DEEP LEARNING TECHNIQUES FOR BREAST CANCER RISK PREDICTION USING IBM CLOUD

A Mini Project report submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

POSHALA SAI DEEPAK **18UK1A05L6**

Under the guidance of

Mr. A. ASHOK KUMAR

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) – 506005

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

WARANGAL

CERTIFICATE

This is to certify that the Mini project report entitled **" DEEP LEARNING TECHNIQUES FOR BREAST CANCER RISK PREDICTION"** is being submitted by **P. SAI DEEPAK (18UK1A05L6)** in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2021- 2022.

Project Guide **HOD**

Mr. A. ASHOK KUMAR **Dr. R. NAVEEN KUMAR**

EXTERNAL

ACKNOWLEDGEMENT

We wish to Jake this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. Prasad Rao**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this mini project in the institute.

We extend our heartfelt thanks to **Dr. R. Naveen Kumar**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the mini project.

We express heartfelt thanks to the Mini Project Coordinator, **Dr. J. Srikanth**, Assistant Professor, Department of CSE for his constant support and giving necessary guidance for completion of this mini project.

We express heartfelt thanks to the guide, **Mr. A. Ashok Kumar**, Assistant Professor, Department of CSE for his constant support and giving necessary guidance for completion of this mini project.

Finally, we express our sincere thanks and gratitude to our family members, friends for their encouragement and outpouring their knowledge and experiencing throughout thesis.

ABSTRACT

Nowadays, Cancer is a disease in which some of the body's cells grow uncontrollably and spread to other parts of the body. Cancer can start almost anywhere in the human body, which is made up of trillions of cells. These cells may form tumors, which are lumps of tissue. Tumors can be "Cancerous" or "Non- Cancerous". Cancerous tumors spread into, nearby tissues and can travel to distant places in the body to form new tumors (a process called metastasis). Cancerous tumors may also call **"Malignant Tumors"**. Non-Cancerous tumors do not spread into nearby tissues. When removed, benign tumors usually don't grow back, whereas cancerous tumors sometimes do. "**Benign Tumors"** are non-cancerous tumors. Breast cancer is one such type of cancer that starts in the breast. Cancer starts when cells begin to grow out of control .Breast cancer cells usually form a tumor that can often be seen on an X-ray or felt as a lump. Breast cancer occurs almost entirely in women, but men can get breast cancer, too. Non-cancerous breast cancer is abnormal growths, but they do not spread outside of the breast. They are not life threatening, but some type of benign breast lumps can increase a woman's risk of getting breast cancer. Any breast lump or change need to be checked by a health care professional to determine if it is benign or malignant (Cancer) and if it might affect your future cancer risk. Here we use **Histopathological Images (**it is the study of changes in tissues caused by disease) to predict the type of breast cancer by training the images with Convolutional Neural Network.

Keywords – Cancerous, Non-Cancerous, Tumors, Benign, Malignant, Lump, Histopathological Images, Convolutional Neural Network.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF CHAPTERS

1. INTRODUCTION.....1

MOTIVATION.....1

DEFINITION.....1

OBJECTIVE OF PROJECT.....2

PURPOSE.....3

2. PROBLEM STATEMENT.....3

3. LITERATURE SURVEY.....4

EXISTING SYSTEM.....4

PROPOSED SOLUTION.....4

4. EXPERIMENTAL ANALYSIS.....6

PROJECT ARCHITECTURE.....6

BLOCK DIAGRAM.....6

SOFTWARE REQUIREMENTS.....7

`PROJECT FLOW.....8

5. DESIGN.....10

CLASS DIAGRAM.....10

USE CASE DIAGRAM.....11

SEQUENCE DIAGRAM 11

FLOWCHART.....12

6. CODE SNIPPETS.....13

MODEL.CODE.....13

HTML CODE 16

APP1.PY CODE.....19

7. CONCLUSION 21

8. APPLICATION.....23

9. ADVANTAGES.....23

10. DIS-ADVANTAGES.....23

11. FUTURE SCOPE.....23

12. REFERENCES.....24

LIST OF FIGURES PAGE NO:

Figure 1: Classification of Histopathological Images 1
Figure 2: Training and Testing using Histopathological images with CNN 5
Figure 3: Project Architecture 6
Figure 4: Block diagram representing the process of CNN 6
Figure 5: Project flow 8
Figure 6: Class diagram 10
Figure 7: Use Case Diagram 11
Figure 8: Sequence Diagram 11
Figure 9: Flow chart 12
Figure 10: .ipynb code describing importing libraries and image preprocessing 13
Figure 11: .ipynb code describing the model initialization and adding convolutional layer 13
Figure 12: Adding layers to the sequential model 14
Figure 13: Adding Hidden and Output layers and training the model 14
Figure 14: Epochs running with 100% accuracy 15
Figure 15: Saving the model 15
Figure 16: Predicting the model 16
Figure 17: HTML Code of index.html describing the styles for webpage 16
Figure 18: Styling the webpage in the body tag 17
Figure 19: Describing the content of the webpage 17
Figure 20: Styling the predict.html 18
Figure 21: Styling the webpage in body tag 18
Figure 22: Content of prediction webpage 19
Figure 23: Python Code 19
Figure 24: Code for prediction in python file 20
Figure 25: Index Webpage 21
Figure 26: Predicting Breast Cancer of type Malignant 21
Figure 27: Index Webpage 22
Figure 28: Predicting Breast Cancer of type Benign 22

1. INTRODUCTION

MOTIVATION

Breast Cancer is a type of cancer that starts in the breast. Breast cancer occurs almost entirely in women, but men can get breast cancer, too. Breast cancer cells usually form a tumor that can often be seen on an x-ray or felt as a lump. Non-Cancerous breast tumors are abnormal growths, but they do not spread outside of the breast. They are not life threating, but some types of benign breast lumps can increase a woman's risk of getting breast cancer. Any breast lump or change needs to be checked by a health care professional to determine if it is benign or malignant(cancer) and if it might affect your future cancer risk. For this type of future risk prediction we use this kind of Breast Cancer Prediction using Histopathological Images with Deep Learning Techniques.

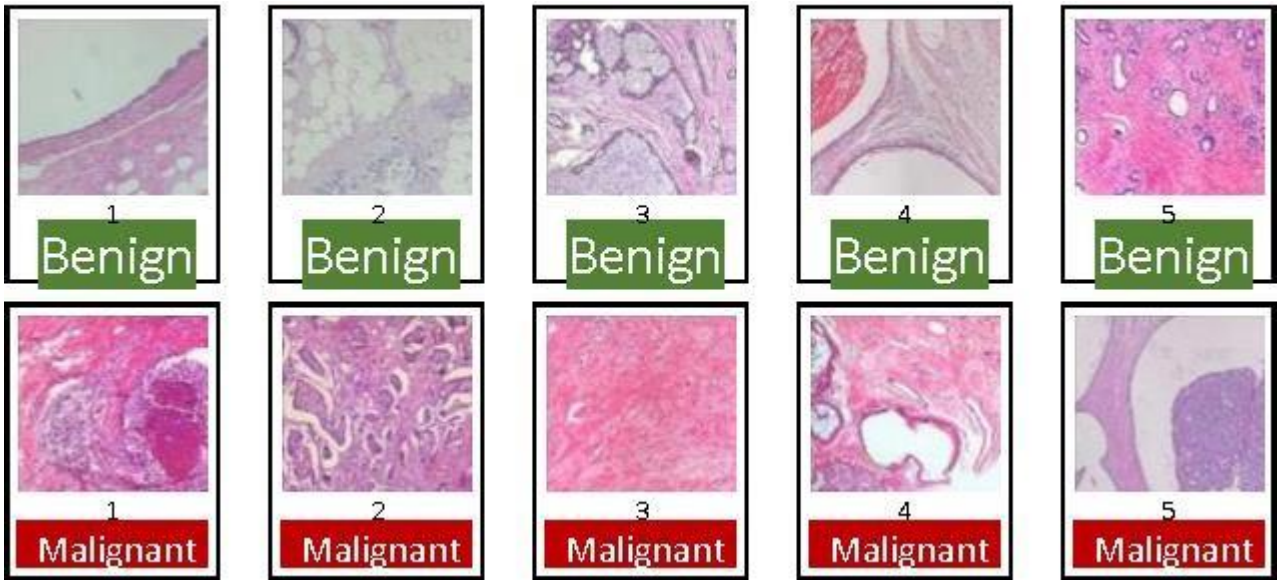

Figure 1: Classification of Histopathological Images

DEFINITION

Breast cancer is cancer that develops from breast tissue. Signs of breast cancer may include a lump in the breast, a change in breast shape, dimpling of the skin, fluid coming from the nipple, a newly inverted nipple, or a red or scaly patch of skin. Risk factors for developing breast cancer include being female, obesity, a lack of physical exercise, alcoholism, hormone replacement therapy during menopause, ionizing radiation, an early age at first menstruation, having children late in life or

not at all, older age, having a prior history of breast cancer, and a family history of breast cancer. About 5–10% of cases are the result of a genetic predisposition inherited from a person's parents, including BRCA1 and BRCA2 among others.

Breast cancer has now overtaken lung cancer as the most commonly diagnosed cancer in women worldwide, according to statistics released by the International Agency for Research on Cancer (IARC) in December 2020. In the past two decades, the overall number of people diagnosed with cancer nearly doubled, from an estimated 10 million in 2000 to 19.3 million in 2020 [1]. Today, one in 5 people worldwide will develop cancer during their lifetime. Projections suggest that the number of people being diagnosed with cancer will increase still further in the coming years, and will be nearly 50% higher in 2040 than in 2020. The number of cancer deaths has also increased, from 6.2 million in 2000 to 10 million in 2020. More than one in six deaths is due to cancer. This reinforces the need to invest in both the fight against cancer and cancer prevention

OBJECTIVE OF PROJECT

Breast cancer happens when cells in your breast grow and divide in an uncontrolled way, creating a mass of tissue called a tumor. The risk of developing breast cancer increases you age and with weight gain. Each year number of deaths is increasing extremely because of breast cancer. It is the most frequent type of all cancers and the major cause of death in women worldwide. Any development for prediction and diagnosis of cancer disease is capital important for a healthy life. Consequently, high accuracy in cancer prediction is important to update the treatment aspect and the survivability standard of patients. Deep Learning techniques can bring a large contribute on the process of prediction and early diagnosis of breast cancer, became a research hotspot and has been proved as a strong technique. The main objective of this research paper is to predict and diagnosis breast cancer, using Convolutional Neural Network i.e., with Deep Learning Technique, and find out accuracy. It is observed that Convolutional Neural Network outperformed the highest accuracy (100%). All the work is done in the Visual Studio Code with Anaconda Environment for python file and html folders, and Google Colaboratory is used.

PURPOSE

Death rate due to the cause of breast cancer is increasing day by day. So detection of Breast cancer is required for the early treatment. For that prediction of breast cancer is developed by using Convolutional Neural Network. By which prediction of breast cancer can lead to take the suitable treatment. So the chance of breast cancer risk can be reduced by prediction.

