# Abstract

Breast cancer (BC) is the tumor that activates in the cells of the breast. Breast cancer can occur in both men and women, but it is for more common in women worldwide. Breast cancer is the second most common malignancy among Nepalese women that places a substantial burden on the Nepalese healthcare system. Breast cancer can be most treated Malignancies if detected early. If not, it is a deadliest one. Use of ICT in health sector will help to detect disease earlier and tries to increase the minimum life expectancy of people of the country. Thus, developing an automatic system for detecting breast cancer would be beneficial for treating the disease without any delay, particularly in remote areas. Also, the classification of the breast cancer as malignant and benign tumor that will further help the doctors in identification of the causes of breast cancer.

The use of Convolutional Neural Network (CNNs) has gained much popularity for disease classification, due to the huge success of deep learning algorithms in analyzing medical images. In addition, features that are learned by the pre-trained CNN model on large number of datasets helps to classify the images with much accuracy.

In this project, we detect breast cancer from chest mammography (also called Mastography) using Convolutional Neural Network (CNN). This project will not only detect if the women have breast cancer or no breast cancer but will further classify the breast cancer as either a malignant or benign tumor. To improve the accuracy and performance of the project, we will implement the image augmentation techniques. Thus, we will use CNN from keras module to develop this model, and calculated accuracy using confusion matrix.

Thus, this project will not only save the time for detection of breast cancer but also help the doctor to better classify the cancer as malignant or benign tumor so that the doctors can be more specific while providing prescription and treatment to the patients.



***Keywords**: Artificial neural network (ANN), Convolutional neural network (CNN), confusion matrix, image processing, RGB.*

# Table of Contents

# List Of Figures

# List Of Tables

# CHAPTER 1: INRODUCTION

## 1.1. Background

Breast cancer is a multifactorial disease that forms in the cells of the breast. Breast cancer can occur in both men and women, but it is for more common in women worldwide. Breast cancer is the second most common malignancy among Nepalese women that places a substantial burden on the Nepalese healthcare system. Breast cancer can be most treated Malignancies if detected early if not it is a deadliest one. Use of ICT in health sector will help to detect disease earlier and tries to increase the minimum life expectancy of people of the country. '**Breast Cancer Prediction System using CNN Algorithm**' is an online based system implemented as web-based application that can be used by an end user (medical personnel) or by an oncologist, so that the user can work out with the interface and find if his/her patient is suffering from cancer which either can be benign or the malignant. The 'Breast Cancer Prediction System**'** is an end user support and online constitution project. Here the system will allow users to get instant result on their breast related issues through an intelligent healthcare system online. The system allows users to share their symptoms and issues in the form of ultrasound image. It then will process user's input images to check for various possible breast related problems that can be associated with it, with two different types of tumours can be encounter early, those are:

1. Benign Tumor

2. Malignant Tumor

A tumor is an abnormal collection of cells. It forms when cells multiply more than they should or when cells don't die when they should. A tumor can be malignant (cancerous) or benign (not cancerous).

## Benign Tumor:

Benign tumors are those that stay in their primary location without invading other sites of the body. They do not spread to local structures or to distant parts of the body. Benign tumors tend to grow slowly and have distinct borders. Benign tumors are not usually problematic. However,

they can become large and compress structures nearby, causing pain or other medical complications. For example, a large benign lung tumor could compress the trachea (windpipe) and cause difficulty in breathing. This would warrant urgent surgical removal. Benign tumors are unlikely to recur once removed. Common examples of benign tumors are fibroids in the uterus and lymphomas in the skin. Specific types of benign tumors can turn into malignant tumors. These are monitored closely and may require surgical removal. For example, colon polyps (another name for an abnormal mass of cells) can become malignant and are therefore usually surgically removed.

### Malignant Tumor:

Malignant tumors have cells that grow uncontrollably and spread locally and/or to distant sites. Malignant tumors are cancerous (i.e., they invade other sites). They spread to distant sites via the bloodstream or the lymphatic system. This spread is called metastasis. Metastasis can occur anywhere in the body and most commonly is found in the liver, lungs, brain, and bone. Malignant tumors can spread rapidly and require treatment to avoid spread. If they are caught early, treatment is likely to be surgery with possible chemotherapy or radiotherapy. If the cancer has spread, the treatment is likely to be systemic, such as chemotherapy or immunotherapy.

Here, This Prediction System predicts the probability and also provide certainty of having cancerous tumors on the body based on available ultrasound images of breast cancer. There are different methods including both supervised and unsupervised learning classifier under deep learning such as neural network,

## 1.2. Problem statement

Breast cancer is very complex and shows multiple clinical manifestation. It might rare happened so many times that you or someone you know need doctor immediately or at earlier stage, but they are not available due to some reason lack of access to effective healthcare services including timely detection services, are some of the reasons why countries like Nepal have high burden of breast cancer. Hence, to make users aware about different tumor detection,

to improve the diagnostic capability and to reduce the possible diagnostic errors, the system Breast Cancer Prediction System can be helpful.

## 1.3. Objectives

The objectives of the project to develop a system are:

- To build a classification model for predicting the types of tumors.
- To ensure and impact the availability of the system.

