CONVOLUTIONAL NEURAL NETWORK IMPROVEMENT FOR BREAST CANCER CLASSIFICATION

CS4099D Project

Endsem Report

Submitted by

TANNA SASIDHAR	**(B170849CS)**
TAMMINEDI AVINASH	**(B170801CS)**
KANDULA VENKATA MANIKANTA REDDY	**(B170963CS)**

Under the Guidance of

Prabu Mohandas

Assistant Professor

Department of Computer Science and Engineering

Department of Computer Science and Engineering
National Institute of Technology Calicut
Calicut, Kerala, India - 673 601

May, 2021

signature
Prabu Mohandas

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT
KERALA, INDIA - 673 601

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that this is a bonafide report of the project work titled

CONVOLUTIONAL NEURAL NETWORK IMPROVEMENT FOR BREAST CLASSIFICATION

done by

TANNA SASIDHAR	**(B170849CS)**
TAMMINEDI AVINASH	**(B170801CS)**
KANDULA VENKATA MANIKANTA REDDY	**(B170963CS)**

of Eighth Semester B. Tech, during the Monsoon/Winter Semester 2020-'21, in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering of the National Institute of Technology Calicut.

16-05-2021 | M PRABU
Date | **Project Guide**

DECLARATION

I hereby declare that the project titled, **Convolutional neural network improvement for breast cancer classification**, is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or any other institute of higher learning, except where due acknowledgement and reference has been made in the text.

Place :Vishakapatnam
Date :16-05-2021

Name :TAMMINEDI AVINASH
Roll. No. : B170821CS

Place :Parvathipuram
Date :16-05-2021

Name :TANNA SASIDHAR
Roll. No. : B170849CS

Place :MARKAPUR
Date :16-05-2021

Name :K.V.MANIKANTA REDDY
Roll. No. : B170963CS

Abstract

Usually doctors have to manually check the mammograms and understand the effected area of Breast tissue. Manual segmentation of mammogram takes lot of time and doesn't guarantee the right result. It is really important to classify the mammogram correctly so that the doctor can give the appropriate treatment to the patient at right time. The algorithm called CNNI-BCC is used to help doctors in breast cancer treatment. The trained CNNI-BBC model identifies the affected regions of breast tissue and also classifies the cancer region. The CNNI-BCC uses a convolutional neural network that improves the breast cancer lesion classification, It can classify the incoming breast cancer medical images into malignant, benign, and no cancer. CNNI-BCC can categorize incoming medical images as malignant, benign or normal patient with sensitivity, accuracy.

ACKNOWLEDGEMENT

In today's world deep learning has great applications and lot of problems are being solved by it. Deep learning is currently receiving a lot of attention due to its application on health care sector. From this project we are learning a lot about deep learning algorithms and how these algorithms are solving today's problems. We are very much thankful to our guide M.PRABHU sir for the opportunity to letting us work with him. We are also thankful to the Department of CSE and faculty for giving this opportunity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the recent years breast cancer has been the one of the most occurring cancers in the women. This is a cancer that develops from breast tissue. In the adult stages of women life, breast tissues consists of milk glands, tissues and fat. In case of breast cancer these breast cells will multiply rapidly. Usually cancer cells do not die at the normal point in their life cycle. This excessive cell growth in the breast tissue causes cancer. The tumor developed uses all the nutrients and energy and deprives the cells around it thereby causing the cancer. Some of the symptoms of breast cancer are pain in the armpits, Lump in the breast, or no breast evolution with the monthly cycle. Some other symptoms include a change in shape of the breast, fluid from the nipple. There are several treatments for Breast cancer. The treatment mainly depends on the kind of breast cancer like benign or malignant. Some of the treatments are Surgery, Chemotherapy, Radiation therapy, Hormonal therapy.

Mammography is a technique used to examine the breast of a woman. A Mammogram is simply an X-ray of the breast tissue. It helps doctors to look for changes in the breast tissue. Abnormal areas in the breast tissue can be found in the mammogram, but doctors won't be able to tell if the abnormal area is cancer or not by just looking at the mammogram. Mammograms help in finding breast cancer at early stages and reduces the risk of dying if proper treatment is given.

Generally, doctors compare new mammograms with old mammograms for identifying changes in the breast tissue, if there are no changes in the mammograms then the chances of cancer will be relatively low. If there are any changes in the mammogram like masses, calcifications then there will be chances of cancer in the breast tissue.