2. PROBLEM STATEMENT

Breast cancer is one of the main causes of cancer death worldwide. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible. The goal is to classify images into two classifications of malignant and benign. As early diagnostics significantly increases the chances of correct treatment and survival. In this application, we are helping the doctors and patients to classify the Type of tumor for the specific image given with the help of Neural Networks.

3. LITERATURE SURVEY

EXISTING SYSTEM

Breast cancer forms in breast cells and is considered as a very common type of cancer in women. Breast cancer is also a very life-threatening disease of women after lung cancer. Breast cancer is categorized into various types according to the cell's appearance through a microscope. The two main types of breast cancer are **(1)** invasive ductal carcinoma (IDC) and **(2)** ductal carcinoma in situ (DCIS). The IDC type is more dangerous, surrounding the entire breast tissue. Most breast cancer patients, approximately 80%, are in this category.

Due to the intrinsic difficulties associated with an image, with meager contrast, noise, and lack of appreciation by the eye, instruments have been prepared to make and improve image processing. Nowadays, Artificial Intelligence (AI), Machine Learning (ML) are the quickest rising areas of healthcare industry. AI and ML are found in the research arena that deals with and improves technological systems to resolve complex tasks through reducing necessity of human intelligence. Breast cancer can be effectively treated through its early detection. Thus, the availability of proper screening methods is important for detecting the initial symptom of breast cancer.

Deep learning (DL) which is part of machine learning family depended on artificial neural networks. DL architectures, such as DNN (deep neural networks), RNN (recurrent neural networks), DBN (deep belief networks), and CNN, are generally applied to the areas likecomputer vision, audio recognition, speech recognition, social network filtering, natural language processing, machine translation, drug design, bioinformatics, medical image analysis, materials scrutiny, Histopathological diagnosis, and board game programs . These new technologies, in particular DL algorithms, can be applied to improve the diagnostic accuracy and efficiency of cancer detection.

PROPOSED SOLUTION

Breast cancer starts when cells in the breast begin to grow out of control. These cells usually form a tumor that can often be seen on an x-ray or felt as a lump. The tumor is malignant (cancer) if the cells can grow into (invade) surrounding tissues or spread (metastasize) to distant areas of the body.

A convolution neural network (CNN) method is proposed in this study to boost the automatic identification of breast cancer by analyzing the Histopathological images. The opportunity that CNN brings to research on medical imaging is not restricted to deep CNN for extraction of the imaging feature. Indeed, a second field that can support medical research is the use of CNN for synthetic image rendering.

CNNs are applied to explore patterns in an image. This is done by convoluting over an imageand looking for patterns. The network can detect lines and corners in the few front layers of CNNs. Via our neural net, however, we can then transfer these patterns down and begin to identify more complex characteristics as we get deeper. This property ensures that CNNs are very effective at detecting objects in images .

The proposed system uses CNNs to detect breast cancer from breast tissue images. The architecture of a CNN has 3 main layers, the convolutional layer, pooling layer, and fully connected layer. The first layer calculates the output of neurons which are linked with local regions. Each one is calculated by a dot product of weights and the region. For image inputs, typical filters are small in area such as 3 × 3.

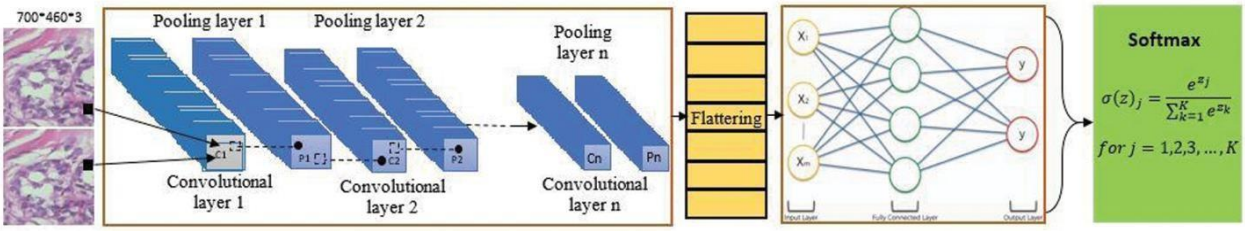

Figure 2: Training and Testing using Histopathological images with CNN

These filters scan the image by a sliding window on the image, while learning the recurrent patterns which arise in any area of the image. The interval between filters is known as the stride. The convolution is extended to overlapping windows if the stride hyper parameter is smaller than the filter dimension. Convolutional layers bring out the features of images with precise positions. If the positions change, even a small amount for any reason, the feature maps will be different. To overcome this problem, the down sampling process must be done at the output of every convolutional layer. With convolutional layers, down sampling can be done by changing the convolution's phase across the image. A more acceptable and common method is to use a pooling layer. Using this process, outputs will be more accurate.

4. EXPERIMENTAL A N A L Y S I S

Breast cancer is one of the main causes of cancer death worldwide. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible.

The goal is to classify images into two classifications of malignant and benign, as early diagnostics significantly increases the chances of correct treatment and survival. In this application, we are helping the doctors and patients to classify the Type of Tumour for the specific image given with the help of Neural Networks.

PROJECT ARCHITECTURE

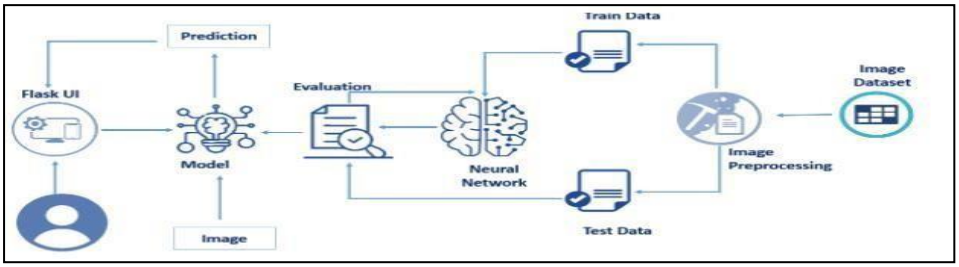

Figure 3: Project Architecture

BLOCK DIAGRAM

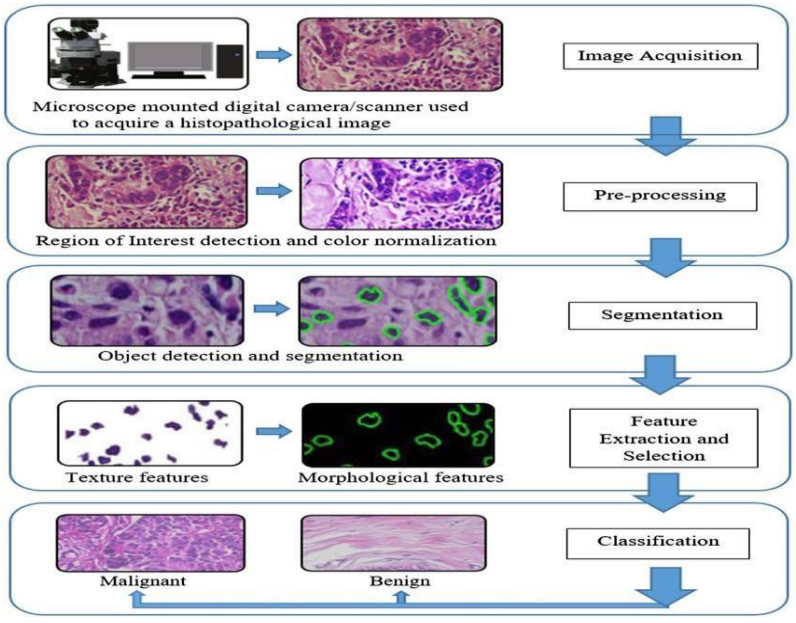

Figure 4: Block diagram representing the process of CNN

SOFTWARE REQUIREMENTS

- Python 3.9:
  - Python is an interpreted high-level general-purpose programming language.
  - Python can be used on a server to create web applications.
- Visual Studio Code:
  - Visual studio code is a source-coeditor made by Microsoft for Windows, linux and macOS.
  - Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.
- IBM Watson Studio
  - IBM Watson® Studio empowers data scientists, developers and analysts to build, run and manage AI models, and optimize decisions anywhere on IBM Cloud Pak® for Data.
- Anaconda Environment
  - The default environment base (path) is used because it consists of multiple libraries and modules.
- Tensorflow and keras modules:
  - Tensorflow and keras is used for the purpose of Convolutional Neural Network model building.
- Flask:
  - Flask is the module used for web framework.
  - Flask provides you with tools, libraries and technologies that allow you to build a web application.
- And other python libraries like NumPy, pandas, etc..,

PROJECT FLOW

1) Data Collection

a) In our project according to project structure, create train test folders and in them, place "0" folder images in benign and place "1" folder images in Malignant in train and test folders respectively as shown in the project structure.

2) Image Preprocessing

a) Import image data generator library and configure it

b) Apply image data generator functionality to train and test datasets

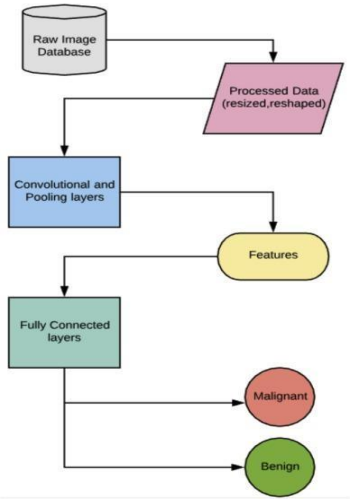

Figure 5: Project flow

3) Model Building

a) Importing the required libraries for model building:

Importing libraries such as keras and tensorflow for using the convolutional neural network.

b) Initialize the model:

Sequential is the model used for CNN, for that initializing the model is required.

c) Add convolution layer:

Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data.

d) Add max pooling layer:

Pooling layer is used for filtering the pixels size.

e) Add flatten layer:

Flatten layer is used for converting multidimensional array to single dimensional array.

f) Compile the model:

Compiling the model takes three parameters: optimizer, loss and metrics.

g) Fit and save the model:

To train, we will use the 'fit()' function on our model with the following parameters: training data, target data, and the number of epochs.

4) Test the Model

a) Import the saved model:

Importing the model that is saved in a plain text file (.h5).

b) Load the test image, preprocess it and then predict and check for results:

Preprocessing the image and predicting the image which is required.

5) Application Building

a) Build a FLASK application:

Flask provides you with tools, libraries and technologies that allow you to build a web application.

b) Build the HTML page and execute it:

HTML page is used for developing the webpage to display the result in webpage.

c) Run the app:

Run the python file such that the pages are rendered and linked to webpage's with a local host.

