating breast cancer feature extraction. They are two process used. First process which is improved the contrast image. Another process is segmentation, which finds the locale for mass identification. They found out good results. The results were the reach qualities expected in every feature extraction. Martin and his group proposed the method for recognition of mas on digitalize mammograms (Martin et l. 2009). They utilized K-means bunching calculation for picture division and dim level co-event grid to portray and break down the surface of divide structures in the picture. The grouping of these structures was attained to through SVM, Which is divided them into 2 gathering; utilizing the shape and composition descriptors, masses and non-messes. The arrangement exactness acquired from that system was 85% [4].

Vishnukumar K. Patel and his group to proposed mammographic images of breast cancer detection by using image enhancement techniques (V.K. Patel et al. 2012). They have used different technique in enhancement algorithm such as frequency domain and spatial domain. A recurrence space smoothing-sharpening system is proposed and its effect is surveyed to supportively enhance mammography image. They have used contrasts enhancement algorithm and using Gabor filter. Gabor filter is used for finding out good image. They took the mammogram image and find out different filter values than after used the PSNR algorithm and found out better results in numerical values. The result pictures turned up new areas which may be of expressive redirection. Combining Gabor figuring with quick Fourier change and overlay cover division showed to be greatly powerful technique for wiping out clamor and updating edges, along these lines improving the sign-to-commotion extent. Superimposition of picture changed using diverse techniques into single picture indicate to 9 profitable upgrade the perceivability and facilities the ID of profitable data to human eye contrast with different routine for separating. Athansiadis I. Emmanouil and his gathering proposed mammographic picture improvement utilizing wavelet based handing and histogram leveling. The purposed of this research was investigate the effectiveness of wavelet transform by processing the DWT detail coefficient with sigmoid function and also used HEMF. They took the 6 parameters and enhancing the images and found out the result 91% overall images. It was watched that the sigmoid wavelet-based channel fulfilled better perception of the bosom skin, the thoracic muscle, vessels, veins and channel, while it improved the separation of the picture representation of average greasy bosoms in alluring level [5].

“Classification of Mammogram Image by using CNN Classifier” by Ketan Sharma and Bobbin Preet, 2016. In this paper they proposed a CAD system named as CNN. They had also compared of CNN with Logistic Regression algorithm [6].

“Whole Mammogram Image Classification with CNN” by Nathan Jacobs, Jinze Liu and Erik Y. Han, 2017. This paper reports preliminary work on developing and optimizing machine learning for whole image classification mammograms. They evaluated 7 different CNN architecture and concluded that combining both data augmentation and transfer learning method with a CNN is the most effective in improving classification performance [7].

“Preprocessing Filters for Mammogram Images” by Kshema and M. Jayesh George, 2017. Preprocessing is the most vital and essential step in the mammogram analysis to improve mammogram picture quality. It is important to redress the mammogram images for further processing and analysis. Filters are utilized to enhance picture quality, evacuate the clamor, saves the edges insides a picture, improve and smoothen the image. Mammogram image corrupted with speckle noise can be better reconstructed with Gaussian filter and adaptive median filter works better for salt and pepper noise [8].

“Mass Classification In mammograms Using Neural Network” by Effa Adrina Azli, Salina Abdul Samad and Modh Faisal Ibrahim, 2017. Different feature affects performance of the classifier so it is important to extract the useful features that are able to distinguish between benign and malignant classes. Besides that the architecture of neural network affects the overall performance of the classification. The architecture with 100 hidden nodes consistently improves the classifier performance by 10% compare to the architecture with only 3 hidden nodes regardless of the input feature fed into the classifier [9].

**CHAPTER 3**

**PROJECT MANAGEMENT**

Time and resource management are the two most important things to be considered for developing an effective project. Beside this, team must be formed and work load must be divided among the team members. To complete the task of developing information management system using face detection and face recognition successfully in a proper managed way and present it in time. We formed a team of five members.

