**Breast Cancer Classification using Deep Learning**

A MINI PROJECT REPORT

***Submitted by***

|  |  |
| --- | --- |
| **NIDIN** | **1905034** |
| **SANJAY PRATAP T K** | **1905046** |
| **YOKESH R S** | **1905062** |
| **ABUTHAGEER** | **2005201** |

***In partial fulfilment for the award of the degree of***

**BACHELOR OF ENGINEERING**

***in***

## COMPUTER SCIENCE AND ENGINEERING



**COIMBATORE INSTITUTE OF TECHNOLOGY**

***(Government Aided Autonomous Institution Affiliated to Anna University)***

## COIMBATORE-641 014

### ANNA UNIVERSITY - CHENNAI 600 025

**MAY 2022**

## BONAFIDE CERTIFICATE

Certified that this mini project **“Breast Cancer Classification using Deep Learning”** is the bonafide work of **NIDIN S (1905034), SANJAY PRATAP T K (1905046), YOKESH R S (1905062), and ABUTHAGEER S (2005201)** under my supervision during the academic year 2021-2022.

### Dr.G.Kousalya, M.E.,Ph.D Dr.S.P.Abirami, Ph.D

#### HEAD OF THE DEPARTMENT, ASSISTANT PROFESSOR,

Department of CSE Department of CSE,

Coimbatore Institute of Technology, Coimbatore Institute of Technology, Coimbatore – 641 014. Coimbatore – 641 014.

Certified that the candidates were examined by us in the project work viva voce examination held on …………………

### Internal Examiner External Examiner

Place: Date:

### TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ACKNOWLEDGEMENT** | **5** |
|  | **ABSTRACT** | **6** |
| **1)** | **INTRODUCTION** |  |
|  | **1.1. INTRODUCTION** | **7** |
|  | **1.2. GOOGLE TREND ANALYSIS** | **8** |
| **2)** | **LITERATURE SURVEY** | **9** |
| **3)** | **METHODOLOGY** |  |
|  | **3.1 INTRODUCTION** | **14** |
|  | **3.2 METHODS USED** | **14** |
|  | **3.3 STEPS OF IMPLEMENTATION** | **16** |
|  | **3.4 SYSTEM SPECIFICATIONS** | **17** |
|  | **3.5 SUMMARY** | **17** |
| **4)** | **SYSTEM ANALYSIS** |  |
|  | **4.1 INTRODUCTION** | **18** |
|  | **4.2 EXISTING SYSTEM** | **18** |
|  | **4.3 PROPOSED SYSTEM** | **18** |
|  | **4.4 SUMMARY** | **18** |
| **5)** | **SYSTEM DESIGN** |  |
|  | **5.1 INTRODUCTION** | **19** |
|  | **5.2 MODEL AND SYSTEM ARCHITECTURE** | **19** |
|  | **5.3 SUMMARY** | **20** |
| **6)** | **IMPLEMENTATION AND RESULT** |  |
|  | **6.1 ANN MODEL** | **21** |
|  | **6.2 EFFICIENTNET B7 MODEL** | **24** |
| **7)** | **CONCLUSION** |  |
|  | **7.1 INTRODUCTION** | **31** |
|  | **7.2 CONCLUSION** | **31** |
|  | **7.3 APPILICATION** | **31** |
| **8)** | **REFERENCES** | **32** |

**ACKNOWLEDGEMENT**

### -----------------------------------------------------------------------------------------------------

**ACKNOWLEDGEMENT**

We express our sincere thanks to our Secretary **Dr.R.Prabhakar** and our Principal **Dr. Rajeswari A** for providing us a great opportunity to carry out our work. The following words are a small part to express our gratitude to them. This work is the outcome of their inspiration and product of plethora of their knowledge and rich experience.

We record the deep sense of gratefulness to **Dr.G.Kousalya**, Head of the Department of Computer Science and Engineering, for his encouragement during this tenure.

We equally tender our sincere thankfulness to our project guide **Dr.S.P.Abirami**, Assistant Professor, Department of Computer Science and Engineering, for her valuable suggestions and guidance during this course.

During the entire period of study, the entire staff members of the Department of Computer Science and Engineering have offered ungrudging help. It is also a great pleasure to acknowledge the unfailing help we have received from our friends.

It is matter of great pleasure to thank our parents and family members for their constant support and co-operation in the pursuit of this endeavor.

# ABSTRACT

**-----------------------------------------------------------------------------------------------------**

**ABSTRACT**

Globally, breast cancer is the most common cancer among women, and the most likely cause of female cancer deaths. High-income countries (HICs) have made the most progress in improving breast cancer outcomes. Between 1990 and 2014, breast cancer death rates dropped by 34% in the US attributable to the combination of improved earlier detection and effective adjuvant therapies. By contrast, breast cancer is an increasingly urgent problem in low- and middle-income countries (LMICs), where historically low incidence rates have been rising by up to 5% per year. The strides seen in HICs have not been mirrored in LMICs where optimal management strategies from wealthy countries cannot be fully implemented due to significant resource constraints related to limited personal resources, underdeveloped health care infrastructure, lack of pharmaceuticals and cultural barriers.

“Breast Cancer Detection” is a simple and effective project that uses efficient machine learning algorithms for early detection and diagnosis of breast cancer by classifying either malignant or benign patients

The model primarily uses Wisconsin Dataset for training the model and classifies using Artificial Neural Network technique.

## INTRODUCTION

**CHAPTER 1**

**INTRODUCTION**

### Introduction:

Breast Cancer is the most affected disease between women all over the world. All most 25% of all cancers with an estimated 1.67 million new cancer cases diagnosed in 2012 and its incidence is increasing day by days. It is also said that Breast cancer is the second leading cause of death for women worldwide.