5. DESIGN

CLASS DIAGRAM

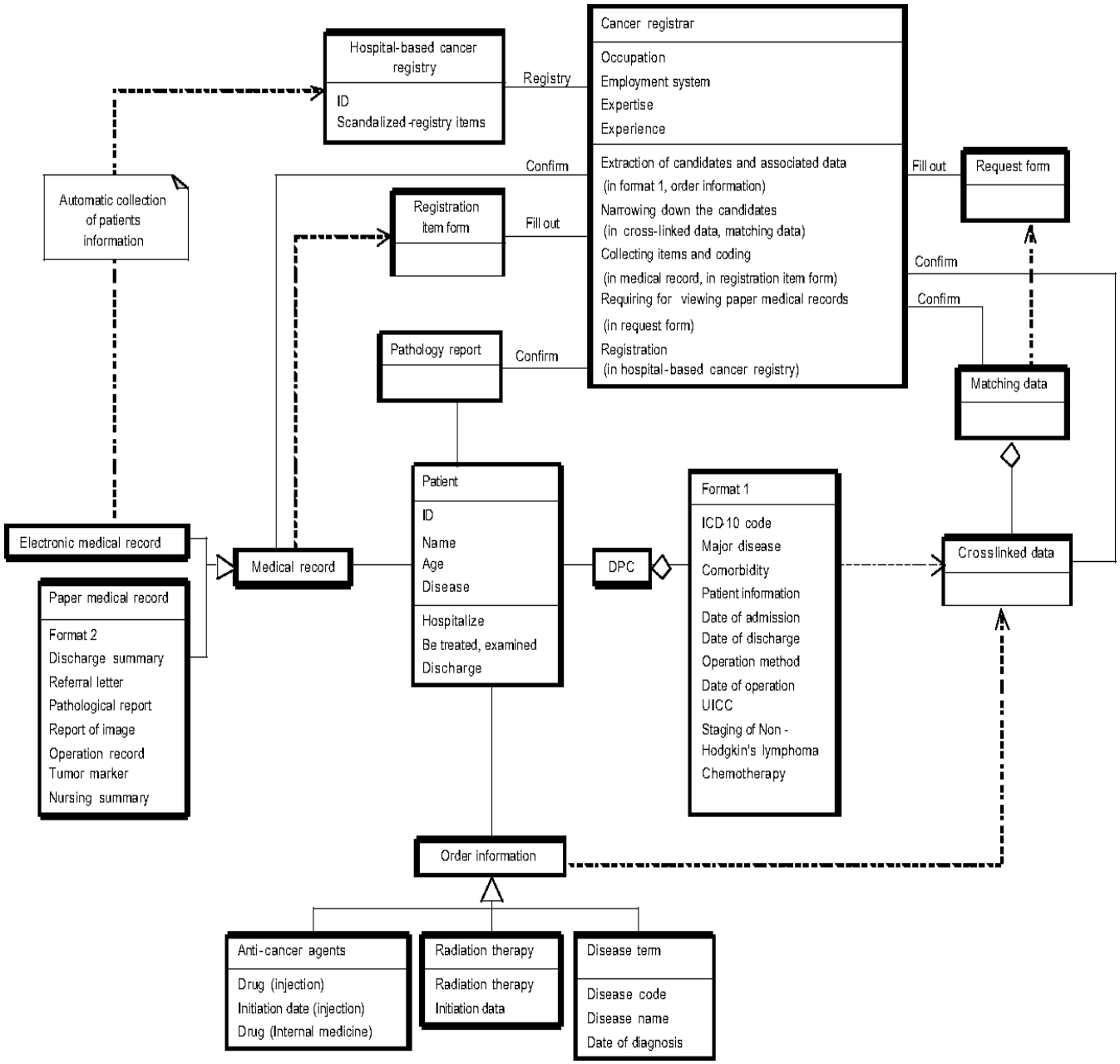

Figure 6: Class diagram

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and

operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of objectoriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram. A region in an organ or tissue which has suffered damage through injury or disease, such as a wound, ulcer, abscess, or tumor.

USE CASE DIAGRAM

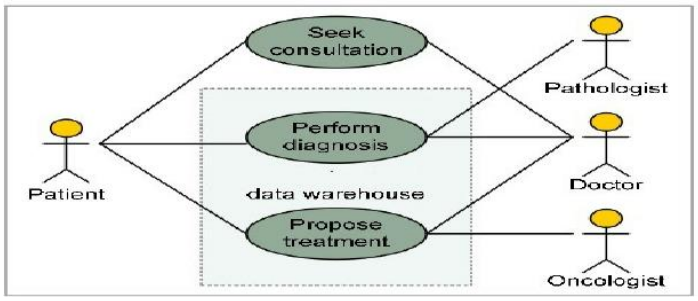

Figure 7: Use Case Diagram

A use case diagram is usually simple. It does not show the detail of the use cases:

- It only summarizes some of the relationships between use cases, actors, and systems.
- It does not show the order in which steps are performed to achieve the goals of each use case.

The use-case diagram corresponding to the project is depicted in Fig 7. There are five users:

- Patient
- Pathologist
- Doctor
- Oncologist

SEQUENCE DIAGRAM

A **sequence diagram** or **system sequence diagram** (SSD) shows object interactions arranged in time sequence in the field of software engineering. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of scenario.

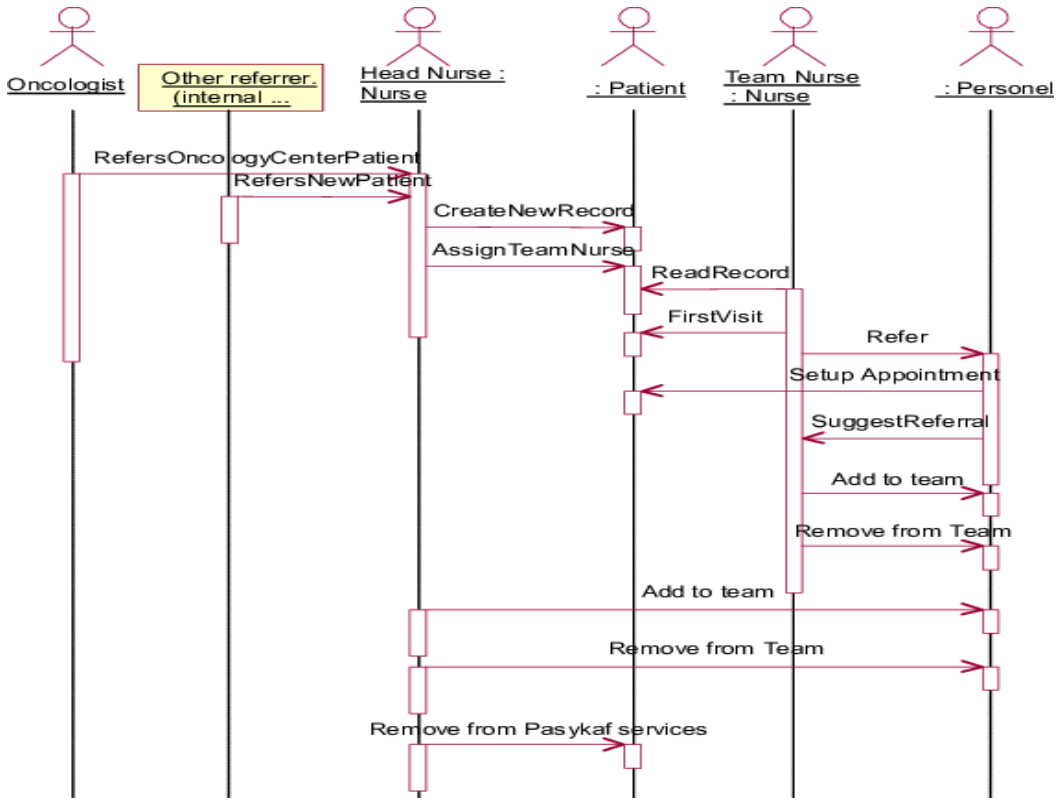

Figure 8: Sequence Diagram

FLOWCHART

A flowchart is a picture of the separate steps of a process in sequential order.