Convolutional Neural Network has brought enormous improvements in the field of Computer vision, mainly in the field of medicine. Even after finding the abnormality, It is difficult for the doctors to classify as cancer or not. The trained CNN models helps doctors in classification of mammogram as cancer or no cancer, or even in classification of the type of cancer as benign or malignant. This classification of cancer by CNN helps the doctors to treat the patient appropriately for the particular type of cancer. The CNN also detects the benign stage of cancer, which is very difficult to find manually by the doctors, benign stages are the early stages of cancer and detecting them helps doctors to treat patients easily thereby saving lives.

Chapter 2

Literature Survey

Breast Cancer Classification is a classification task, which involves categorising mammograms. This might be more difficult than it seems as it usually depends on the mammogram. It involves various studies on Deep Learning, especially on computer vision sector.

Deep learning is a class of Artificial Neural Network, which is also part of Machine Learning, like Multilayered Human Cognition system. There are vast number of applications of Deep Learning in this Health Sector.

However, [1] Accessibility of Big Data, improved processing control with GPU, numerous impediments of Artificial Neural Networks have been solved with Deep Neural Networks. These approaches showed good performance in imitating humans. Medical imaging is also utilizing to identify structural abnormalitites and classify them into disease types and categories. In the context of Picture Archiving and Communication Systems, CAD systems were applied.

In recent years, Swetha Saxena et al. [2] from NIT, Bhopal proposed Machine Learning Methods for Diagnosis of Computer Aided Breast Cancer using Histopathology. In this, cancer effected tissues are extracted and observed under microscope then compared with Histopathology slide which contains intricate visual patterns that are used to identify the cancer as benign or malignant. For this pattern recognition, Machine Learning models with CAD systems were applied. [9]It also gives the process of classification i.e, Preprocessing, feature extraction and selection, classification and Analysis of performance. [11]Through this preprocessing, the dataset has to be increased so as to improve the accuracy of the model.

Puja Gupta, Shruti Garg et al. [3] from BITS, Ranchi proposed a model that uses various Parameters for Breast Cancer Prediction techniques. In this, six different ML Algorithms were used to classify the tumor cells and predicted their accuracies.
Karabatak et al.[7]. used association rules along with Neural Network in order to train the model then applied cross validation to increase accuracy.
Payam et al.[8] used some Data Preprocessing techniques besides Data reduction in order to increase the data set which there by increases the accuracy. It also stated that a classifiers with any kind of classification Machine Learning model can predict more or less equally, so selection of an appropriate model for any problem is difficult.

Anji Reddy, Sudheer Reddy[4] from NIT, Silchar proposed a model that detect the Breast Cancer by levaraging Machine Learning. They introduced a new method called Deep Neural Networks with Support Value (DNNS) to get better quality Mammograms in order to fix some other performance parameters. The main idea behind this method is to improve the quality of mammograms for better recognition/prediction which there by increases accuracy. In this they also used rotation technique to increase the Data Set by rotating the Mammograms by 90, 180, and 270.

Pin Wang and Co et al.[5], proposed a model that uses Cross task extreme learning machine for classifying Breast Cancer images using Deep Convolutional features. In this, they build a special structure called Hybrid Structure which is a Double Deep Transfer Learning learning(D2TL) and a new machine called Interactive Cross table extreme learning machine. This machine significantly uses both feature representation ability of Convoluted Neural Networks and classification ability of ELM[13, 14]. It give higher accuracy than expected since it uses hybrid structure for classification. This also stated that the limited number of Mammograms won't be sufficient for effective distinguishing the cancer cells using Deep Convolution Neutral Network. Hence, increasing the Data-set by Pre processing to better results is mandatory.

Y Wang and Co et al.[6] proposed a Machine Learning model that classifies Breast Cancer in Automated Breast ultrasound using multi view Convolution Neural Network with Transfer learning. This speaks about the importance of Computer-Aided Diagnosis(CAD) in classifying Mammograms. In this method, CAD systems observes the same extracted Breast lesions in different mammographic views and gives useful features independently. It is similar to pre-processing but here there is no manual preprocessing step and the model directly extract features from lesion patch.