## 1.4. Scope and Limitations

- The implementation of this system helps the cancer patient to detect the tumor earlier with the help of Oncologist and mammographic (ultrasound) images.

There are some limitations with the projects which are follows:

- The system is limited to predict the probability of either having or not having cancerous tumors, but it can't be used by the patient themselves.
- This system is based on the mammographic images, but it can't differentiate if the supplied image is mammographic or another x-ray image.

## 1.5. Development Methodology

For the completion of the project successfully, there is great role of System development methodology. For this project, Agile methodology is selected. The task of the system has been broken down into manageable units which assists to focus on high quality development, documents, and testing. Frequent testing and reviews help to improve the quality of the system. Similarly, Agile is flexible to change even in late stage which is beneficial in case of any change in requirements of our system.

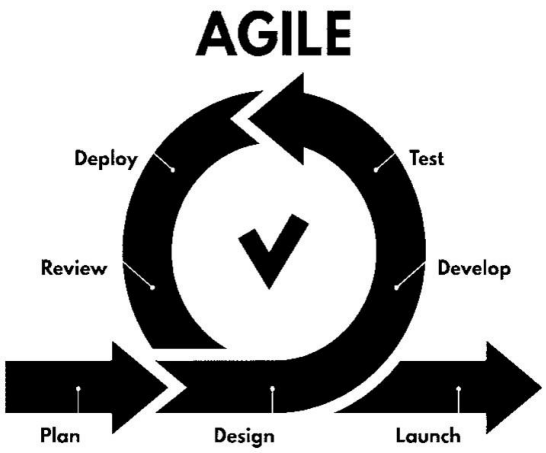



Figure 1: Agile methodology

## 1.6. Report Organization

The report is separated into different chapters which are as follows:

The report organization of each chapter that have on the report it arranges will refer to the specific format and it easy for readers to understand the whole of report. The report is started with **chapter 1** that explains about introduction, problem statement, objectives, scope and limitations and methodology. The next **chapter 2** explains about the literature review related the paper research for the system development. Then, **chapter 3** discussed about the system analysis where requirement analysis and feasibility analysis are required in system development. Then, **chapter 4** deals the project design and modelling are the core part of the development process. Then, **chapter 5** discussed about the implementation and testing of the system. Lastly, **chapter 6** conclude on the system development.

# Chapter 2: Background Study and Literature Review

## 2.1. Background Study

There are various ways to detect breast cancer including Mammography, Magnetic Resonance Imaging (MRI) Scans, Computed Tomography (CT) Scans, Ultrasound, and Nuclear Imaging. Although, none of these techniques gives a completely correct prediction of cancer. Tissuebased diagnosis is mainly done with a staining methodology. In this procedure elements of tissues are colored by some staining element, usually hematoxylin and eosin (H&E). Cell structures, types, and other foreign elements are stained accordingly and are easily visible under high resolution. Pathologists then examine the slide of stained tissues under a microscope or using high-resolution images taken from the camera. For the detection of tumors, a histopathology test is essential. It is an old method used to predict invasive cancer cells from H&E-stained tissues. There are various shortcomings for this procedure as it involves intra-observer variation, cancer cells and tissues can also have multiple appearances, and many other figures in cells have the same hyperchromatic features, which make identification difficult. The choice of area is also a factor as the process is done only on a small area of tissue, so the chosen area should be in the tumor periphery. The above-mentioned problems can be addressed by using deep learning strategies. Deep learning is a popular subset of machine learning technology that is inspired by the working of the human brain to analyze unstructured patterns.

Breast Cancer Detection and classification will provide better detection for breast cancer cases from mammography. This will reduce the chances of mistakes during the diagnosis of either a malignant or benign tumor. On implementation of this project, it will not require a trained and expert medical radiologist or specialist to determine if the person has Breast Cancer or not. Since it is difficult to differentiate between malignant or benign tumors, utilization of such a model will help the doctors to correctly determine the accurate cause at the earliest.

So, to know about the most leading cancer in the world and its detection techniques, its screening techniques, the effect of late detecting about the tumors which leads to the untimely death. So, the background behind our project is to develop the system which helps the cancer patient in early detection of tumors which can be treated earlier.

## 2.2. Literature review

The research works done in breast cancer detection customarily follow either image processing and segmentation techniques or uses machine intelligence algorithms on the extracted features to classify the images as malignant or benign. There are very few instances where both the image processing and machine intelligence methodologies are used. [1]

Some previously completed paper had used different types of neural network for classification mammogram image in breast cancer. S.Das[2] used the mammographic image for the prediction of breast cancer by using artificial neural networks for classification and prediction.[2]

The wavelet neural network is employed for breast cancer diagnosis [3]. Both neural networks are designed for general decision-making purpose, so both need to setup many parameters than CNN which is designed for purpose recognize visual imagery. Same number of hidden layers, standard neural networks working on processing visual imagery need have 3x106 parameters. But for CNN, it only needs around 600 parameters for processing visual imagery. H.Zhou et. Al. [4] suggested a mammogram image detection using CNN. But the accuracy percentage is too low for a medical side solution which is around 60% for all classes' detection, 75% for only masses class, and 100% for only calcification. Except only calcification argument, all argument and mass only argument can further be improved the accuracy to get a better performance.[4]