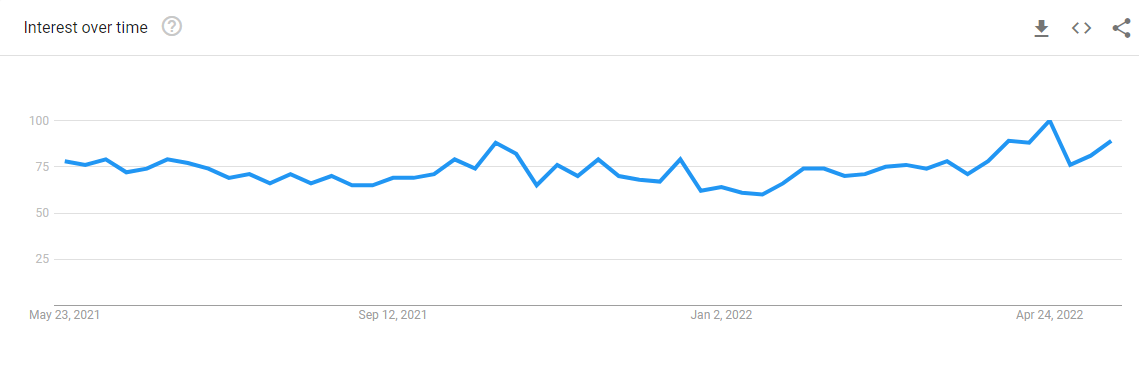
Early detection of cancer can reduce the risk of deaths for cancer patients**.** To increase the survival rate the early diagnosis of breast cancer and a trustworthy detection model is required.

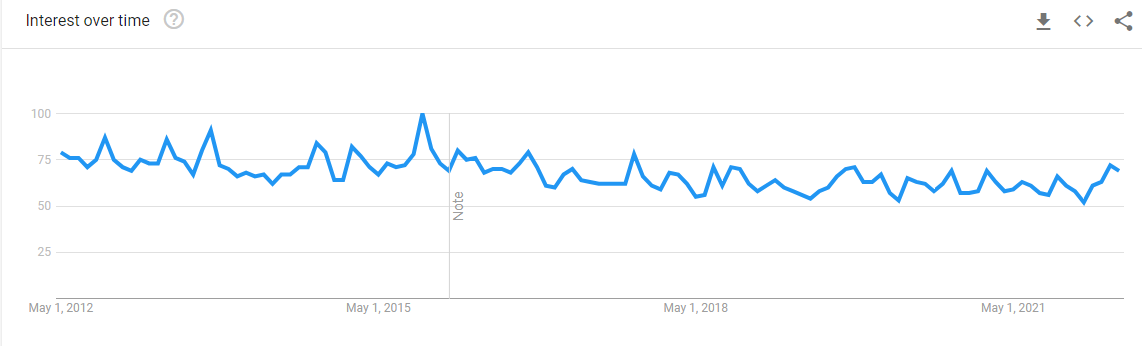
The goal of the research is to identify and classify Malignant and Benign patients by proposing a simple and efficient model for early detection of breast cancer with minimal error percentage.

This model uses an Artificial neural network and produces Accuracy of 98% and f1 score as 98% - Benign and 97% - Malignant .

### Google trend analysis:

The diagram shown below displays the google trend analysis about the domain of this project.





.

## LITERATURE SURVEY

**----------------------------------------------------------------------------------**

## CHAPTER 2 LITERATURE SURVEY

# 2.1 Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm (2016)

**Authors:** Moh’d Rasoul Al-hadidi,Abdulsalam Alarabeyyat,Mohannad Alhanahnah.

#### Description

In this paper, we proposed a new method to detect the breast cancer with high accuracy.This method consists of two main parts, in the first part the image processing techniques are used to prepare the mammography images for feature and pattern extraction process. The extracted features are utilized as an input for a two types of supervised learning models, which are Back Propagation Neural Network (BPNN) model and the Logistic Regression (LR) model .The LR model resulted in Accuracy of 45% with 750 features and BPNN with 93.7 % which is comparatively extremely good but with only 240 features.

#### Advantages

* + 1. They have used image processing to obtain the features for training the prediction model and used have BPNN and Logistic regression.

**Disadvantages**

1 .The accuracy rate of the model greatly affects with the increase in the features and the maximum accuracy they could bring out was 93.7% only.

# 2.2 Efficient Approaches for Accuracy Improvement of Breast Cancer Classification Using Wisconsin Dataset (2017)

**Authors:** Shajib Ghosh ,Jubaer Hossain, Dr.Shaikh Anowarul Fattah , Dr. Celia Shahnaz, Asir Intisar Khan

#### Description:

This paper deals with different statistical and deep learning analysis of Wisconsin Breast Cancer Database for improving the accuracy in detection and classification of breast cancer based on different attributes. Applying Naïve Bayes, SVM, Logistic Regression, KNN, Random Forest, MLP and CNN classifiers, higher accuracy is obtained which is up to 98% to 99%.

#### Advantages:

1. The proposed goes through a lot of classification models and compares the accuracy for each.

#### Disadvantages

1. Harder to optimize the network as convergence time significantly increases.

# 2.3 Breast Cancer Diagnosis Using deep learning Algorithm (2018)

**Authors:** Naresh Khuriwal , Dr. Nidhi Mishra

**Description:**

In this paper they’ve proposed the deep learning method convolutional neural network that mostly used for classification of images dataset. It is basically divided into three parts first collection of dataset and applied pre-processing algorithm for scaled and filter data then split dataset for training and testing purpose and generate some graph for visualization data. They’ve concluded deep learning technology is a good way for diagnosis breast cancer with Wisconsin Breast Dataset. This database provides 569 rows and 30 features in the dataset. In this paper they’ve **just used 11 features** for diagnosis .After the implementation they’ve achieved 98.67% accuracy.