**3.1 Team Members**

1. Bina Kandel (0720311)
2. Niru Manandhar (0720321)
3. Rabin Prajapati (0720329)
4. Sajid Shrestha (0720337)
5. Sujan Katwal (0720346)

**3.2 Gantt chart**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S.N | Week  Strategies | 1st week | 2ndweek | 3rd week | 4th week | 5th week | 6th week |
| 1) | Research and analysis |  |  |  |  |  |  |
| 2) | Establishing Goals |  |  |  |  |  |  |
| 3) | Design a model |  |  |  |  |  |  |
| 4) | Coding and Debugging |  |  |  |  |  |  |
| 5) | Implementation & testing |  |  |  |  |  |  |
| 6) | Documentation |  |  |  |  |  |  |

Table 3.2: The expected Gantt chart of the project

**CHAPTER 4**

**METHODOLOGY**

A convolutional Neural Networks is a combination of many types of layers. Input layer is the image data that we input by pixel. In this model, histology image is the input, so it will have 50 x 50 input data for the CNN. In hidden layer, there are many layers included such as convolutional layer, rectified linear unit (ReLU) layer, Pooling layer, Fully-connected (Dense) layer.

**4.1 Block Diagram:**

Test Images

Preprocessing

(Segmentation)

CNN

(Feature Extraction)

CNN based classifier

Prediction results

Training Dataset

Preprocessing

(Segmentation)

CNN

(Feature Extraction)

Loss function

Trained CNN

Gradient Descent

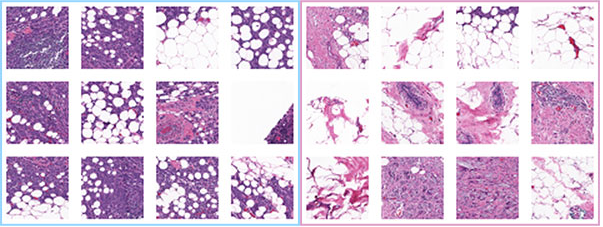
Back propagation

Fig 4.1: Training Fig 4.2: Testing

**Training Dataset:**

Positive

Negative



The dataset was originally curated by [Janowczyk and Madabhushi](https://www.ncbi.nlm.nih.gov/pubmed/27563488) and [Roa et al.](http://spie.org/Publications/Proceedings/Paper/10.1117/12.2043872) but is available in public domain on [Kaggle’s website](https://www.kaggle.com/paultimothymooney/breast-histopathology-images).

The original dataset consisted of 162 slide images scanned at 40x. For our system we have converted the images to 50×50 pixel.

Slide images are naturally massive (in terms of spatial dimensions), so in order to make them easier to work with, a total of 277,524 patches of 50×50 pixels were extracted, including:

* 198,738 negative examples (i.e., no breast cancer)
* 78,786 positive examples (i.e., indicating breast cancer was found in the patch)

Each image in the dataset has a specific filename structure. An example of an image filename in the dataset can be seen below:

10253\_idx5\_x1351\_y1101\_class0.png

We can interpret this filename as:

* Patient ID: 10253\_idx5
* x-coordinate of the crop: 1,351
* y-coordinate of the crop: 1,101
* Class label: 0 (0 indicates no IDC while 1 indicates IDC)

**Building Breast Cancer Dataset:**

Our breast cancer image dataset consists of **198,783 images**, each of which is 50×50 pixels. We have splitted our deep learning breast cancer image dataset into training, validation, and testing sets.

Convolution

Convolution

Pooling

Pooling

Fully Connected

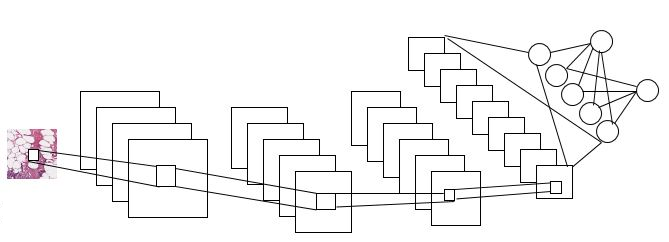


Fig 4.3 : CNN steps

Convolutional layers applied specified number of convolution filters to the image. For each sub region, the layer performs a set of mathematical operations to produce a single value in the output feature map. Convolutional layers then typically apply a ReLU activation function to the output to introduce nonlinearities into the model. This model has an input of 50 x 50 matrixes and filters it into 32 features image by using kernel size of 3 x 3 filter with ReLU activation. Then the result is pooled to take only important features from the matrices.