Lo et al.[5] proposed an automated mass detection scheme on automated breast ultrasound images using watershed transform. Ye et al. [6] reported a mass detection algorithm that consists of three major steps: active contour-based segmentation, feature extraction and classification. [7]

Similarly, Naive Bayes Classifier has been applied to Wisconsin Prognostic Breast Cancer (WPBC) dataset (UCI Machine Learning repository: http://archive.ics.uci.edu/ml/machinelearning-databases/breast-cancer-wisconsin/),
concerning several 198 patients and a binary decision class: non-recurrent-events of 151 instances and recurrent-events of 47 instances. The input features contain 12 relevant attributes describing the characteristics of cell nuclei. The testing diagnosing accuracy was about 74.24% in accordance with other well-known Machine

Learning techniques.[8] Authors analyze the performance of supervised learning algorithm such as Naive Bayes, SVM Gaussian RBF kernel, RBF neural networks, Decision tree J48 and simple CART. These algorithms are used for classifying the breast cancer datasets WBC, WDBC, Breast tissue from UCI Machine learning Repository (http://archive.ics.uci.edu/ml). They conducted their experiments using WEKA tool. In which the accuracy percentage of Naive Bayes algorithm for WBC dataset yields to be 96.50%, for Breast tissue dataset comes to be 94.33% and for WDBC dataset it is 92.61%.[9]

## 2.3. Domain Research

As we have planned to implement CNN (Convolutional Neural Network) to develop a "Breast Cancer Prediction" there has been some work already done in the past through this algorithm. CNN is known as feed forward neural network which works by inspecting images.

### 2.3.1. CNN (Convolutional Neural Network)

CNNs are applied to explore patterns in an image. This is done by convoluting over an image and 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, as shown in Figure 2.1. 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 $3 \times 3$, $5 \times 5$, or $8 \times 8$. 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 hyperparameter is smaller than the filter dimension. A detailed visual explanation of neural networks (NNs) is shown in Figure 2.2.

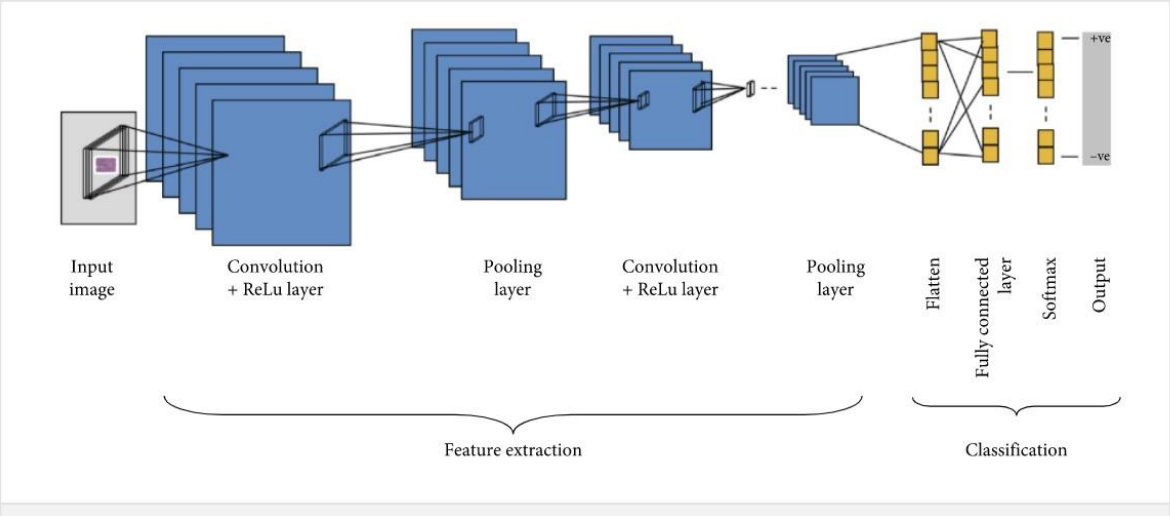



Figure 2: Typical CNN architecture for automatic detection of IDC breast cancer.