Dragana Djilas and Co et al. [10] proposed a model that compares three methods for early detection of breast cancer. Those three methods are Breast MRI, digital Mammography and breast tomosynthesis. These three methods are taken for analysing Breast Cancer Classification and compared the results of those methods. There was a notable difference in the performance of three methods.
Mammography gave low performance when compared to Breast tomosyn-

thesis because of higher background noises. Hence, it was concluded that performance using digital Mammograms can be increased by removing those background noises.
Kwang Gi Kim et al.[12] gave a research paper that deals about Deep Learning. It describes what Convolution is and explained the motivation behind the process called Convolution in Neural Networks and also explains the process of Pooling. It also describes different applications of different Deep Learning models.

Chapter 3

Problem Definition

In the earlier CNN models, the model is only used to classify whether the cancer is present or not, but that model didn't know the level of cancer if it is present. Further models were build to classify the breast cancer as bengin, malignant or healthy. This type of classification helps the doctors to give appropriate treatment to the patient based on the level of breast cancer. So, there will be better chances for the patient to survive if patient is given the appropriate treatment. This model also accounts for the increase in accuracy than that of the previous models.

Problem Statement 1: Convolution Neural Network based classifier model is build so as to classify the mammograms as Benign, Malignant or healthy person. This model just simply implements direct mammograms of benign, malignant and healthy people without any pre-processing.

Input : mammogram of size 1024 X 1024 pixels.
Output : Benign, malignant, or healthy person.

Problem Statement 2: Preprocessing the benign and malignant mammograms to build a CNN model. This process cannot preprocess the healthy patient mammogram as they donot have any lesions.

Input : 1024 x 1024 pixel Mammogram image.
Output: 8 images each of size depending on radius of the lesion.

Problem Statement 3: Build a CNN model to classify benign or malignant using preprocessed images. Compare the accuracies of CNN models without image pre-processing and with image pre-processing.

Input : Pre-processed mammograms.
Output: Benign, malignant.

Chapter 4

Methodology

We have gone through the paper thoroughly and we also made some research about the breast cancer. From this paper we figured out all the problems and the corresponding inputs and outputs to each of the following problem.

CNNI-BCC is used for detecting lesions. This helps for diagnosing breast cancer. Model classifies mammograms into malignant, benign and healthy.

CNNI-BCC consists of (1) feature wise pre-processing, (2)Convolutional neural network -based classification.

4.1 Mammogram pre-processing

Mammograms that are used in this model are of high size i.e 1024 x 1024 pixel. So, lot of computational power is required inorder to process the whole mammogram size. Hence, it is necessary to reduce the process time of CNN model by Pre-processing the input mammograms. For this, there is a process called Mammogram Pre-processing. It divides larger images into smaller images then these images are rotated every angle(0-360) and flipped vertically so as to increase the Dataset. Hence, a single input mammogram generates multiple mammograms, which are used for the training of the model.

4.2 Convolutional Neural Network Classification

In this step, the mammogram and the features from fwda go through convolution layer, ReLU layer, polling layer and fully connected layer. With softmax function we will classify the tumor as benign, malignant or healthy person.