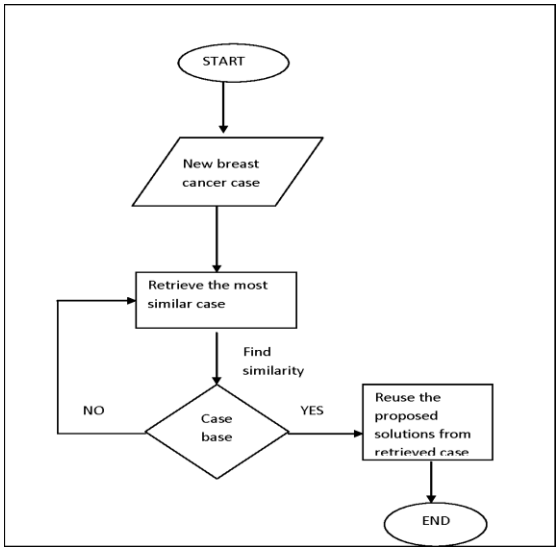

Figure 9: Flowchart

6. CODE SNIPPETS

MODEL CODE

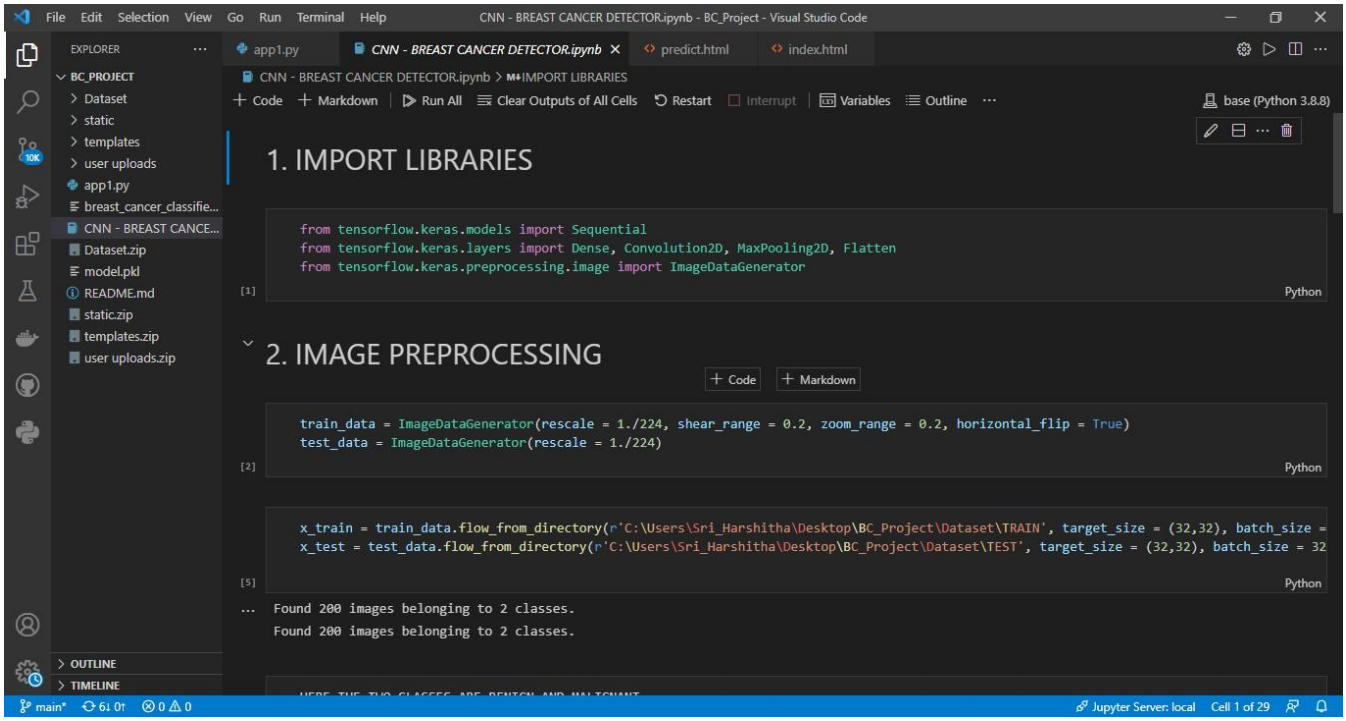

Figure 10: .ipynb code describing importing libraries and image preprocessing

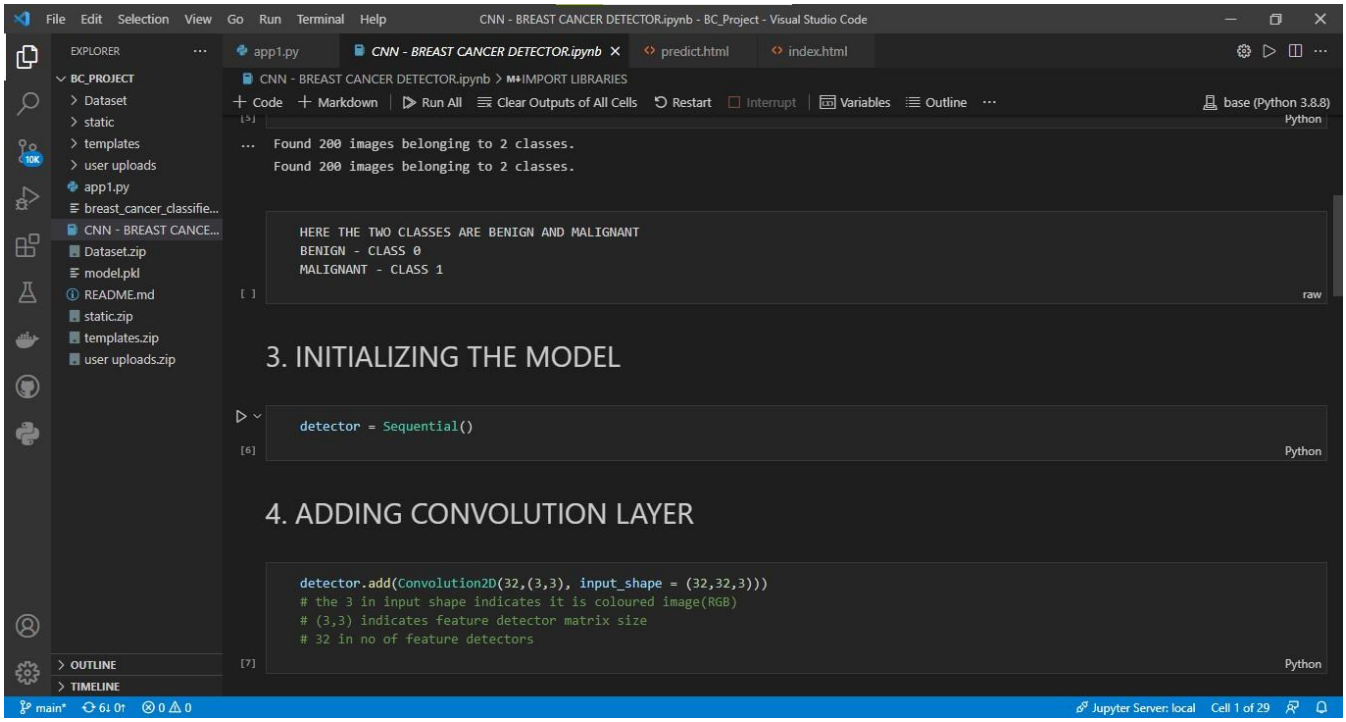

Figure 11: .ipynb code describing the model initialization and adding convolutional layer