#### Advantages:

1. Uses Convolutional Neural network faster and better Accuracy

**Disadvantages:**

* + - 1. Uses only 11 features of given 30 features.

**2.4 Breast Cancer Detection Based on Deep Learning Technique (2019)**

**Author:** Nur Syahmi Ismail and Cheab Sovuthy.

**Description:**

In this paper, they’ve used deep learning technique using VGG16 and ResNet50 network to classify between normal tumour and abnormal tumour using IRMA dataset and have implemented for normal and abnormal breast cancer detection. The classification methods were evaluated using three performance evaluations which are precision, recall, and accuracy rate. The best result of classification accuracy was VGG16 with 94%. Compared to ResNet50 with 91.7% in term of accuracy.

**Disadvantage:**

1. Usage of already available Model
2. Poor accuracy.

# Breast Cancer Prediction Using Machine Learning (2020)

# Author: Ramik Rawal.

# Description:

# In this paper they have used four algorithms SVM, Logistic Regression, Random Forest, and KNN to predict the outcome the breast cancer. It broadly speaks about three domains. First domain is prediction of cancer before diagnosis, second domain is prediction of diagnosis and treatment and third domain focuses on outcome during treatment. The accuracy obtained by SVM (97.13%) is better than the accuracy obtained Naïve Bayes and k-NN that have an accuracy that varies between 95.12 % and 95.28 %. It was also clear with SVM having highest value of correctly classified instances and the lower value of incorrectly classified instances than the other classifiers.

# Advantage:

# Better preprocessing Technique

# Average accuracy of different models was high.

# Disadvantage:

# Efficiency of benign class and malignant class got the highest for two different algorithms such as SVM for benign class and K-NN for malignant class

**METHODOLOGY**

**CHAPTER 3**

**METHODOLOGY**

**3.1 INTRODUCTION**

This chapter explains the entire methodology that has been used in the system in detail. The methods used are Artificial Neural Network and Convolutional Neural Networks. It also covers the hardware and software specifications requirements that have been used to work on this system.

**3.2 METHODS USED**

**3.2.1 Convolutional Neural Network:**

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

Convolutional layers are the major building blocks used in convolutional neural networks.

A convolution is the simple application of a filter to an input that results in activation. Repeated application of the same filter to an input result in a map of activations called a feature map, indicating the locations and strength of a detected feature in an input, such as an image.

The innovation of convolutional neural networks is the ability to automatically learn a large number of filters in parallel specific to a training dataset under the constraints of a specific predictive modeling problem, such as image classification. The result is highly specific features that can be detected anywhere on input images.

**3.2.2 Artificial Neural Network:**

The term "Artificial Neural Network" is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another; artificial neural networks also have neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes. An artificial neural network has three or more layers that are interconnected. The first layer consists of input neurons. Those neurons send data on to the deeper layers, which in turn will send the final output data to the last output layer. All the inner layers are hidden and are formed by units which adaptively change the information received from layer to layer through a series of transformations. Each layer acts both as an input and output layer that allows the ANN to understand more complex objects. Collectively, these inner layers are called the neural layer. The units in the neural layer try to learn about the information gathered by weighing it according to the ANN’s internal system. These guidelines allow units to generate a transformed result, which is then provided as an output to the next layer.

An additional set of learning rules makes use of backpropagation, a process through which the ANN can adjust its output results by taking errors into account. Each weight is updated proportionally to how much they were responsible for the error. Hence, the error is used to recalibrate the weight of the ANN’s unit connections to take into account the difference between the desired outcome and the actual one. In due time, the ANN will “learn” how to minimize the chance for errors and unwanted results.

**3.3 STEPS OF IMPLEMENTATION:**

**1. Dividing the dataset:**

The dataset that is used in the ANN model is WISCONSIN dataset and the histopathological images of malignant, benign and normal classes are given as input to the EfficientNet B7 model for classification.

The Wisconsin dataset is first divided into 70 % for training the ANN model and the remaining 30% for testing or validating the model. The WISCONSIN dataset consists of 33 features and 569 rows of data.

**2. Preprocessing the dataset:**

The two unwanted features of the WISCONSIN dataset namely the ‘id’,’Unnamed’ columns are removed from the dataset. The diagnosis class feature is encoded with 0 for Benign and 1 for Malignant. MinMaxScaler() is imported from sklearn library and the data is scaled in the particular common range of values.

And for the images the mask images removed from the dataset ,which are unnecessary and the total image size reduces to 780.Whenever the image is passed into the trained model we divide the pixel of each image by 224x224 to reduce the image size , indeed reducing the image size.

**3. Build and train the model:**

The CNN network consists a set of customized layers such as conv2D, max-pooling layer, activation layer, dropout layer, dense layer. The augmented images were fed to the network which in turn returns the feature learnt from the image to the output layer which detects the diseases present in the particular leaf.

**4. Output:**

The ANN which is trained using the WISCONSIN dataset gives the accuracy and the predicts for any input value of the same features that is used to train the model. And the EfficientNet B7 model which is trained with the image dataset gives the prediction either of three classes Benign or Malignant or Normal for any histopathological image given as input.

**3.4 SYSTEM SPECIFICATIONS:**

**Hardware specifications:**

● Operating System: Windows 10

● Processor : Ryzen 3/ Intel i5

● Hard disk : 500 GB

● RAM : 8GB (minimum)

**Software specifications:**

● IDE : Google Colab, Jupyter Notebook

● Coding Language : Python

**3.5 SUMMARY:**

All the methods used are explained in detail in this chapter. The tools used in the current implementation are also given above. It also provides the steps of implementations with a brief description of the process under each step.

**SYSTEM ANALYSIS**

**CHAPTER 4**

**SYSTEM ANALYSIS**

**4.1 INTRODUCTION**

This chapter explains the existing system and the proposed system in detail.