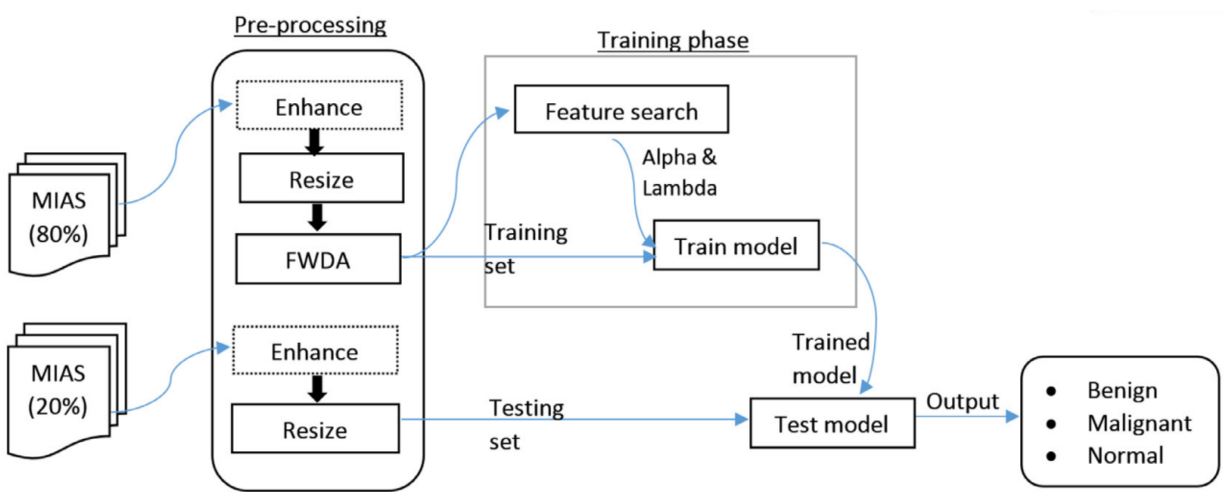

Figure 4.1: CNNBS

The above diagram shows how CNNBS process occurs. Initially, we split MIAS Dataset and preprocess the input mammograms, which includes Enhance, Resize and FWDA then these mammograms are sent into training phase where we train the model and the remaining mammograms are used for testing. Hence, it classifies the input mammograms into Benign, Malignant and Normal.

4.3 Packages Used

1. OS
This package is mainly used to perform tasks on operating system.
This package provides functions for manipulating directories like fetching its contents, identifying and changing the current directory, etc.

2. from keras.models import sequential

The Sequential model is used to create a way of deep learning models. We create an instance of the Sequential class and multiple layers are created and added to this model. The Sequential model is really useful for developing deep learning models in any kind of situations, but it also come with some limitations.

3. from keras.layers -dense, conv2d, maxpool, flatten, dropout
These are the main layers used to build the convolutional neural network model.
4. numpy: NumPy is a library used to work with arrays in python.

5. from pil import image PIL is Python Imaging Library, it is malinly used for opening, manipulating, and saving many different image files.

6. from sklearn.model.selection import train_test_split
train-test-split is a function used to split the dataset for training and testing,

this function also allows us to partition the dataset into required percentages (eg: 70 percent training and 30 percent testing).

7. from keras.utils import to-categorical keras. utils. to-categorical Converts a class vector to binary class matrix.

4.4 Design Model 1

4.4.1 Accessing Dataset

Info.txt : In this file, we have metadata related to the breast cancer all-mias dataset.

The contents of 'Info.txt' are image-id, background tissue, type of abnormality present, the abnormality's severity, x-axis, y-axis coordinates of the abnormality centre, and the abnormality's radius measured in pixels.

Extract-label: we are going to create a dictionary of image-id to type of cancer for every image. We need to map every image-id to benign(1), Malignant(2), No cancer(3) using type or abnormality data(column 3) present in 'Info.txt' and returning the dictionary.

Extract-Image: in this function we are creating a dictionary of image-id to image. We are resizing every image to the same size for later usage. To get each image we use os.getcwd (to get current directory path). For image in the directory we map image-id to resized image and return the dictionary.

Split: In this function, we make two dictionaries 'labels' and 'images'. We get these dictionaries from the extract-label and extract-image functions.

Now we create two numpy arrays ‘X’ and ‘Y’ in ‘X’ we have all the images in the order of their ids and in ‘Y’ we have the corresponding labels(1 or 2 or 3 corresponding to benign, malignant or no cancer). Now we divide the data into training and testing using the train-test-split function. We divide 80 percent of data into training and 20 percent of the data is kept for testing. Finally, we return the X-train, Y-train, X-test, Y-test values.

4.4.2 Building Model

Convert y-train and y-test into one hot encoding using to-categorical function.
Size of the image sent into the cnn is 64 x 64 (resized).

model = Sequential()
Sequential model is used for a stack of layers where each layer has exactly one input tensor and one output tensor.

model.add(Conv2D(filters=32, kernel-size=(3, 3), activation='relu'))
The numbers of filters that convolutional layers will learn from is 32. For each filter of size 3x3 dot product with all the sub matrices of size 3x3 in the input image. Every node of the images will go through the relu activation function ($f(x) = \max(x, 0)$) output matrix will be formed.

model.add(MaxPool2D(pool-size=(2, 2)))
In the pool layer for every stride(2x2), all the maximum values from each of the stride will generate a new output matrix.

Similarly, add another 2 sets of Conv2D layer with filter size 32 with window-size of (3, 3) with ReLU activation function, a pooling layer with pool size (2, 2) and output matrix is sent to the flatten layer.

model.add(Flatten())
In a flatten layer all the nodes will form a pile or stack.

model.add(Dense(3, activation='softmax'))
In Convolution Neural Network models, we mostly use softmax activation function for multi-class classification. The softmax function outputs either 0 or 1.