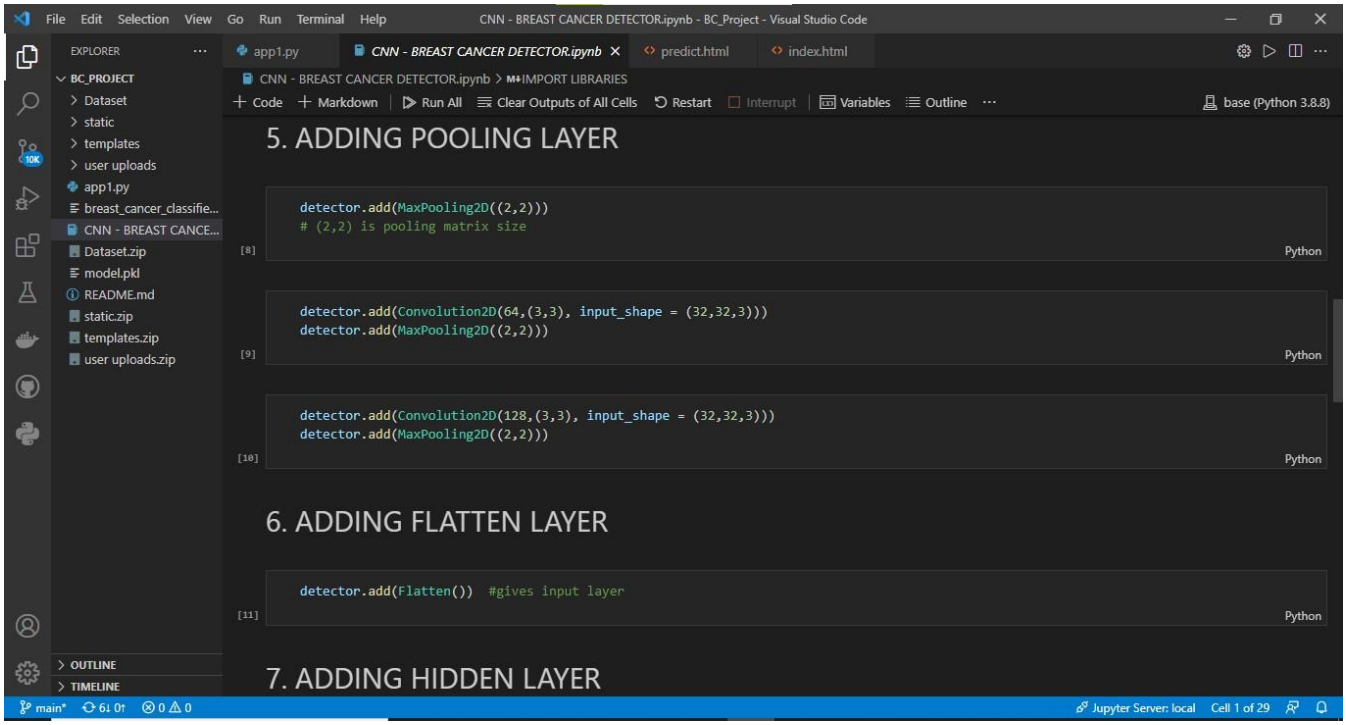

Figure 12: Adding layers to the sequential model

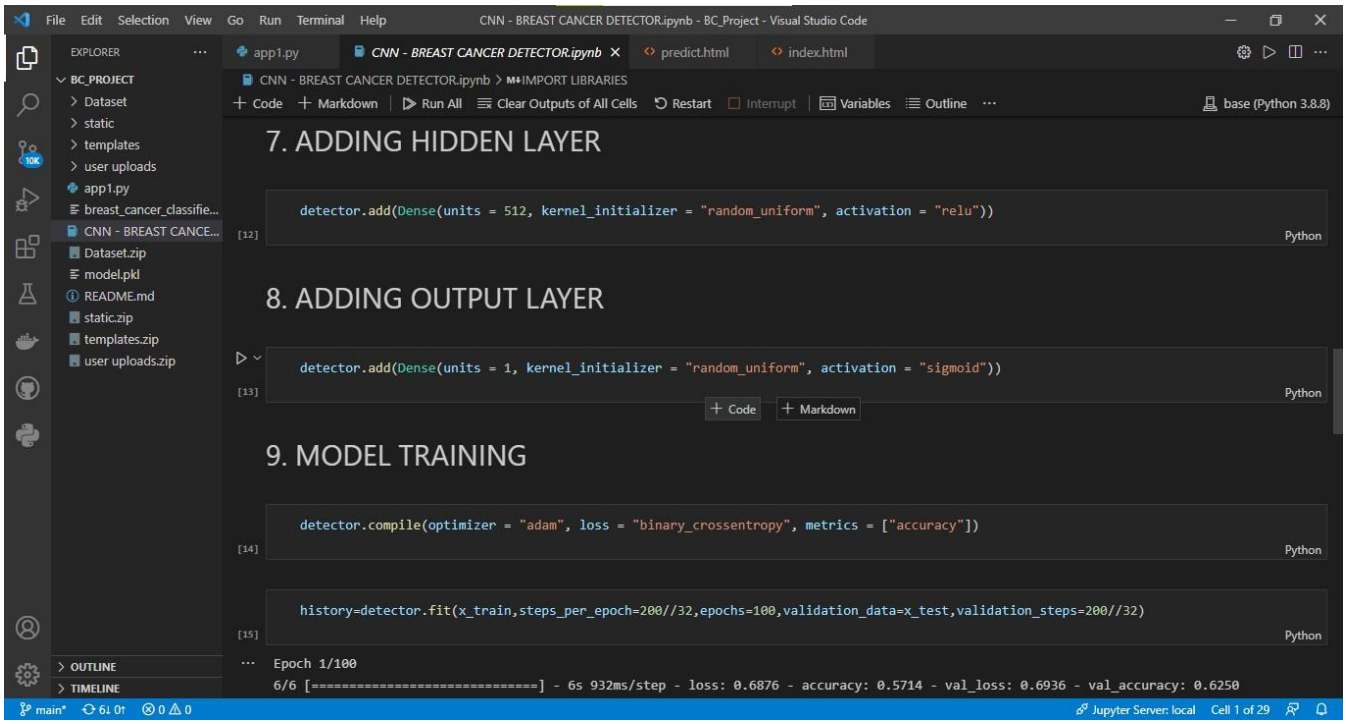

Figure 13: Adding Hidden and Output layers and training the model

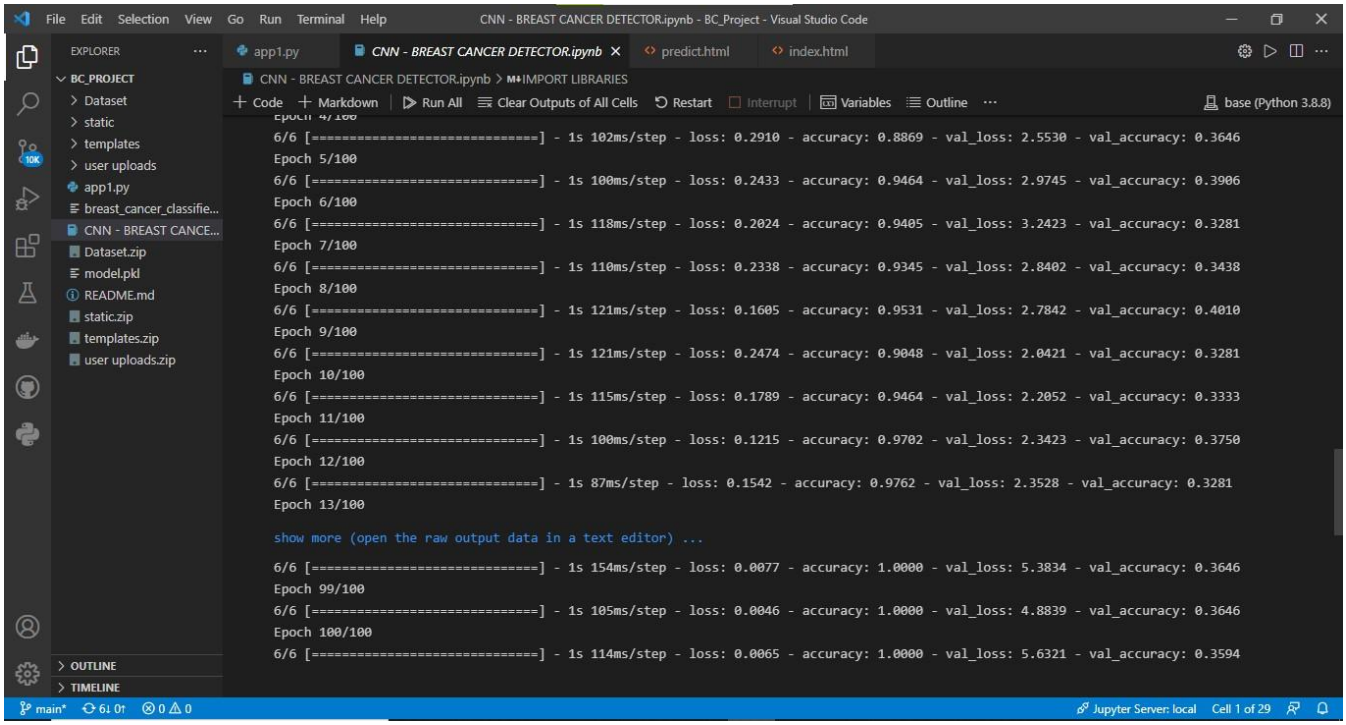

Figure 14: Epochs running with 100% accuracy

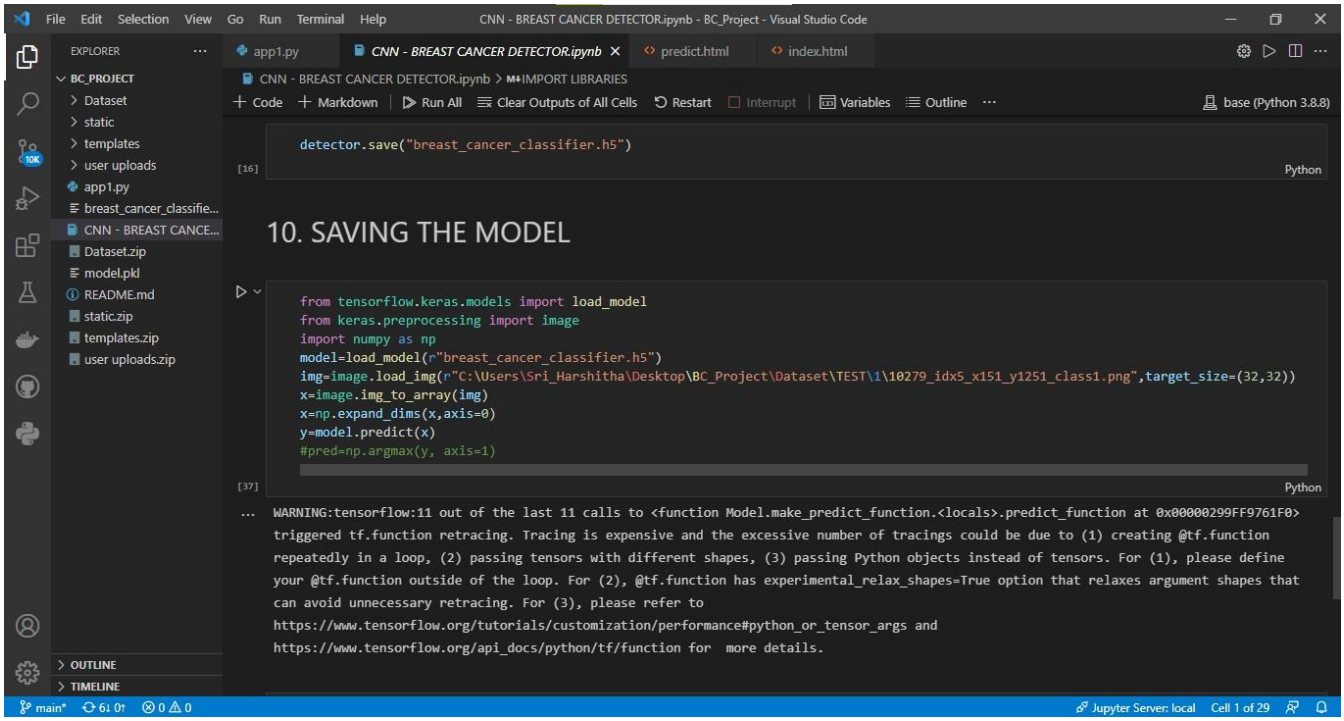

Figure 15: Saving the model

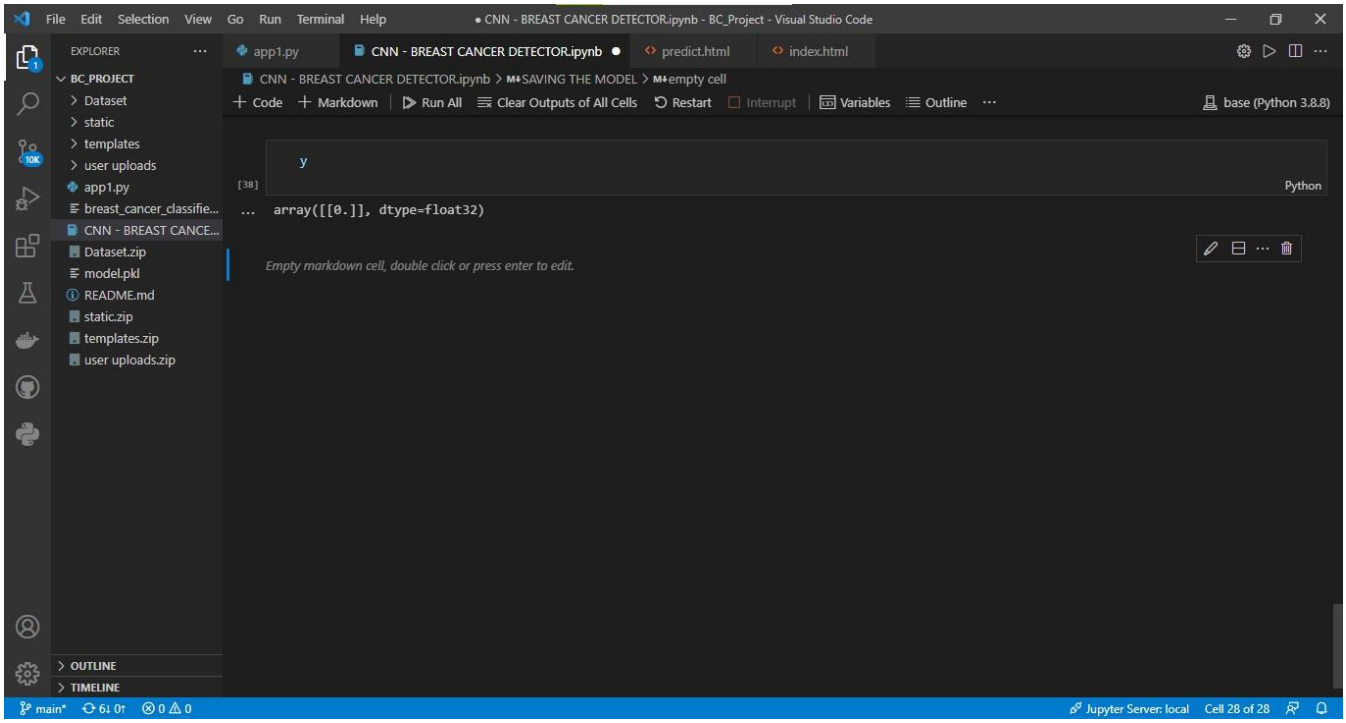

Figure 16: Predicting the model

HTML CODE OF INDEX