**4.2 EXISTING SYSTEM**

The existing system for Breast cancer Classification include just only the predictor using the simple learning algorithms and have ended up with good accuracy, on the other hand with few input as training images and with is insufficient to really prove a given data benign or malignant and less accuracy.

**4.3 PROPOSED SYSTEM**

Our model uses EfficientNet b7 as a base model for training and testing the data, which are most advanced model developed with CNN and really a good model for classification using image datasets. The model uses 780 image datasets including benign, malignant and normal for training the model gives best result for further classification on input of any histopathological image.

**4.4 SUMMARY:**

This chapter gives an overview of the existing system and the proposed system. The proposed system uses powerful Convolutional Neural Networks of EfficientNet B7 to accurately predict for the given image sample as benign or malignant or normal sample.

**SYSTEM DESIGN**

**CHAPTER 5**

**SYSTEM DESIGN**

**5.1. INTRODUCTION:**

This chapter discusses the flow of this system.

**5.2. MODEL AND SYSTEM ARCHITECTURE:**

1. **ANN Model :**

Read Wisconsin dataset

Data Pre-processing

Data Visualization

Splitting Data for training and Testing

Training ANN model

Testing Model

1. **EfficientNet B7:**

Pre-Processing the image

* **Create EfficientNEt B7 model**
* **Don’t include top layers**
* **Freeze all the weights**
* **Append own layers for Transfer Learning**
* **Compile model**

Prepare target

Building the Model

Training model

Evaluate Model

Train the model again without freezing the weights

Evaluate Model

Prediction

Normal

Malignant

Benign

Not harmful

Very virulent or infectious

**5.3. SUMMARY:**

This chapter provides the overall system design and model architecture which facilitates better understanding.