Adam optimizer is used in the compilation of the model because we need to categorize and we have multiple classes, the loss is calculated with "categorical-crossentropy".

Adam Optimizer : Unlike other optimisers, In Adam optimizer a learning rate is maintained for each network weight. It uses Root Mean Square Propagation for optimizing.

Compile the model and train the model using fit function. Predict the cancer for all the X_test values and calculate the accuracy using evaluate function.

modeltesting.py Initially we have stored the cnn model in "bcc_cnn.pkl" using pickle package and load this model in this file. Take single image as an input and resized it and send it to the model to classify as begnin, malignant and no cancer.

4.5 Preprocessing

4.5.1 Dataset Preprocessing

The pre-processing of mammogram images is an important task before training a convolutional neural network model. The pre-processing basically consists of Noise cancellation, contrast enhancement and breast segmentation. The raw mammogram images contain noise that can be removed by some noise cancellation techniques like FNLM denoising algorithm. Breast segmentation usually clears the background areas and labels of the mammograms. There should be some difference between the background pixels and foreground pixels of the mammogram image. While pre-processing we should make sure important information in the mammogram image is not lost.

In this model, for pre-processing initially we cut the mammograms into 128 x 128 pixels and rotated each image for every angle (0 to 360).

4.6 Data Augmentation

Developing a convolutional neural network requires a sufficiently large dataset to train the model, most of the standard datasets available like the MIAS have very small data to train and test. The processing of the large mammogram is also computationally huge for the personal computers. So, the ROI's (Region Of Interest) are segmented from the mammograms to reduce the computation of the CNN model. With the help of the data available in the MIAS dataset, given the center of the lesion and the radius of the lesion, the ROIs can be carved out of the mammogram images using the python tool called pillow. The ROIs from the benign and malignant mammogram images are cropped with the center and radius and are then rescaled to a particular resolution (x * x) and are stored in a folder for further processing. If the dataset has very little data for training the CNN model then, Data augmentation is an effective solution to increase the generalization and performance of the CNN model, In Data augmentation we create new sample images by applying some image transformations like flipping and rotations to increase the dataset. In flipping, the mirror images of the present images are generated and are added to the dataset. In rotation, the whole dataset along with the mirror images are rotated to angles 90, 180 and 270 degrees. The images can be rotated to other angles also so as to increase the size of the dataset for training and testing. This type of augmentation generates relevant training samples as the cancerous tumors captured in the mammograms can be in any orientations(in any angle).

Crop: From the info file in the MIAS dataset we get the center and radius of each of the lesions for benign and malignant mammograms. If there is no center and radius in the tuple of info file then it is no cancer mammogram, such mammograms should be excluded from the preprocessing of the dataset. The mammograms with radius and center are benign and malignant will be cropped by crop method from image class accordingly.

Crop() method is imported from the PIL library. It takes left, top, right, bottom as arguments in order to adjust the size of the image in a rectangular shape.

If the coordinates of the center are x, y from the left-bottom of the image (given in the dataset) and radius is r. im.crop((x - r, y - r, x + r, y + r)) is used to crop the image with coordinates as Left x - r, top y - r, right x + r and bottom y + r margins.

Rotate_image(angle, dir):
This function is used to rotate the image in the required angle. The rotate() method is imported from PIL library and it takes the Number of Degrees as argument and rotates the image in counter clockwise direction present in a directory 'dir'.
Eg: img.rotate(angle).save("C://Users//dell//Desktop//bccn//out//" + str(angle)

+ " " + image)

We rotated our images present in the DataSet in different directions like 90, 180 and 270 and we saved the images in different folders which will be used for further classification.

Eg: Rotate_images(90, out2)
Rotate_images(180, out2)
Rotate_images(270, out2)

Mirror(): This function will flip every image from left to right image to right to left image present in a directory.

Transpose(): method was implemented in mirror function. This transpose method flips image and saves the image in another directory.

Eg: img.transpose(Image.FLIP_LEFT_RIGHT).save("C://Users//dell//Desktop//bccn//out//" + " rotate " + image).

The mirror images generated by flipping are also rotated by 90, 180 and 270 degrees and are added to the final dataset. This is relevant because the tumors in the flipped images vary from the original images and the rotations vary as well.

4.7 Design Model 2

4.7.1 Accessing Dataset

Preprocessed Dataset: Initially we extracted the region of interest(ROI) of all the benign and malignant mammograms and we created a new dataset of the cropped Mammogram and we use these regions of interests for further classification.