- **INDEX.HTML**

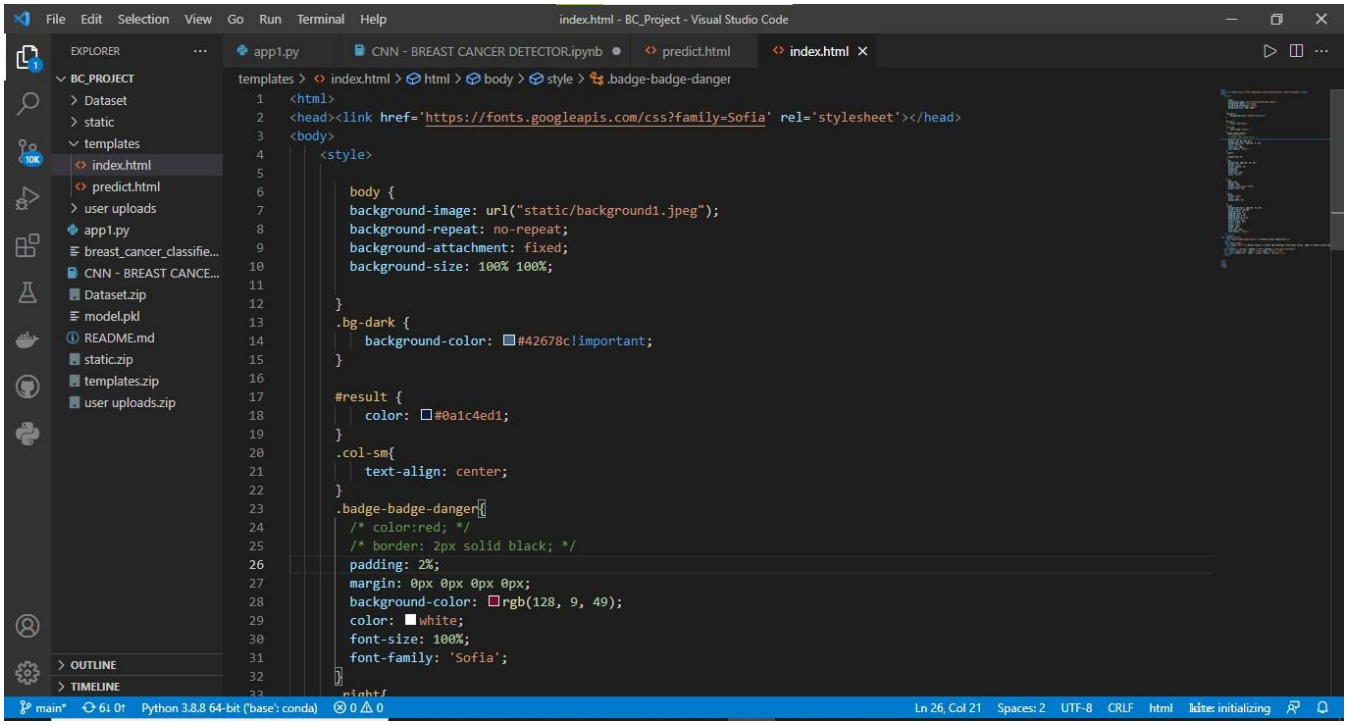

Figure 17 HTML Code of index.html describing the styles for webpage

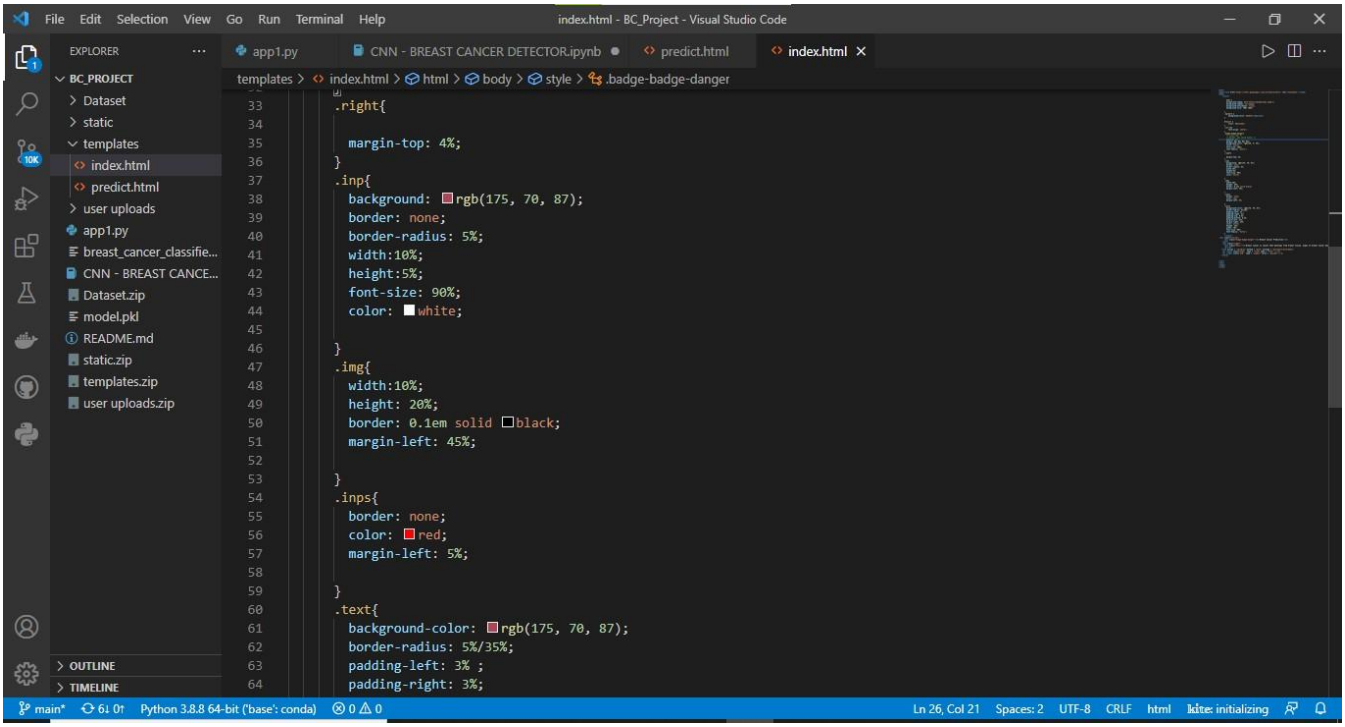

Figure 18: Styling the webpage in the body tag

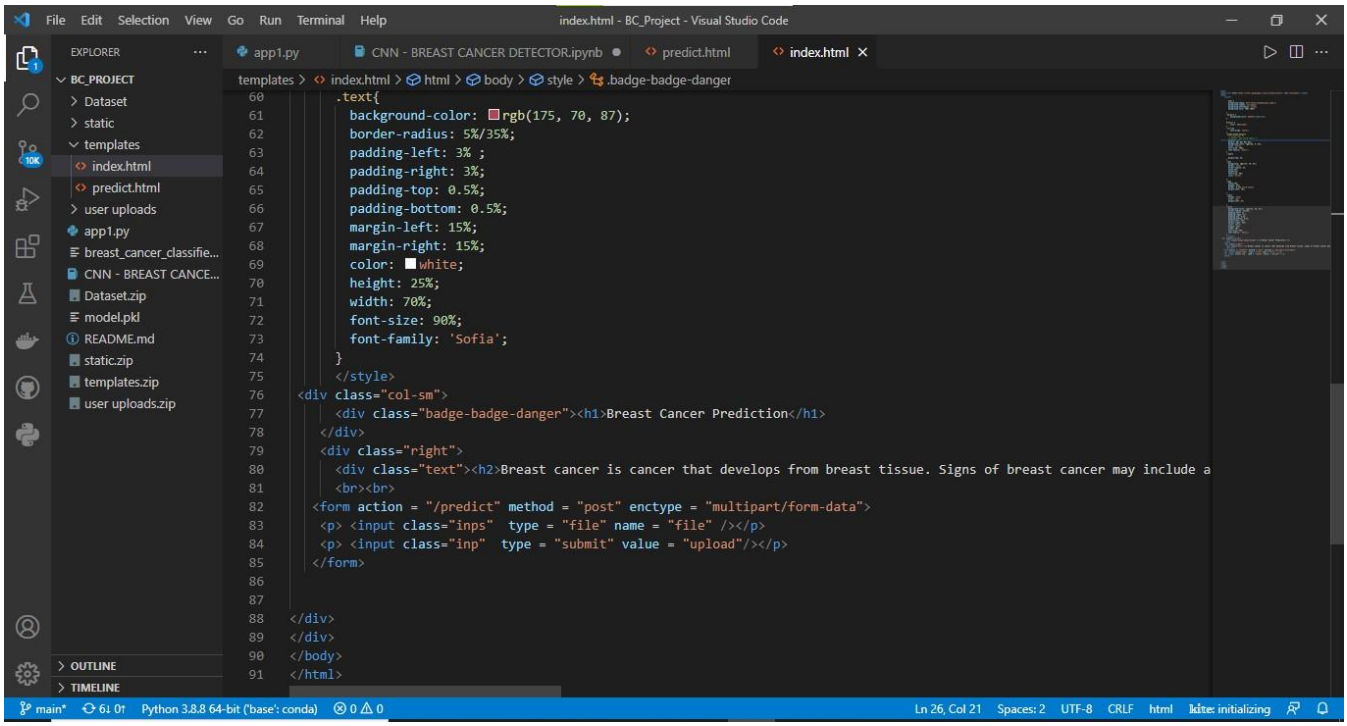

Figure 19: Describing the content of the webpage

- **PREDICT.HTML**

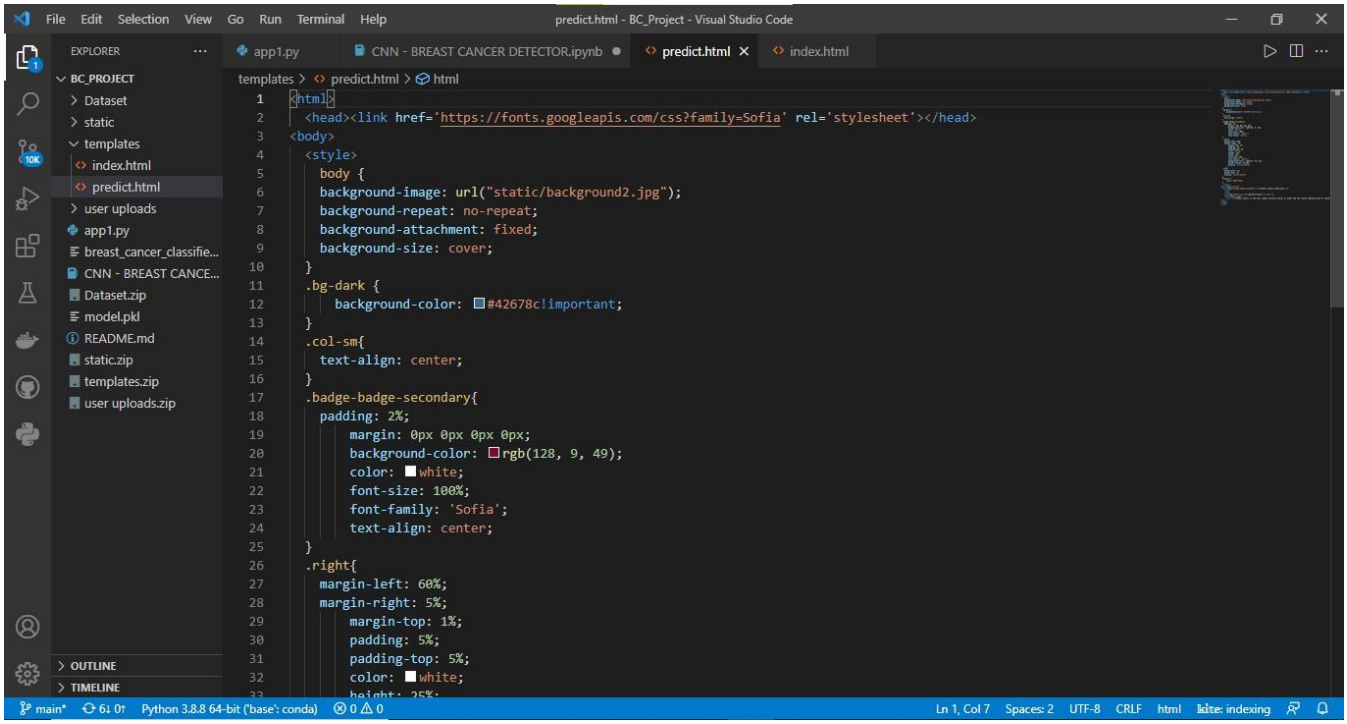

Figure 20: Styling the predict.html

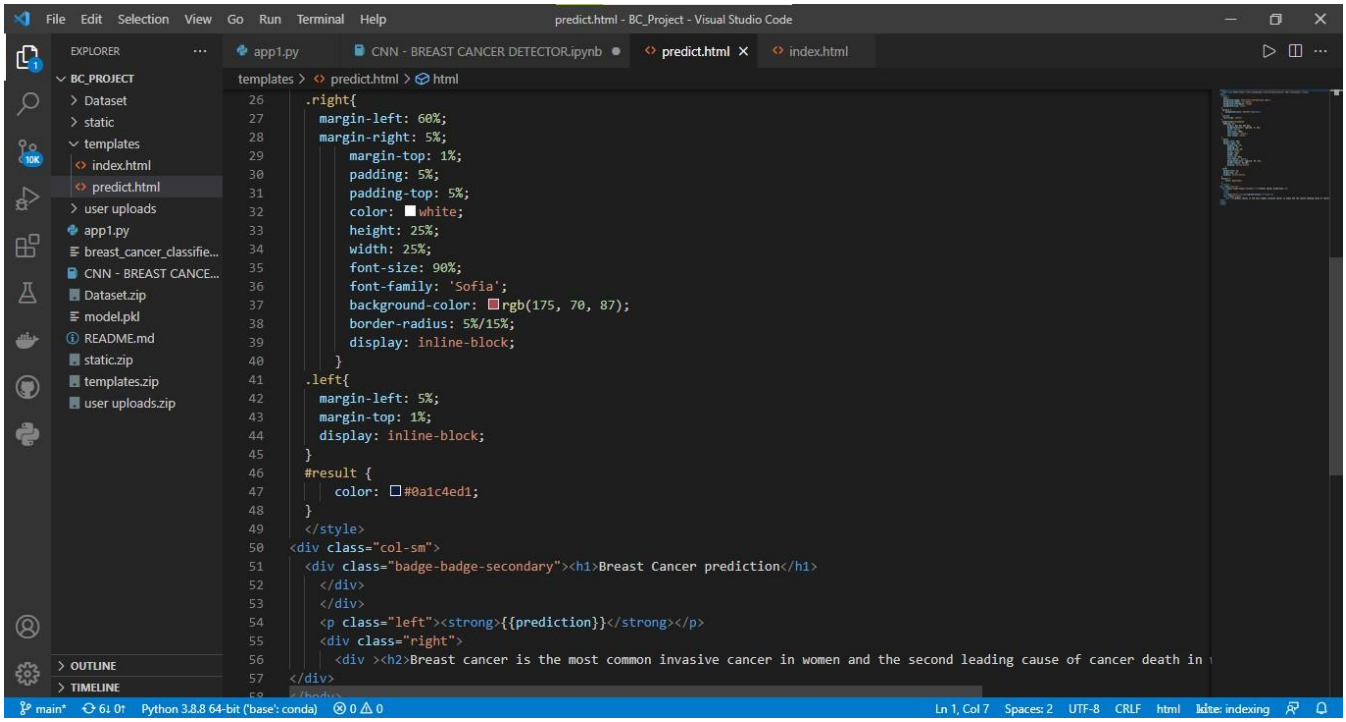