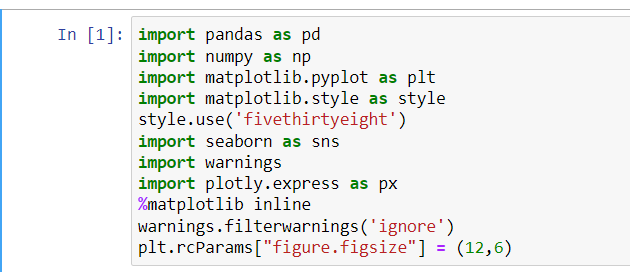
**IMPLEMENTATION AND RESULT**

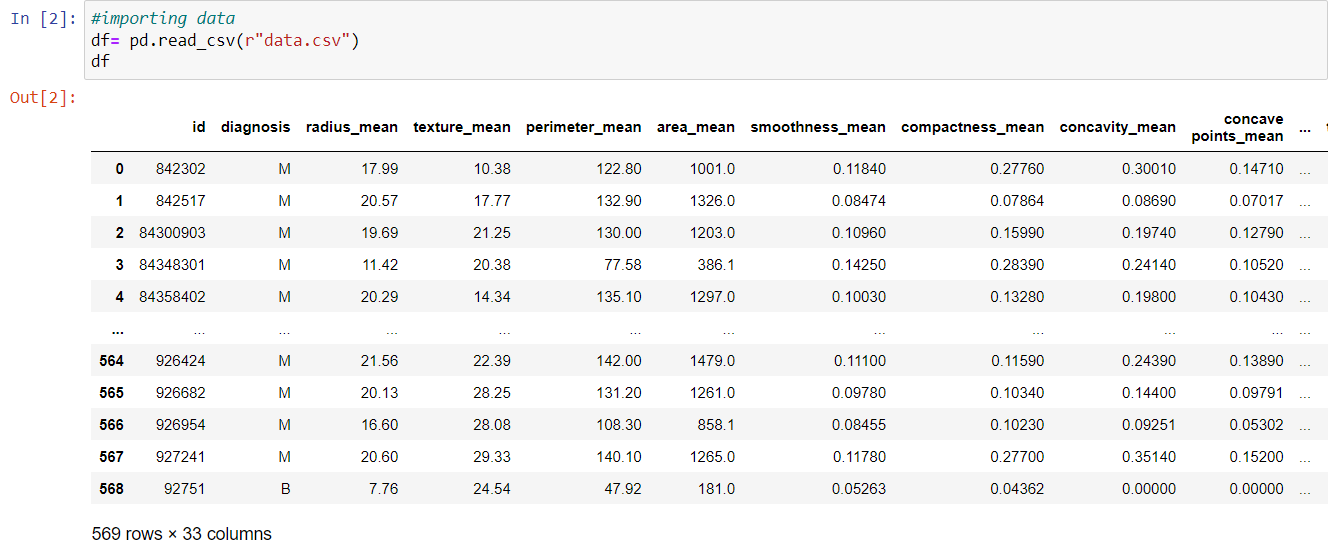
**CHAPTER 6**

**IMPLEMENTATION & RESULT**

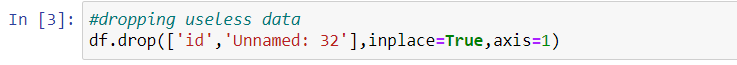
**6.1) ANN MODEL:**

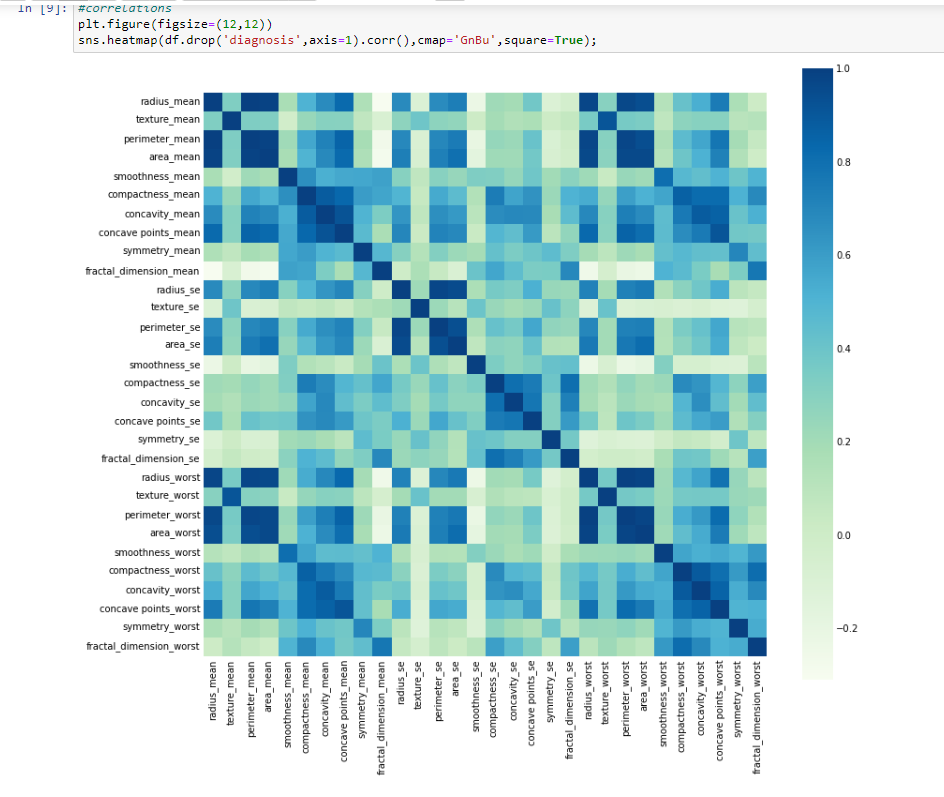
**Fig.6.1.1 Importing libraries and dataset:**

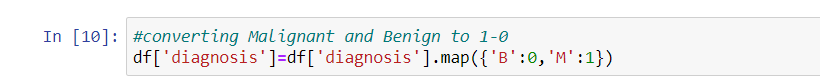




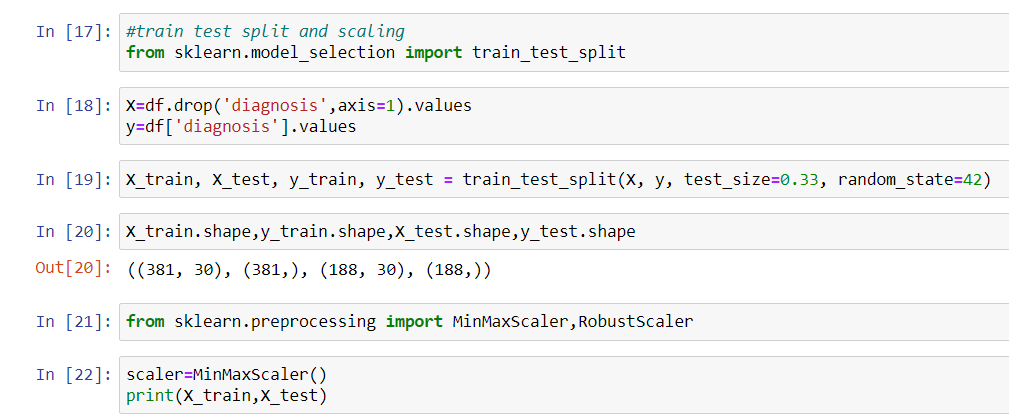
**Fig.6.1.2 Data Pre-processing and Data Visualization:**