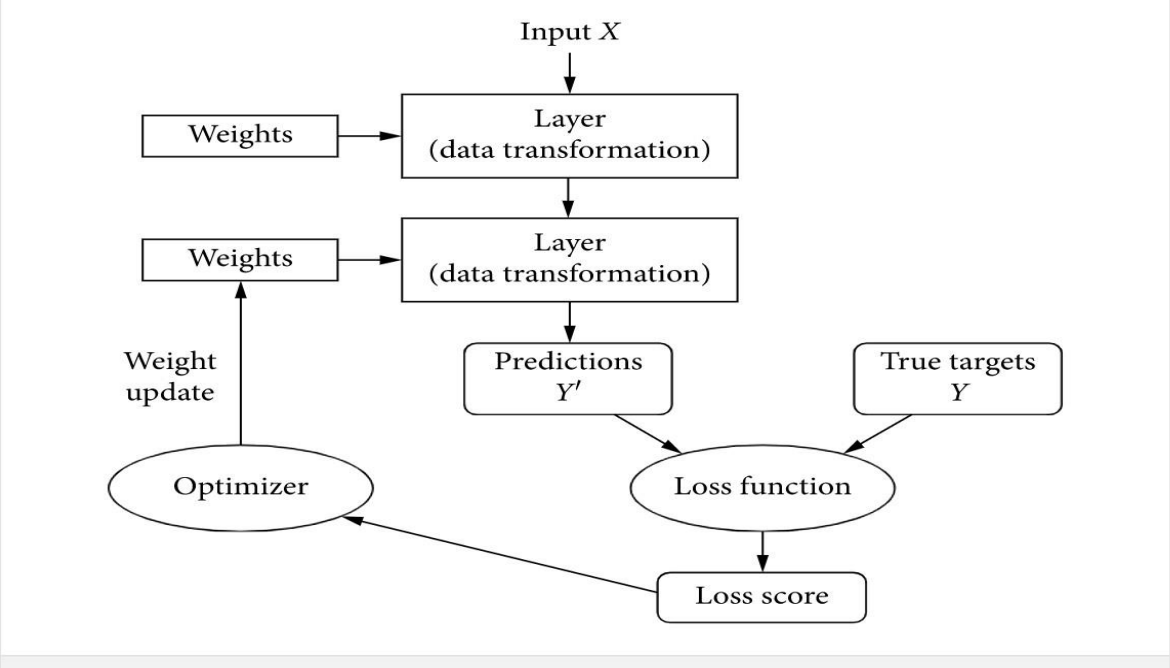



Figure 3: Detailed process of a neural network (NN).

### i. Convolutional Layer

The purpose of this layer is to receive a feature map. Usually, we start with low number of filters for low-level feature detection. The deeper we go into the CNN, the more filters we use

to detect high-level features. Feature detection is based on ‘scanning’ the input with the filter of a given size and applying matrix computations to derive a feature map.

### ii. Pooling Layer

The goal of this layer is to provide spatial variance, which simply means that the system will be capable of recognizing an object even when its appearance varies in some way. Pooling layer will perform a down sampling operation along the spatial dimensions (width, height), resulting in output such as [16x16x12] for pooling size= (2, 2).

### iii. Fully Connected Layer

In a fully connected layer, we flatten the output of the last convolution layer and connect every node of the current layer with the other nodes of the next layer. Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks and work in a similar way.

# Chapter 3: System Analysis

## 3.1. System Analysis

System analysis involves requirement analysis: Functional Requirements (Illustrated using use case diagram/use case descriptions), Non-Functional Requirements and Feasibility Analysis.

### 3.1.1. Requirement Analysis

The system requirement collection specification of the project consists of functional and nonfunctional requirement.

## I. Functional Requirements

Functional requirement are statements of services that system provide. First the user must provide the attributes and symptoms they have faced then the system predicts the most probable diseases further process is explained by the use case diagram. The user interface is simple to use.

**Input data**: System provides the user/admin to input the data that later feed to algorithm.

**View result**: Result can be viewed by user using system which shows the prediction of types of tumors as benign, malignant.

**Add dataset**: In the system, admin can add dataset, train it, and test it.

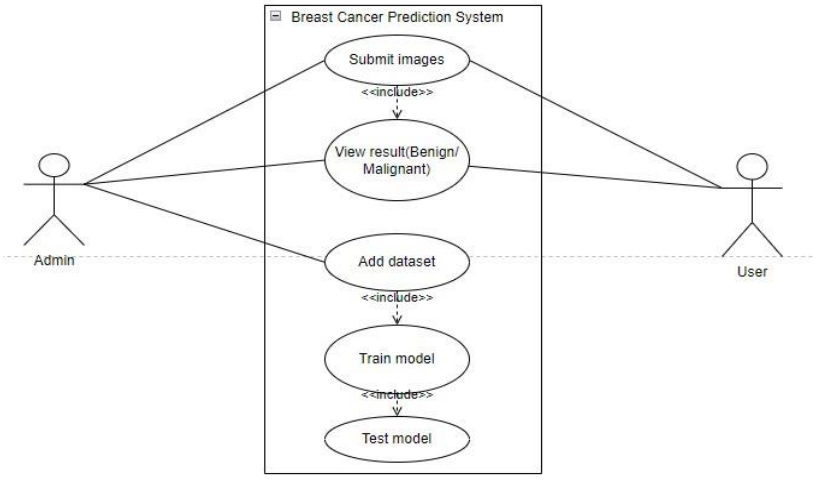



Figure 4 Use Case of the Proposed System

## ii. Non-Functional Requirements

Some of the non-functional requirements of Breast Cancer Prediction System are summarized as following:

- **Availability:** The system can be accessed by anyone and anywhere having PC and Internet connection.
- **Maintainability:** The system can be restored to operational status after a failure occurs, so the system is highly maintainable.
- **Scalability:** The system can be enhanced for other diseases in future.
- **Reliability:** The reliability of the system is defined by the overall accuracy of the system and the ability of the system to predict a correct output from the given input data.
- **Ease of use:** The system has a user-friendly interface so that any user can use the system without facing any difficulties.

## 3.1.2. Feasibility Analysis

Here, we have studied all the feasibility aspects of the project under consideration to check out if the project is feasible with the decided requirements and availability of information, technologies, and budget.