Accessing Data-set: Initially we get the current directory address from getcwd() function and then using the walk() function, we get all the mammogram image file names from the current directory. We then make a dataframe using info.txt file in the all-mias dataset, which contains all the corresponding data of the mammograms. We modify the data frame by removing unwanted columns and we reset indexes.

Label Encoding: We create a list for labeling the mammogram images, If the severity is benign we encode it as 1 and if the severity is malignant we encode it as 0.

4.7.2 Data Augmentation:

We augment the region of interest of the mammograms so as to increase the size of the dataset to make a good CNN model. For every image of the dataset, initially we read the image using imread() function and then resize the image to a size of 224 x 224 pixels using resize() function. For each mammogram we rotate the mammogram 360 degrees(0 to 360) using the function getRotationMatrix2D(). getRotationMatrix2D() function is in the cv2 package and it takes coordinates of the center of the image, and the particular angle to be rotated as its attributes. All the 360 images generated by the initial image will be labeled the same as the initial image. Now the final dataset size is initial dataset size times 360.

Splitting dataset for training and testing:we split the dataset for training and testing using the train_test_split() function of sklearn.model_selection, of the dataset 80 percent is used to train the model and remaining 20 percent is used to test the model. Now we convert the splitted lists into np arrays using np.array() function. The final training set size after data augmentation is 35136 images. The final testing set size after data augmentation is 8784 images.

4.7.3 Building Model

Initially, we create a sequential model using sequential() function from keras. Then we add two convolution layers using the function Conv2D() from 'keras.layers' with 32 and 64 as batch sizes respectively and with kernel size 3 x 3 and activation function as "ReLU". Then we add a max pooling layer using the function MaxPool2D() from 'keras.layers' with pool size 2 x 2. Again we add one convolution layer with batch size 64 and with kernel size 3x3 and activation function as "ReLU" and we add max pooling with pool size 2 x 2. Now, we add a Dropout layer using the function Dropout() from 'keras.layers' with Dropout Rate as 0.5. Then we add a Dense layer with batch size of 64 using the function Dense() from 'keras.layers' with activation function as "ReLU".

The main importance of the Dropout layer is that it prevents CNN models from overfitting. This dropout technique selects some neurons and ignores them during training and those neurons are "dropped-out" randomly in order to avoid overfitting. This implies that activation for the downstream neurons is removed temporarily in the forward pass and there won't be any updation of weights for those neurons in the backward pass. Dropout technique is the best technique among all the regularization techniques for CNN models. Now we compile the model with an optimizer as Adam, with loss as Binary cross entropy and accuracy as a metric and then we train the model with the training set, validation_split = 0.2 and batch size of 64 and 10 epochs. Finally we evaluate the model using the testing set. After compilation, we store the model as a pickle file (pickle file name is ‘bcccnn_fimg.pkl’).

Chapter 5

Results

5.1 Proposed Model Results

In problem statement 1, we created a CNN model to classify benign, malignant or no cancer. This method does not involve any preprocessing as we cannot preprocess no cancer images. For this model we got an accuracy of 57% for the three way classification. In problem statement 3, we created a CNN model to classify benign and malignant images. In this model, we preprocessed and augmented the images and are sent to the model for training. After evaluating the model we got accuracy of 94% and with a loss value of 0.18. Successfully implemented the above built classification algorithms for recognition of breast cancer mammograms.
Table 5.1 contains Confusion Matrix after Evaluating the final model.

Actual predicted	Positive	Negative	Total
Positive	3617	248	3865
Negative	261	4658	4919
Total	3878	4906	8784

Table 5.1: Confusion Matrix

5.2 Performance Analysis

Model1: classification of Benign, Malignant and no cancer without preprocessing.

Model2: classification of Benign or Malignant with preprocessing.

Table 5.2 depicts the performance comparison of different Evaluation Parameters of two proposed model.

model	Accuracy(%)	Precision	Recall	F1
model1	57.0	0.2000	0.1111	0.1428
model2	94.2	0.9494	0.9469	0.9481

Table 5.2: Comparison in proposed models

5.3 Comparison with References

5.3.1 Comparison with learned research papers

Research paper 1:

The Pre-processing and Image segmentation techniques from the paper "Convolutional neural network improvement for breast cancer classification[1]", are used to extract Region of Interest from the mammograms.