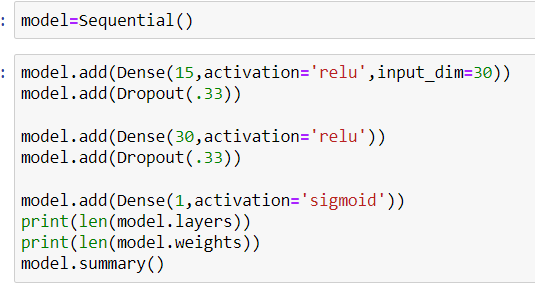


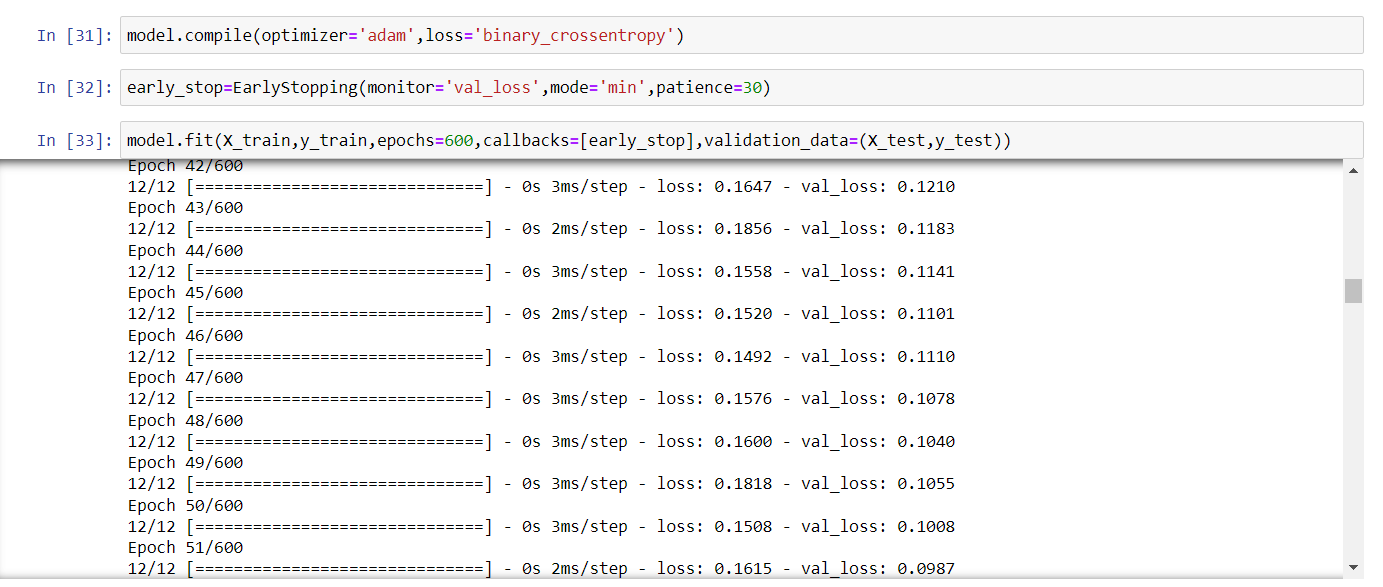


**Fig.6.1.3 Splitting dataset for train and test:**

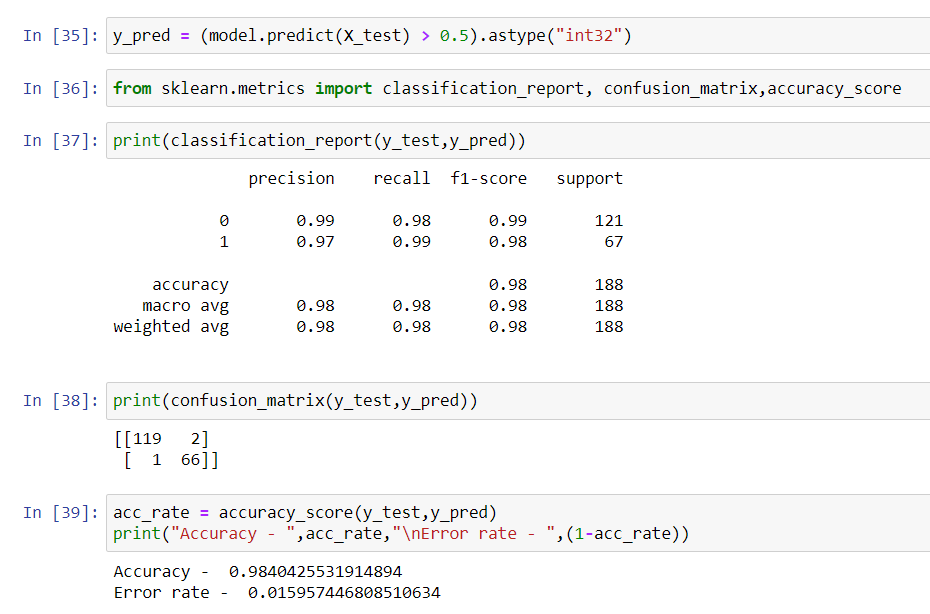


**Fig.6.1.4 Train the model:**



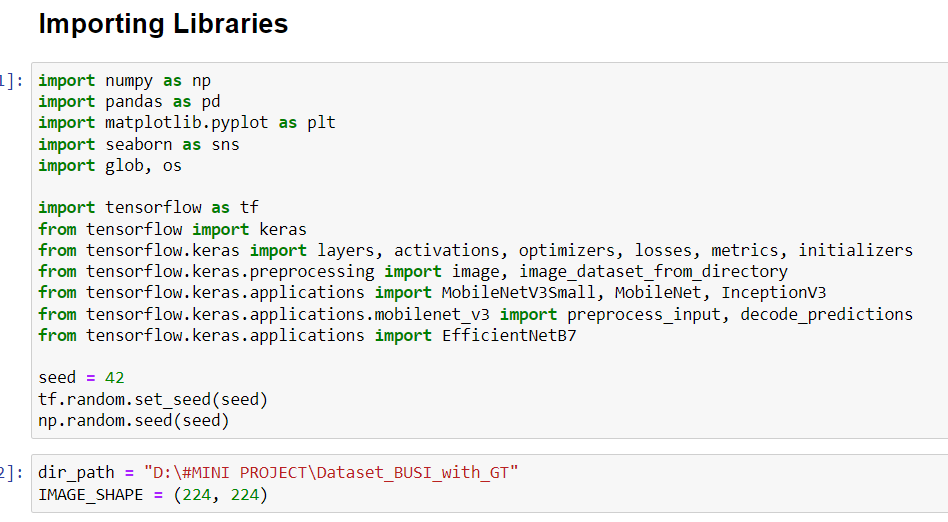


**Fig.6.1.5 Testing the Model and producing the confusion matrix:**

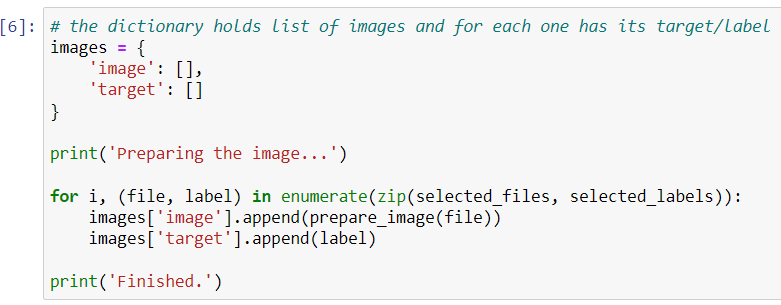
****

**6.2)EfficientNet B7 model :**

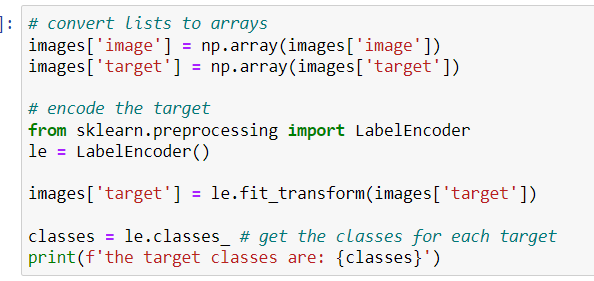
**Fig.6.2.1 Importing libraries and dataset:**