### I. Technical feasibility:

The technical resources available to the project is focused here. Evaluation of the hardware, software, and other technology requirements of the proposed system are involved here. This will help in determining the technical resources meet the capacity and whether the technical team can convert the ideas into the working system. Hence our system is technically feasible because:

- The hardware and software required are easy to install and handle.
- The system supports interactivity with GUI.
- It requires less system resources.
- Have User-friendly interfaces.

### ii. Operational feasibility:

This involves undertaking a study to analyze and determine whether and how well the needs of the organizations can be meet by completing the project. This will analyze how a project

satisfies the requirements that are identified in the requirement analysis phase of system development.

The system is operationally feasible as it's very easy for the End users to operate it and contains all the requirements that are promised to keep. It only needs basic information about Windows platform.

**iii. Economic feasibility:**

This involves a cost/benefits analysis of the project, helping organizations determine the cost, viability and benefits associated with a project before financial resources are allocated. This will consider the following:

- The cost of hardware and software for the project.
- The cost to conduct a full system investigation.

The proposed system is economically feasible as it does not require enormous amount of money to be developed. The system will deliver fast and effective automated environment instead of slow and error prone manual system, thus reducing both time and manpower spent in running the system. And the system will have GUI interface and very less user-training is necessary to learn it.

**iv. Schedule feasibility:**

The importance of a feasibility study is based on organizational desire to "get it right" before committing resources, time, or budget. A feasibility study might uncover new ideas that could completely change a project's scope. It's best to make these determinations in advance, rather than to jump in and to learn that the project won't work. Conducting a feasibility study is always beneficial to the project as it gives you and other stakeholders a clear picture of the proposed project.

## 3.1.3. Analysis

**I. Class Diagram**

To model the static structure of the system, a class diagram is created. In UML, a class diagram is the static structure diagram which maps out the structure of the system by showing systems classes, their attributes, methods and finally the relationship among the objects. It is one of the

popular types of UML structure diagrams. It helps us to document the system architecture as they describe what the modeled system must possess. It can directly be mapped with objectoriented languages and hence is used widely during programming.

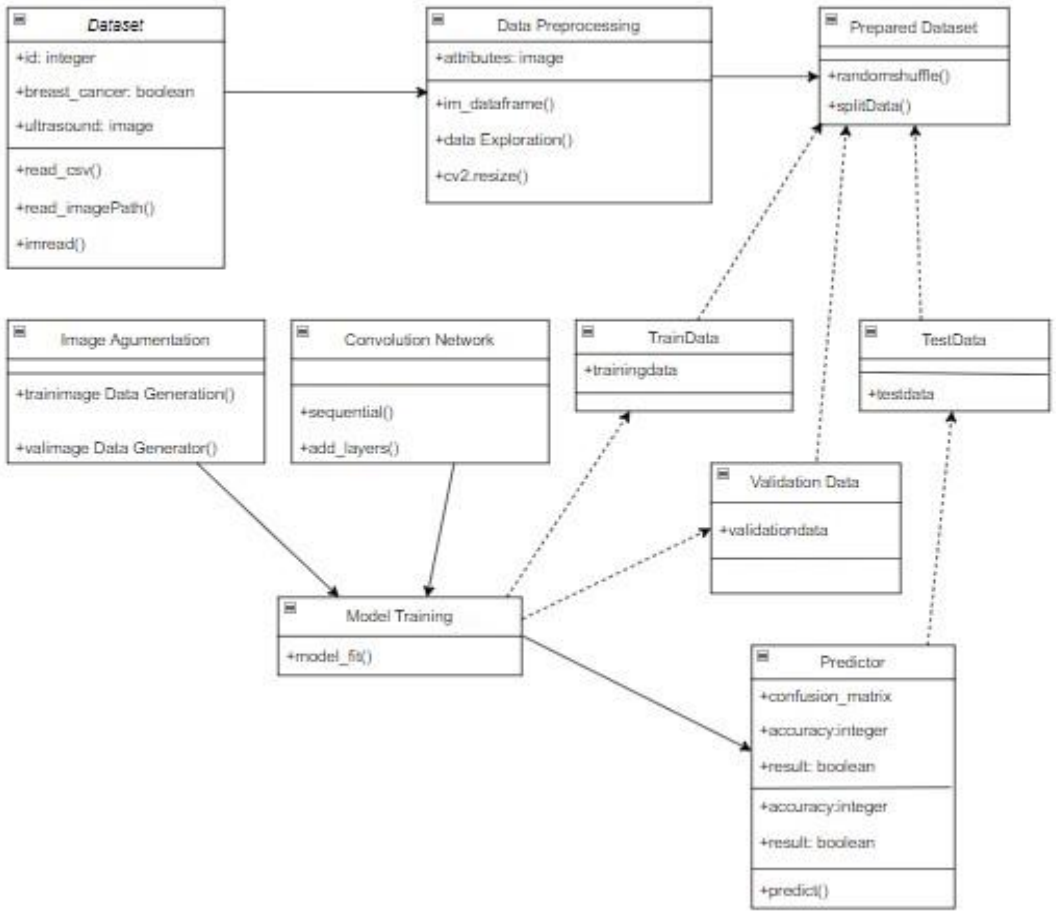



Figure 5: Object modelling using Class Diagram

**ii. Sequence Diagram**

To visualize how the operations are carried out in the system, sequence diagrams are created. It is an interaction diagram which show how and in what order, a group of objects work together. They focused on time and show the order of interaction visually. The vertical axis of the diagram is used to represent the time, what messages are sent and when. It helps to plan and understand the detail functionality of the developed or developing system. It shows

different operations at higher level of abstraction than activity diagram. It is slightly complex diagram than activity diagram. It is also known as event diagram.