Fig 4.4 : Convolution

utio

There are three main types of layers used to build CNN architectures; (1) convolutional layer, (2) pooling layer, and (3) fully connected (fc) layer.

* **Input layer:** This layer loads whole breast cancer histology images and produces outputs that feed to the first convolutional layer. The input images are composed of three 2D arrays in the 8-bit depth of red-green-blue channels.
* **Convolutional layer:** This layer extracts features by computing the output of neurons that connect to local regions of the input layer or previous layer. The set of weights which is convolved with the input is called filter or kernel. The size of every filter is 3 × 3, 5 × 5 or 7 × 7. Each neuron is sparsely connected to the area in the previous layer. The distance between the applications of filters is called stride. The hyperparameter of stride is set to 2 that are smaller than the filter size. The convolution kernel is applied in overlapping windows and initializes from a Gaussian distribution with a standard deviation of 0.01. The last convolutional layer is composed of 128 filters that initialize from Gaussian distributions with a standard deviation of 0.0001. The values of all local weights are passed through ReLU.
* **Pooling layer:** Max pooling is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.

This is done to in part to help over-fitting by providing an abstracted form of the representation. As well, it reduces the computational cost by reducing the number of parameters to learn and provides basic translation invariance to the internal representation.

Max pooling is done by applying a max filter to (usually) non-overlapping sub regions of the initial representation.

Fig 4.5 : Max Pool

utio

* **Fully Connected:** Neurons in a fully connected layer have full connections to all activities in the previous layer, as seen in regular Neural Networks. Their activities can hence be computed with a matrix multiplication followed by a bias offset.

**ReLU Activation**

In a neural network, neuron output or activation function generally has the form f(x) = tanh(x) or the sigmoid function for some input x. These neuron activation functions are by definition saturating because they ultimately map their input to an output between a fixed range such as [−1, 1]. Neurons that use non-saturating activation functions such as f(x) = max(0,x) are called Rectified Linear Units and can be used in a Convolutional Neural Network.

Fig 4.6 : ReLU activation

utio

**Softmax Function**

Softmax function, also known as softargmax or normalized exponential function, is a function that takes as input a vector of K real numbers, and normalizes it into a probability distribution consisting of K probabilities proportional to the exponentials of the input numbers. That is, prior to applying softmax, some vector components could be negative, or greater than one; and might not sum to 1; but after applying softmax, each component will be in the interval (0,1) , and the components will add up to 1, so that they can be interpreted as probabilities. Furthermore, the larger input components will correspond to larger probabilities. Softmax is often used in neural networks, to map the non-normalized output of a network to a probability distribution over predicted output classes.

for i =1, …,K and z = ()

**Cost Function**

A Cost Function/Loss Function evaluates the performance of a machine learning algorithm. The Loss function computes the error for a single training example while the Cost function is the average of the loss functions for all the training examples.

**Binary Cross entropy**

Binary cross entropy is a loss function used on problems involving yes/no (binary) decisions. For instance, in multi-label problems, where an example can belong to multiple classes at the same time, the model tries to decide for each class whether the example belongs to that class or not.

L(y,) = -

Where, ŷ is the predicted value.

Binary cross entropy measures how far away from the true value (which is either 0 or 1) the prediction is for each of the classes and then averages these class-wise errors to obtain the final loss.

**Gradient Decent**

Gradient descent is an iterative method. We start with some set of values for our model parameters (weights and biases), and improve them slowly. To improve a given set of weights, we try to get a sense of the value of the cost function for weights similar to the current weights (by calculating the gradient). Then we move in the direction which reduces the cost function. By repeating this step thousands of times, we’ll continually minimize our cost function.

**Back Propagation**

Back propagation is based around four fundamental equations. The back propagation equations provide us with a way of computing the gradient of the cost function. Back propagation in the form of an algorithm is given below:

1. **Input**x**:** Set the corresponding activation  for the input layer.
2. **Feed forward:** For each l=2,3,…,L compute

And)

1. **Output error:** Compute the vector C ⊙ ()
2. **Back propagate the error:** For each *l*=*L*-1,*L*-2,…,2 compute

*(*

1. **Output:** The gradient of the cost function is given by

We compute the error vectorsbackward, starting from the final layer. The backward movement is a consequence of the fact that the cost is a function of outputs from the network. To understand how the cost varies with earlier weights and biases we need to repeatedly apply the chain rule, working backward through the layers to obtain usable expressions.