****

**Fig.6.2.2 Image Pre-processing:**

****

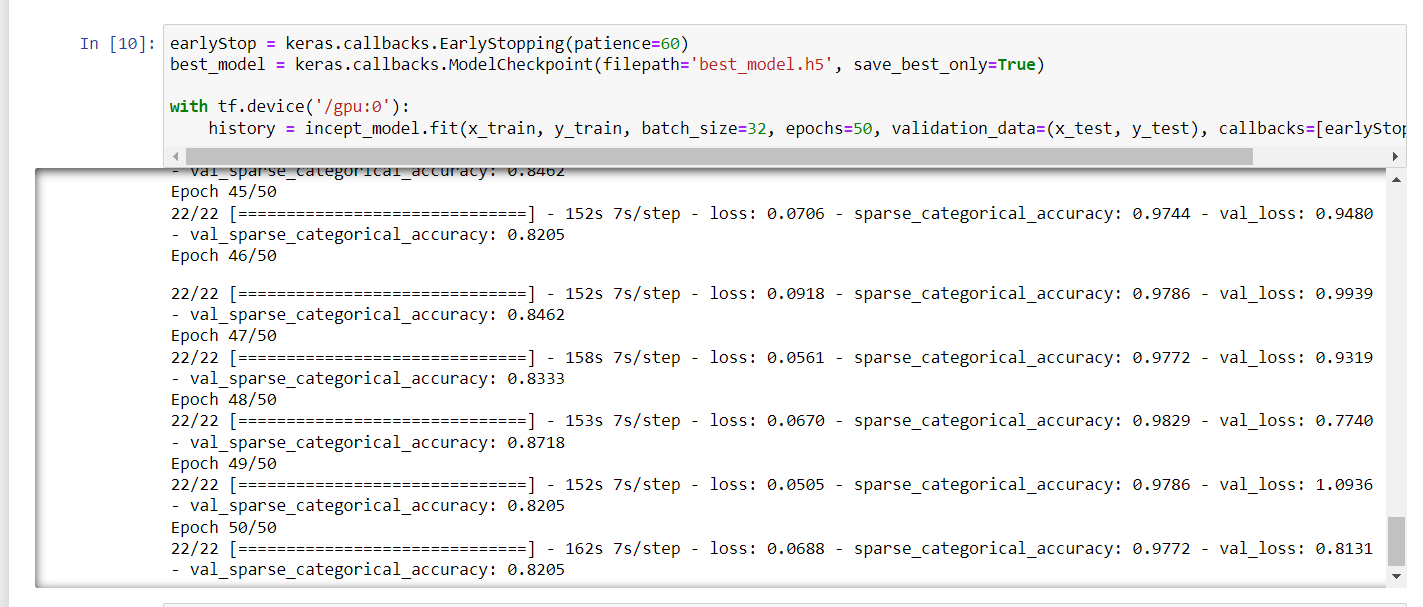
**Fig.6.2.3 Identifying images and encoding class names:**

****

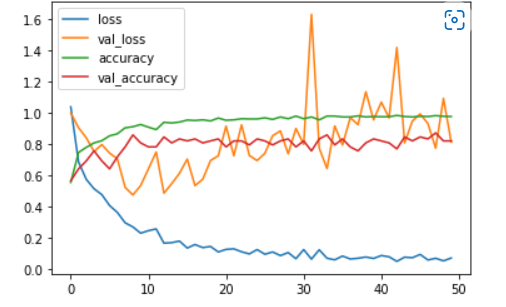
**Fig.6.2.3 Build the model :**

****

**Fig.6.2.4 Train the model :**

****

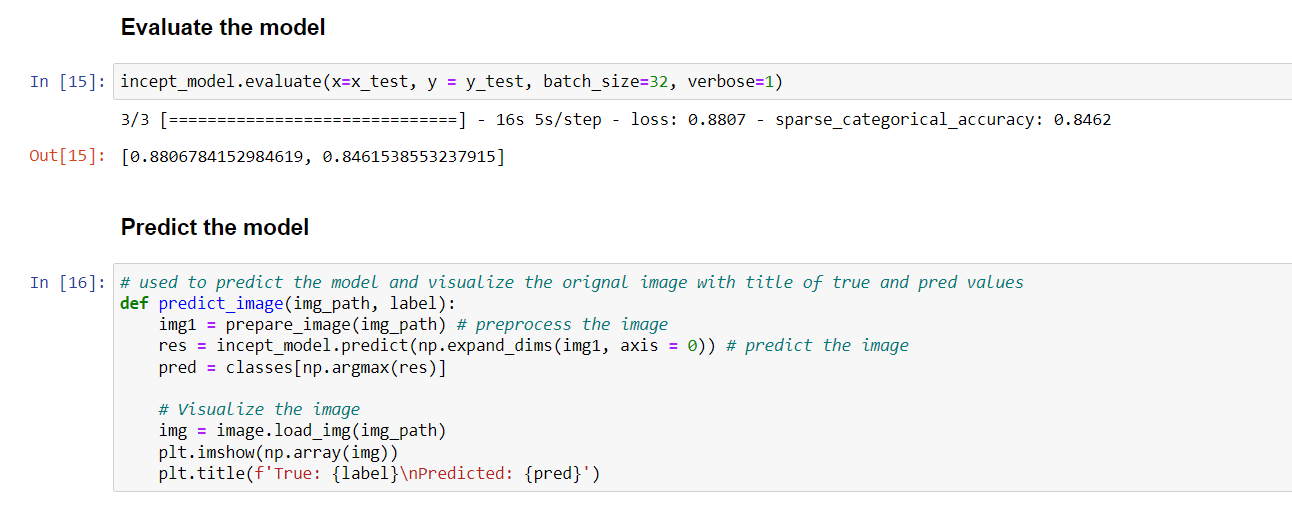
**Fig.6.2.4 plot for Loss, value loss, accuracy and value accuracy.**