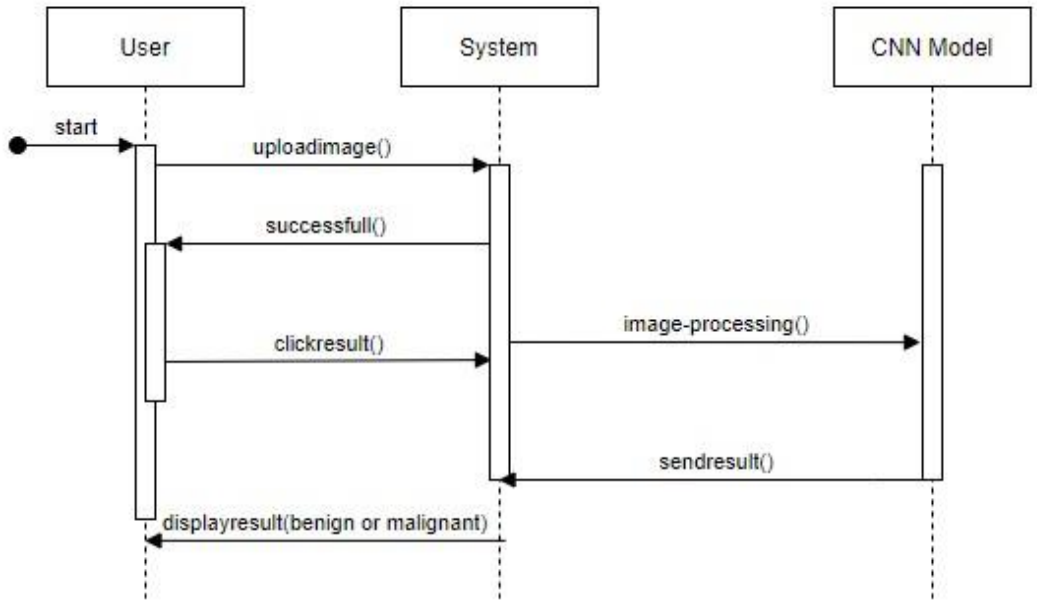



Figure 6: Sequence Diagram for Dynamic Modelling

The above diagram shows how the user upload image and system process it with the help of CNN model and it send result which is displayed as either normal or benign or malignant.

**iii. Activity Diagram**

To describe the dynamic aspects of the system, the activity diagram is created. It is the graphical representation of the workflow of the system from one activity to another activity. The logic of the operation shown in the class diagram is described by an activity diagram. The control flow which can be branched, sequential, or concurrent is drawn from one operation to another and different elements such as fork, join, etc. are used in activity diagram to deal with all type of flow control.