**Adam Optimizer**

Adam can be looked at as a combination of RMSprop and Stochastic Gradient Descent with momentum. It uses the squared gradients to scale the learning rate like RMSprop and it takes advantage of momentum by using moving average of the gradient instead of gradient itself like SGD with momentum. Let’s take a closer look at how it works.

Adam is an adaptive learning rate method, which means, it computes individual learning rates for different parameters. Its name is derived from adaptive [moment](https://en.wikipedia.org/wiki/Moment_(mathematics)) estimation, and the reason it’s called that is because Adam uses estimations of first and second moments of gradient to adapt the learning rate for each weight of the neural network

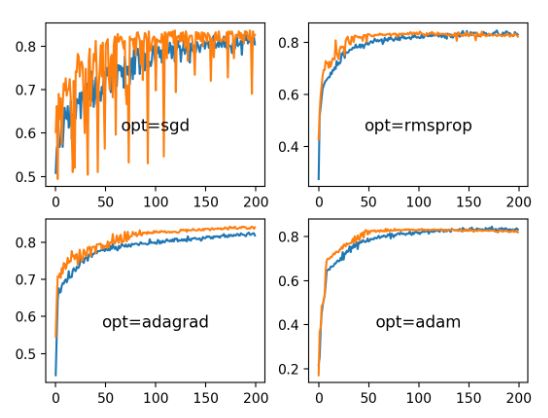


Fig 4.7 : Performance of different optimizer

**Dropout**

Dropout refers to ignoring units (i.e. neurons) during the training phase of certain set of neurons which is chosen at random. By “ignoring”, I mean these units are not considered during a particular forward or backward pass.

More technically, at each training stage, individual nodes are either dropped out of the net with probability 1-p or kept with probability p, so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed.

We need dropout to literally shut-down parts of a neural networks so that it prevent our neural network from over fitting

Fig 4.8 : Standard Neural Net

Fig 4.9 : After applying dropout

**Learning Rate**

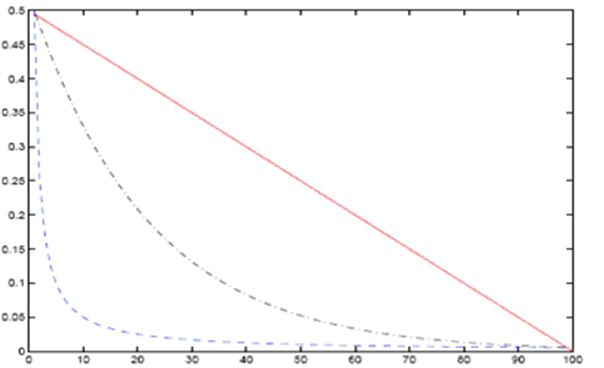
The learning rate is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated. Choosing the learning rate is challenging as a value too small may result in a long training process that could get stuck, whereas a value too large may result in learning a sub-optimal set of weights too fast or an unstable training process. The standard weight update formula used by nearly all neural networks is given by

W = W + α \* gradient

The learning rate, α, controls the “step” we make along the gradient. Larger values of \alpha imply that we are taking bigger steps. While smaller values of α will make tiny steps. If α is zero the network cannot make any steps at all (since the gradient multiplied by zero is zero).

A network is then trained for a fixed number of epochs without changing the learning rate. This method may work well in some situations, but it’s often beneficial to decrease our learning rate over time. Instead of keeping our learning rate constant throughout the whole we can decrease our learning rate, thereby allowing our network to take smaller steps. This decreased learning rate enables our network to descend into areas of the loss landscape that are “more optimal” and would have otherwise been missed entirely by our learning rate learning

Learning rate = init\_lr \* (1/1+decay\*iterations)



Learning Rate

Time Step

Fig 4.10 : Linear Decay Learning Rate

**Image augmentation**

Image data augmentation is a technique that can be used to artificially expand the size of a training dataset by creating modified versions of images in the dataset.