Research paper 2:

The Data Augmentation techniques from the paper "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system" are used for the classification.

Resulted of all these Research Papers are noted in Table 5.3 and compared with proposed model.

model	Accuracy(%)	Precision	Recall	F1
Proposed model	94.2	0.9494	0.9469	0.9481
Research paper 1	90.5	92	-	-
Research paper 2	97	-	-	-

Table 5.3: Comparsion with Learned Research Papers

5.3.2 Comparison with other models

Different other models are adopted to verify the accuracies with the proposed model and accuracies are noted in Table 5.4.

Model/Method	Accuracy	Reference
Proposed model	94.2	-
KNN	97.3	[5]
RCNN Clasifier	91.3	[4]
CNN with Support Value	97.21	[4]

Table 5.4: Comparison with other models

Chapter 6

Conclusion and Future work

Our research project mainly focuses on Breast Cancer Classification on MIAS Data set. The research that we did has shown the process of Breast Cancer Classification can be made more accurate with proper Pre-processing and Data Augmentation techniques. Initially, a three way classification of breast cancer model is developed to classify benign, malignant or no cancer without data augmentation, then we built a classification model with Pre-processing and Data Augmentation techniques to classify begnin or malignant, We have observed a considerable change in the accuracies of the two models.

As part of performance analysis we compared our model with other models of similar techniques. We have referred the Pre-processing and Image Segmentation techniques from [1], and we referred Data Augmentation technique from [4]. We can say that, with theses techniques our model outperforms some of the other model results shown above.

We can extend this project to classify more severity levels of Cancer other than begnin and malignant, so that it helps Medical Experts to treat the disease in much accurate way for more each severity levels.

References

[1] K. S. S. Fung Fung Ting, Yen Jun Tan, "Convolutional neural network improvement for breast cancer classification," *Expert Systems With Applications*, vol. 120, pp. 103–115, 2018.

[2] M. G. Shweta Saxena, "Machine learning methods for computer-aided breast cancer diagnosis," *Journal of Medical Imaging and Radiation sciences*, vol. 51, pp. 182–193, 2020.

[3] S. G. Puja Gupta, "Breast cancer prediction using varying parameters of machine learning models," *Procedia Computer Sciences*, vol. 171, pp. 591–601, 2020.

[4] S. R. K. Anji Reddy Vakaa, Badal Sonia, "Breast cancer detection by leveraging machine learning," *ITC express*, 2020.

[5] Q. S. Y. L. a. S. L. J. W. L. L. a. H. Z. b. Pin Wanga, , "Cross-task extreme learning machine for breast cancer image classification with deep convolutional features," *Biomedical Signal Processing and Control*, vol. 57, 2020.

[6] y. . Y. C. H. Z. G. Y. J. y. YI WANG, 1 EUN JUNG CHOI and S.-B. KO, "Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning," *Ultrasound in Med*, vol. 46, pp. 1119–11132, 2020.

[7] I. M. Karabatak M, "An expert system for detection of breast cancer based on association rules and neural network," *Expert systems with Applications*, vol. 36, pp. 3465–3469, 2009.

[8] A. A. PayamZarbakhsh, "Breast cancer tumor type recognition using graph feature selection technique and radial basis function neural network with optimal structure," *Journal of Cancer Research and Therapeutics*, vol. 14, pp. 625–33, 2018.

[9] L. R. R.Thoma, "Histology image analysis for carcinoma detection and grading," *Computer Methods and Programs in Biomedicine*, vol. 107, pp. 538–556, 2012.

[10] D. P. Dragana Djilas, Sasa Vujnovic, "Breast mri,digital mammogrphy and breast tomosynthesis: comparison of three methods for early detection of breast cancer," *Computer Methods for Health Sector*, 2016.

[11] C. P. Fabio A. Spanhol, Luis S. Oliveria, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, pp. 1415–1462, 2016.

[12] K. G. Kim, "Deep learning," *The MIT Press*, vol. 22, pp. 351–354, 2016.

[13] K. K. J. H. Xialong Sun, Juyoung Park, "Novel hybrid cnn-svm model for recognition of functional resonance images," *IEEE International on SMC*, 2017.

[14] K. L. Mingxing Duan, Kenli Li, "An ensemble cnn2elm for age estimation," *IEEE Transactions on Information Forensics and Security*, vol. 13, pp. 758–772, 2017.