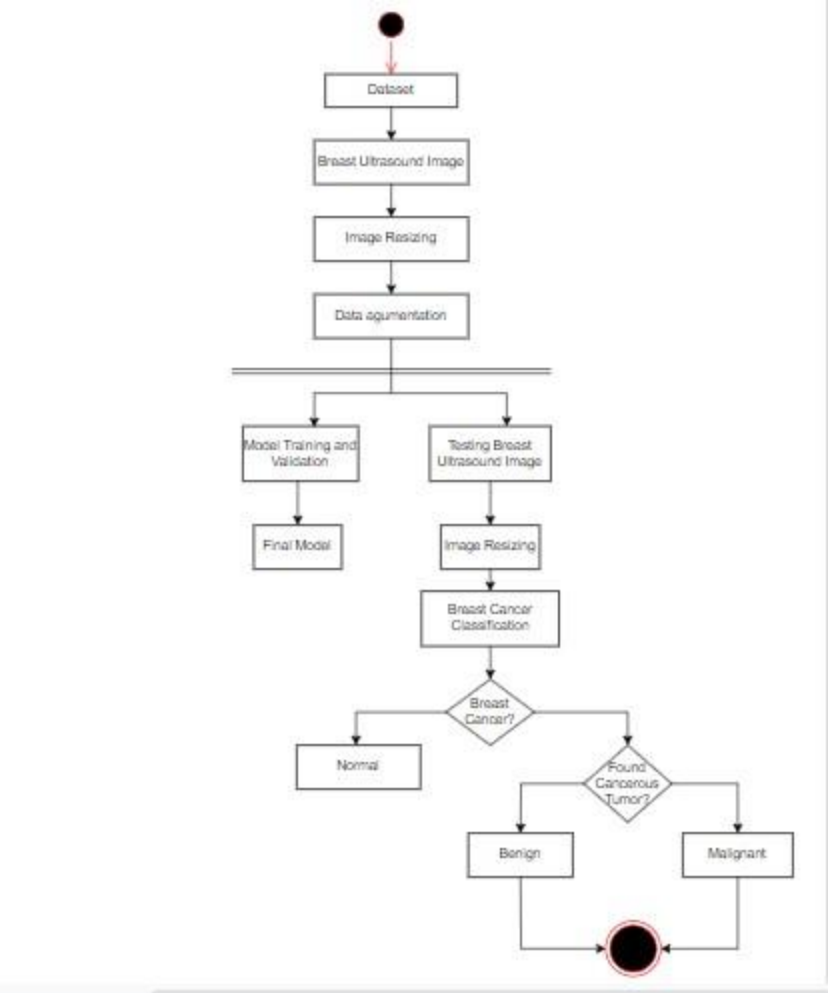



Figure 7: Activity Diagram for Process Modelling

## 3.2. Hardware Requirements

Processor : Above Dual Core

RAM : 1GB+

Hard Disk :20 GB

Input Device : Standard Keyboard and Mouse

## 3.3. Software Requirements

Operating System : Windows10, LINUX, MacOS

IDE : SPYDER, Visual Basic Code, Anaconda with Python3, Jupyter Notebook.

## 3.4. Software Development Methodology

For the completion of the project successfully, there is great role of System development methodology. For this project, Agile methodology is selected. The task of the system has been broken down into manageable units which assists to focus on high quality development, documents, and testing. Frequent testing and reviews help to improve the quality of the system. Similarly, Agile is flexible to change even in late stage which is beneficial in case of any change in requirements of our system.

# Chapter 4: System Design

## 4.1. Design Goals

Under our model, the goal of our project is to create a design to achieve the following:

### 4.1.1. Accuracy

Only accurate outcomes can help to make this model a good one. It can be reliable only when all the outcomes are correct and can be trusted. As this data is required for healthcare purposes, it is important that no errors occur.

### 4.1.2. Efficiency

The model should be efficient as there is no requirement of manual data entry work or any work by doctors. It takes less time to predict outcomes after CNN algorithms have been uses on the data.

## 4.2 Component Diagram

It depicts the general operation and other component of prediction process and represent a system physical component.

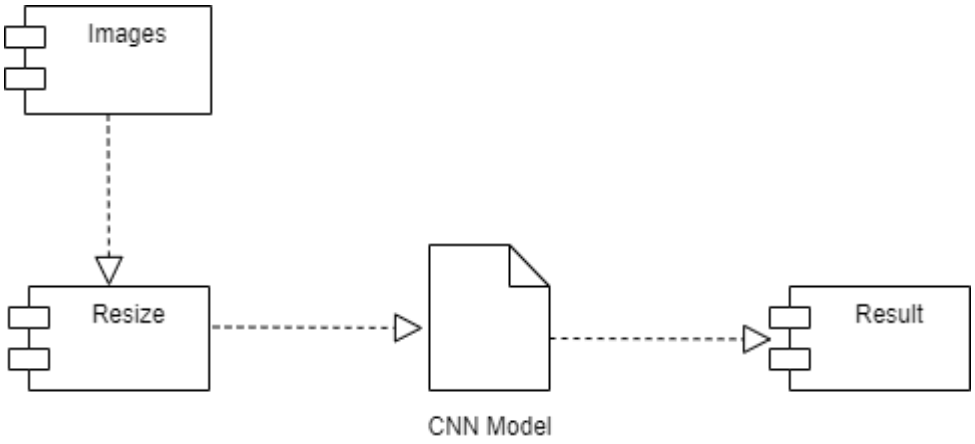



Figure 8: Component Diagram

## 4.3. Algorithms Details

The algorithm used in our system is Convolutional neural network.

**Convolutional Neural Network**

Neural network is a model inspired by how brain works. A neural network consists of neurons which is a mathematical function that gathered and classify information as per specific architecture. It includes layers of interconnected nodes which is a perceptron and is like a multi linear regression. The interconnected neurons consist of weight numeric values that are tuned to the model while training, which will act to the actual recognition or predication of the objects. The interconnected neurons respond to a neural network consisting numerous neurons obtained by a series of input weights related with the structure from previous neurons. The application of neural network is, forecasting, fraud detection, disease detection and many more.

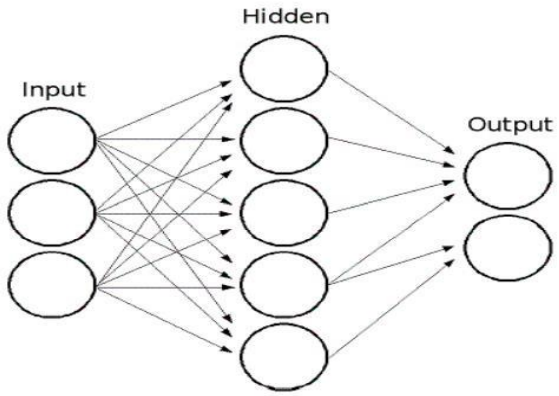



Figure 9: Component Diagram

The first layer of the input layer is called an input layer which identifies the primitive patterns, the second hidden layer identifies the first layers i.e., input layer patterns and the last and third layer called output layer identifies the second layer patterns. Convolutional Neural Network (CNN) is a deep learning algorithm that take an input image, assign learnable biases and weights to numerous objects in the image and differentiate one from the other. As comparing to other classification algorithm, the preprocessing needed in CNN is much lower. CNN are used in object detection, plant disease detection, fraud identification and many more. CNN is like the connectivity of neurons pattern in human brain and motivated by the visual cortex organization.