****

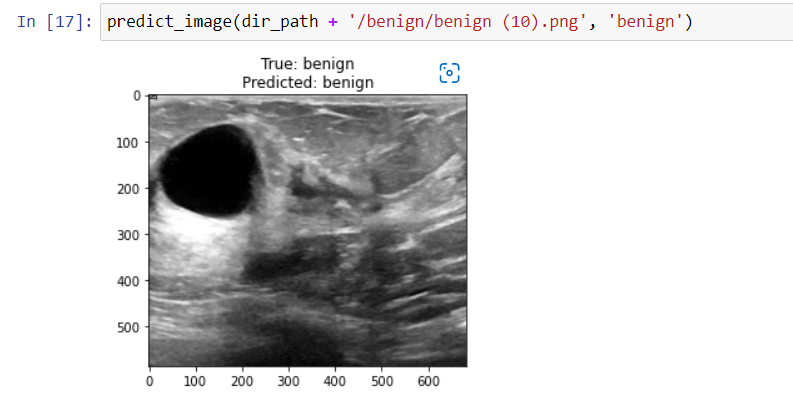
**Fig.6.2.5 Training the Model again (Transfer Learning).**

****

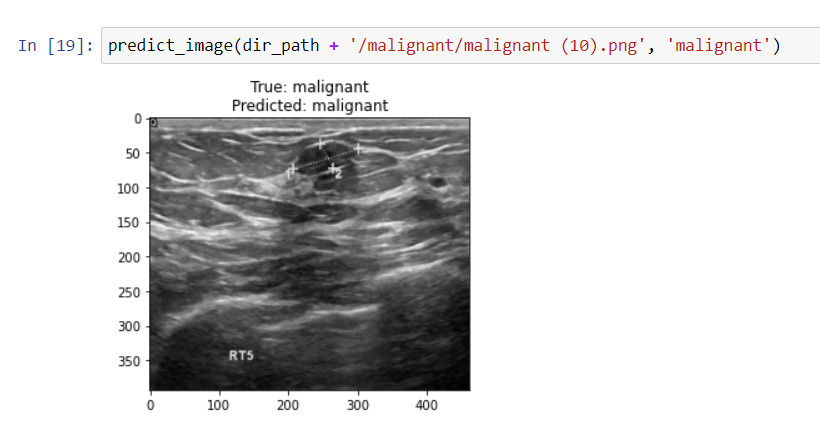
**Fig.6.2.6 Evaluate and predict the model.**

****

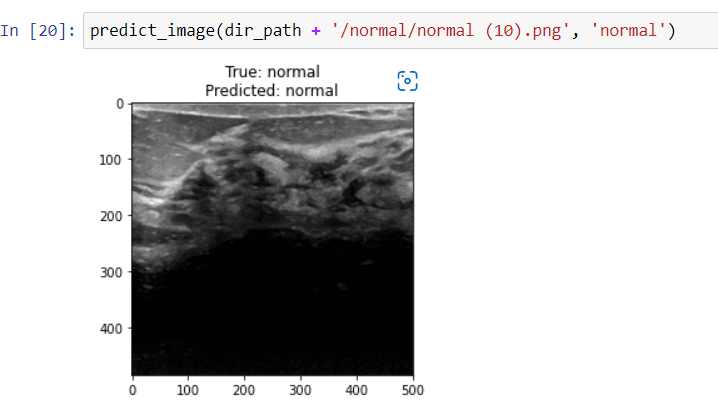
**Predicted samples.**

****

**Fig.6.2.7**

****

**Fig.6.2.8**

****

**Fig.6.2.9**

**CONCLUSION**

**CHAPTER 7**

**CONCLUSION**

**7.1 INTRODUCTION:**

This section gives the conclusion of the system, its applications in the real world.

**7.2 CONCLUSION:**

The project uses a deep learning technique convolutional neural system in that, a latest model of EfficientNet B7 which is trained of using the image datasets and another model of Artificial neural Network that is trained using the Wisconsin dataset. Both the models can be used for classification of breast cancer. The given datasets is properly pre-processed and used for training and testing the model.

**7.3 APPLICATION:**

The trained model can be used in the hospitals for diagnosing the patients whether they are affected with breast cancer or not. If affected what is the stage of the tumour either benign or Malignant, which will be helpful for the early diagnosis and could save lives of the people.

**REFERENCES**

**CHAPTER 8**

**REFERENCES**

[1] Moh’d Rasoul Al-hadidi, Abdulsalam Alarabeyyat, Mohannad Alhanahnah Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm 2016 9th International Conference on Developments in eSystems Engineering.

[2] Shajib Ghosh, Jubaer Hossain , Dr. Shaikh Anowarul Fattah, Dr. Celia Shahnaz , Asir Intisar Khan Efficient Approaches for Accuracy Improvement of Breast Cancer Classification Using Wisconsin Database 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh.

[3] Naresh Khuriwal , Dr Nidhi Mishra **Breast Cancer Diagnosis Using Deep Learning Algorithm** International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018).

[4] Nur Syahmi Ismail , Cheab Sovuthy Breast Cancer Detection Based on Deep Learning Technique *Universiti Teknologi PETRONAS* Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia.

[5] Hari Krishna Timmana, Rajabhushanam C Breast Malignant Detection using Deep Learning Model Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020) IEEE Xplore Part Number: CFP20V90-ART; ISBN: 978-1-7281-5461-9.

[6] Ramik Rawal , BREAST CANCER PREDICTION USING MACHINE LEARNING 2020 JETIR May 2020, Volume 7, Issue 5.