* **rotation\_range :** Int. Degree range for random rotations.
* **brightness\_range :** Tuple or list of two floats. Range for picking a brightness shift value from.
* **zoom\_range** : Float or [lower, upper]. Range for random zoom. If a float, [lower,upper]=[1-zoom\_range,1+zoom\_range]
* **horizontal\_flip :** Boolean. Randomly flip inputs horizontally.
* **vertical\_flip :** Boolean. Randomly flip inputs vertically.
* **Rescale :** rescaling factor. Defaults to None. If None or 0, no rescaling is applied; otherwise we multiply the data by the value provided (after applying all other transformations).
* **preprocessing\_function :** function that will be implied on each input. The function will run after the image is resized and augmented. The function should take one argument: one image (Numpy tensor with rank 3), and should output a Numpy tensor with the same shape.

**4.2 Tools and Platform**

* Python as platform for coding
* Flask micro framework

**CHAPTER 5**

**RESULT AND DISCUSSION**

* 1. **Discussion:**

The CNN Breast Cancer model prepared is trained on CPU for approximately 20 hours for 111,010 **images which result is as shown below:**

Fig 4.11 : Training and Validation Accuracy/ Loss

We can see that our model achieved **84% accuracy** in which we classified benign/no cancer correctly 93% of the time but we classified malignant/cancer only 68%.

Our sensitivity measures the proportion of the true positives that were predicted correctly with accuracy of 84.44% whereas specificity measures our true negatives which were predicted with accuracy of 83.14%.

**Confusion Matrix**

Fig 4.12 : Confusion Matrix

* **True Positives (TP)** - These are the correctly predicted positive values which mean that the value of actual class is yes and the value of predicted class is also yes.
* **True Negatives (TN)** - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no.

False positives and false negatives, these values occur when your actual class contradicts with the predicted class.

* **False Positives (FP)** – When actual class is no and predicted class is yes.
* **False Negatives (FN)** – When actual class is yes but predicted class in no.

**Breast Cancer Classification Report**

Fig 4.13 : Breast Cancer Classification Report

* **Accuracy** - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best.

Accuracy = TP+TN/TP+FP+FN+TN

In our case, accuracy is about 84%.

* **Precision** - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

Precision = TP/TP+FP

In our case, precision is about 86%.

* **Recall**(Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class.

Recall = TP/TP+FN

In our case, recall is about 84%.

* **F1 score** - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it’s better to look at both Precision and Recall. In our case, F1 score is 0.701.

F1 Score = 2\*(Recall \* Precision) / (Recall + Precision)

In our case, F1score is about 85%.

**5.2 Remaining tasks**

The remaining tasks for this project can be summarized as:

* Hyperparameters tuning to improve accuracy
* To reduce class imbalance

**CHAPTER 6**

**REFERENCES**

**Paper in Conference**

[1] Sachin Prasad N and Dana Houserkova ,”The role various modalities in breast imaging”, 2007.

[2] Jaya Sharma, R. P. Tewari and J. K. Rai, “Identification of Preprocessing Technique for Enhancement of Mammography Images “, 2014.

[3] American Cancer Society, “Cancer prevention and early detection facts and figures”, 2003.

[4] Hala S, Ali E, “Feature Extraction Values for Breast Cancer Mammography Images”, IEEE 353-340, 2010.

[5] Kingsbury N, “Wavelet Transforming in image processing. Signal Analysis and Prediction: 27”, 1998

[6] Ketan Sharma and Bobbin Preet, “Classification of Mammogram Images Using CNN”, 2016.

[7] Nathan Jacobs, Jinze Liu and Erik Y. Han,“Whole Mammogram Image Classification with Convolution Neural Networks”, 2017.

[8] Kshema and M. Jayesh George, “Preprocessing Filters for Mammogram Images”, 2017.

[9] Effa Adrina Azli, Salina Abdul and Mohd Faisal Ibrahim, ”Mass Classification In Mammogram Using Neural Networks ”, 2017.

[10] Rikiya Yamashita, “Convolution neural networks: An overview and application in radiology”, 2018.

**APPENDIX**

Appendix A.1 : Home Page

Appendix A.2: Prediction

Appendix A.3 : Training Process

Appendix A.4 : Console Output

Appendix A.5 : Model Summary