Figure 21: Styling the webpage in body tag

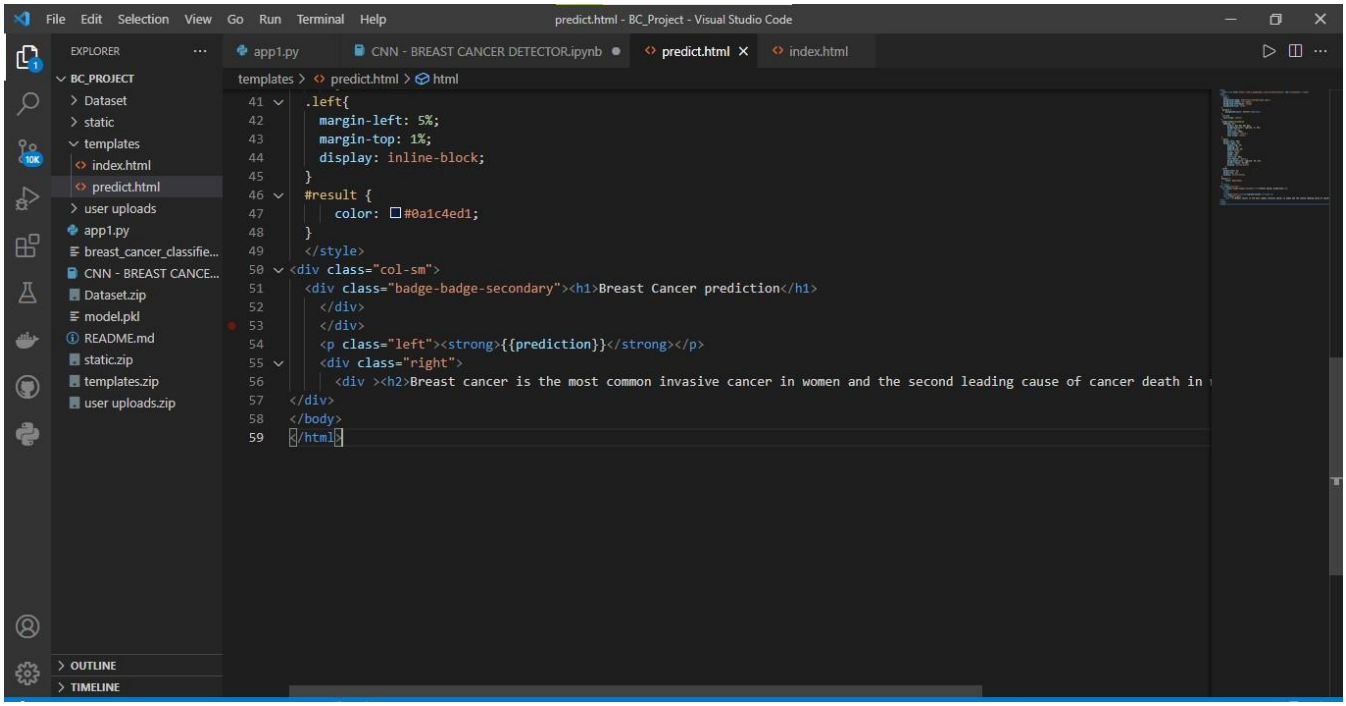

Figure 22: Content of prediction webpage

APP1.PY CODE

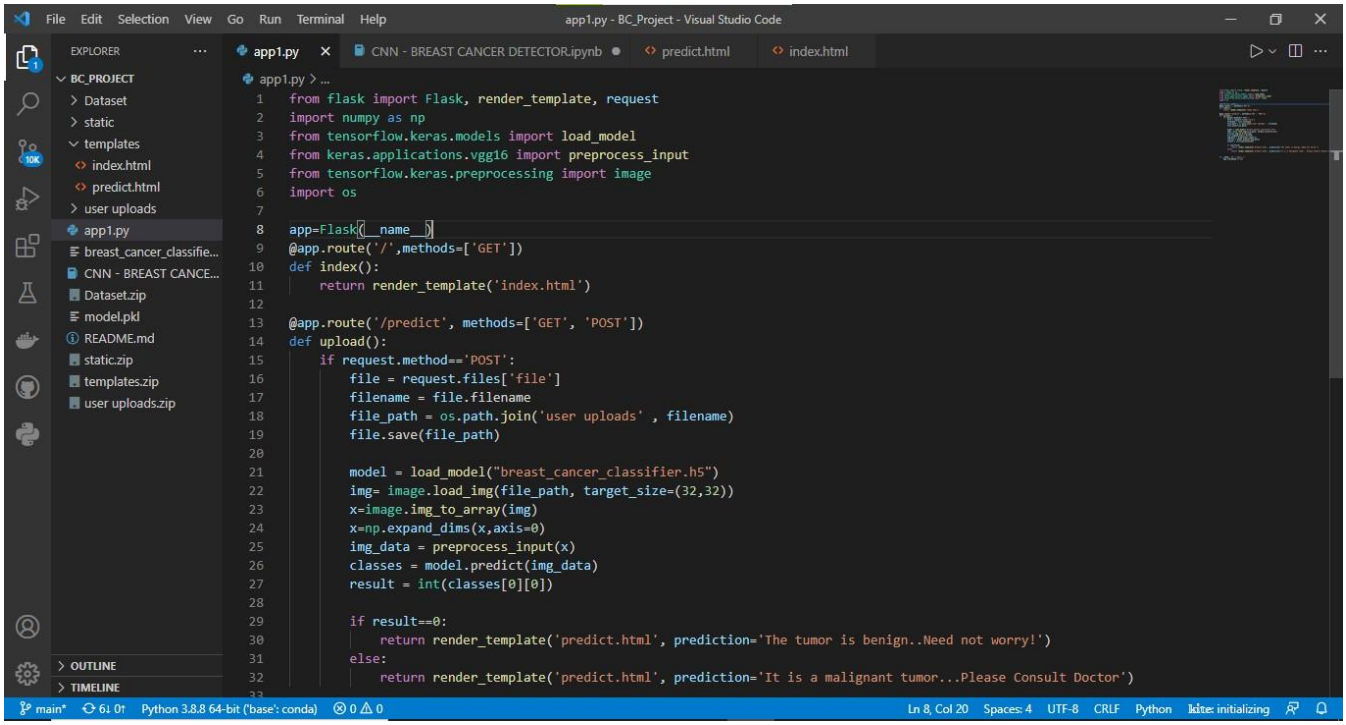

Figure 23: Python Code

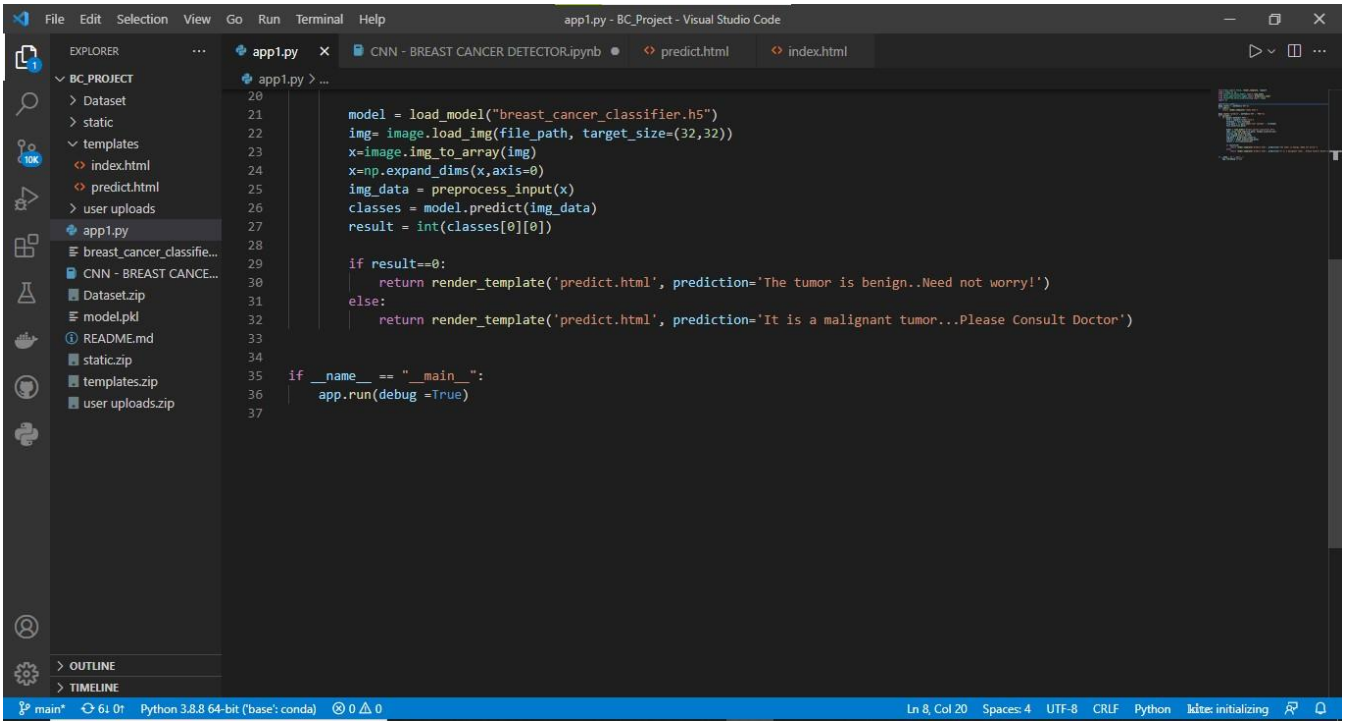

Figure 24: Code for prediction in python file

7. CONCLUSION

The following steps listed above are performed by our team, and herewith we attach snaps of our web page we achieved.

1. CLASS 1 (MALIGNANT)

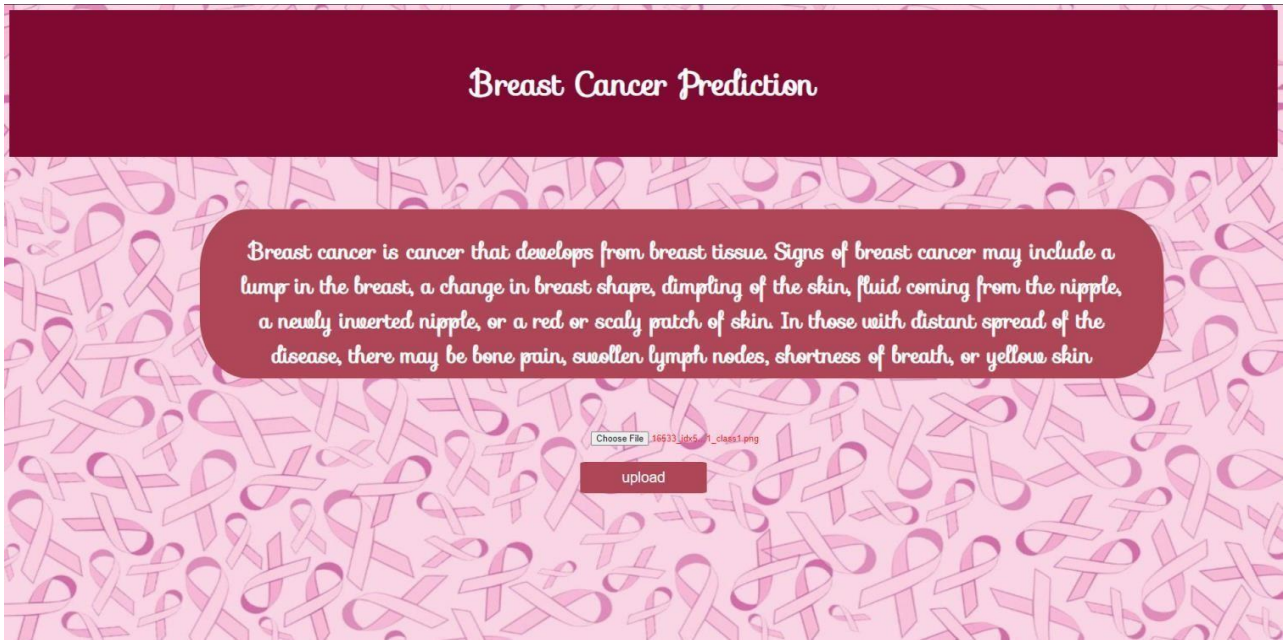

Figure 25: Index Webpage

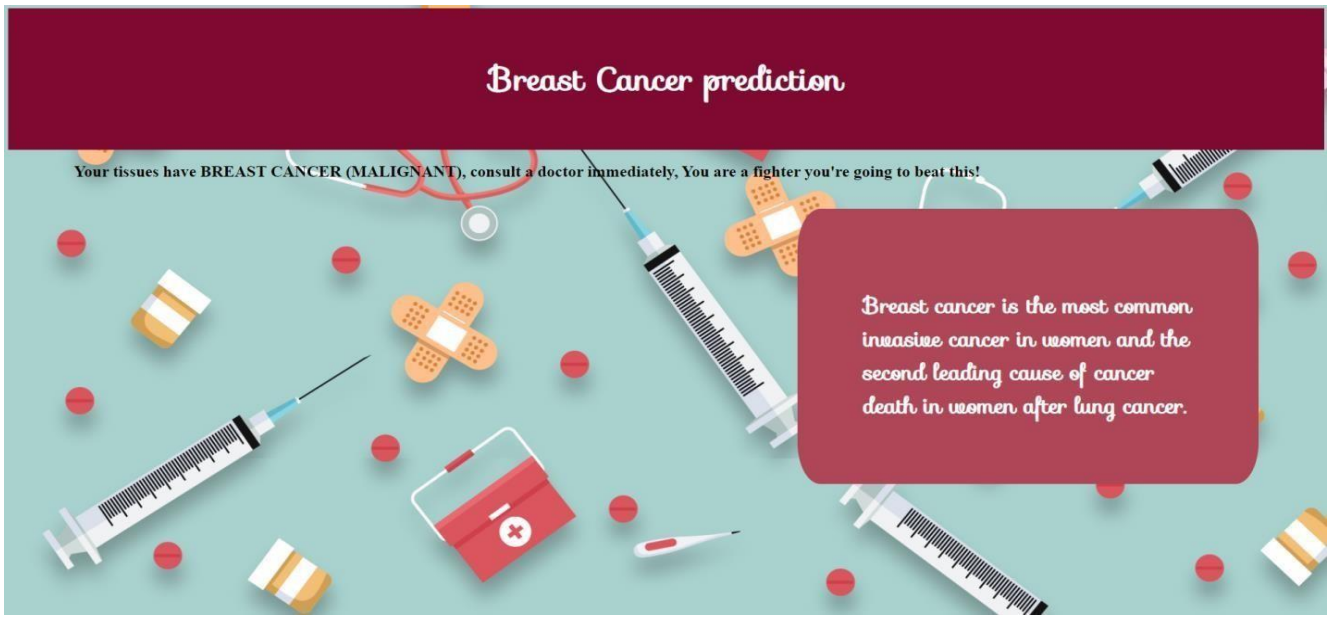

Figure 26: Predicting Breast Cancer of type Malignant

2. CLASS 0 (BENIGN)

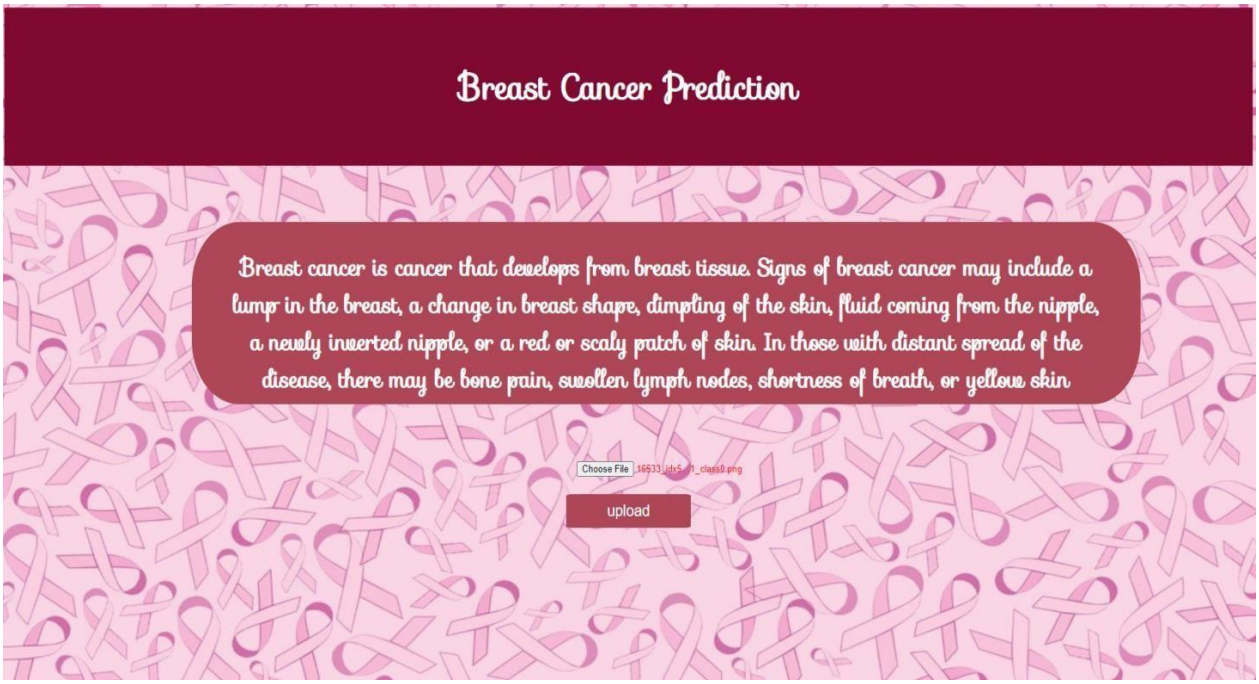

Figure 27: Index Webpage

Figure 28: Predicting Breast Cancer of type Benign

8. APPLICATION

The following application could be used in a better understanding with if patients past history details are also uploaded, and the accuracy can be increased if the system is fed to a large no of and variety of dataset. With various technologies and AI(artificial intelligence) applications coming up in the near future we can add a lot of advantage to the current system and get more realistic results. We can also apply a series of techniques involving scanned images as well as other related data points together for the classification of the dataset .Artificial Intelligence (AI)techniques could also be used to understand the different regions of the breast during classification.

9. ADVANTAGES

- Helps in medical department such as in laboratories for predicting the type of cancer.
- By the reports we can analyze the stage of breast cancer that patient is suffering from.

10. DIS-ADVANTAGES

- Cannot predict images other than Histopathological type Images.
- If any images other than Histopathological images are given then it automatically prints the default prediction that has been predicted previously.

11. FUTURE SCOPE

- In this system we have only considered Histopathological Images, but in further future prediction we can use mammograms, Breast ultrasound, Breast MRI scan.
- Here we used malignant and benign breast cancer types but in future we can use another types such as IDC(+ve and -ve).

12. REFERENCES

- https://smartinternz.com/Student/guided_projects
- https://www.kaggle.com/paultimothymooney/breast-histopathology-images
- Boosting Breast Cancer Detection Using Convolutional Neural Network (hindawi.com)
- W3Schools Online Web Tutorials