**Layers in CNN**

- **Convolutional Layer • Pooling Layer • Fully connected Layer**

**Convolutional Layer:** This is a first layer and a main building block of CNN which is used to extract the numerous features from the input images. The mathematical operation is executed between the pixels of input image and a filter of a size S*S. This layer contains filters and padding. In this layer, basically two problems occur, one is while reducing the size of input image the important features may be loss and another is increase of the same dimension. This problem is solved by using padding. There are 2 types of padding, they are valid padding and zero padding. Each filter will be moved from the output's top-left corner, but only 1 filter is move at a time. After that achieved, the filter moves one position downward and the filters strides one position at a time from left to right. This way, it is repeated until the filter reach to

the bottom right. The result obtained from this process is called filter map or feature map. Then the ReLu activation function is used to increase the non-linearity in our images.[5]

**Pooling Layer:** The second layer of the CNN is pooling layer which is used to decrease the resolutions or dimensions of the features map obtained from convolution layer. It decreases the number of parameters to learn, and computation power performed in network.

There are 2 types of pooling layers, they are:

- **Average pooling:** It calculates the average value of each selected box on features map.
- **Max pooling:** It calculates the maximum value of selected box on features map. (Jason, 2019)

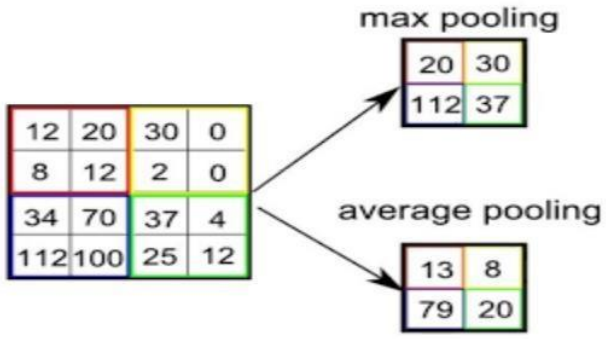



Figure 10: Types of pooling layer

**Fully connected Layer:** The last layer of the CNN is called fully connected layer. It is a simple feed forward neural network. The input to the last layer is the output from last pooling layer, which is flattened and then move into fully connected layer. The process of changing dimensional matrix into vector is called flatten

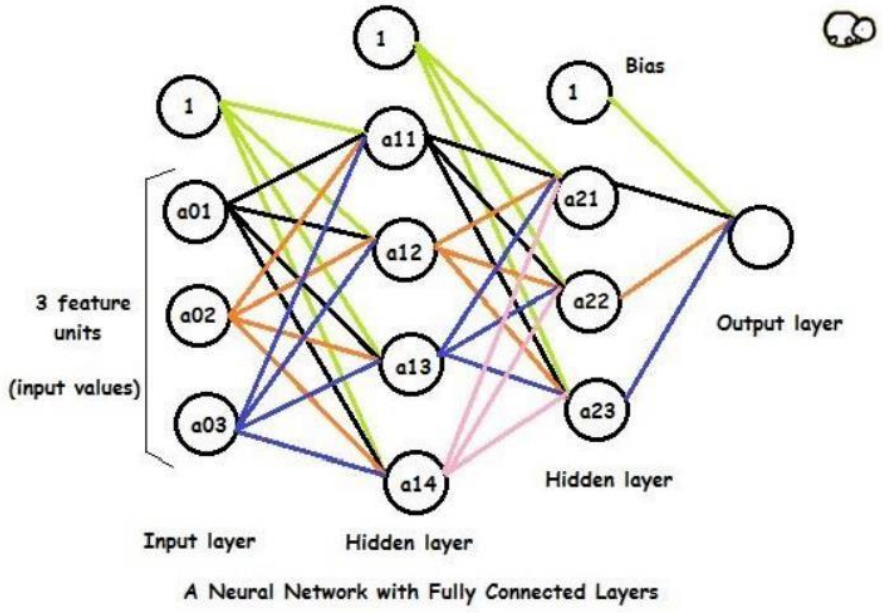



Figure 11: Fully connected layer

From the above figure, a01, a02, a03 are the three input features in the network. These input features assigned with weight and bias and sent to hidden layer. In the hidden layer, the summation of inputs features, weight and bias will perform. Then we put activation function in the summation i.e., f (wa+b), while a is the input, w is the weight and b is the bias. The activation function used in our system is ReLu. The output from the 1st hidden layer is send to next layer assigned with weight and bias. Then this process will continue in hidden layers. In output layer we used SoftMax activation function because it converts the output values in normalized probability distribution. If the error rate is maximum in output layer, then the network will back propagate until the global minimum value is found. The process of finding global minima value is called gradient descent.

**Activation Functions**

They are mathematical equations which determines the output of a neural network model. It takes decision whether a neuron should be fired or not by performing the calculation of weighted sum and bias with it. The aim of activation function is to determine non-linearity into the output of neurons. Some of the activation functions are:

**Hyperbolic Tangent function**

This function usually used in classification between 2 classes which values lies between -1 to 1.

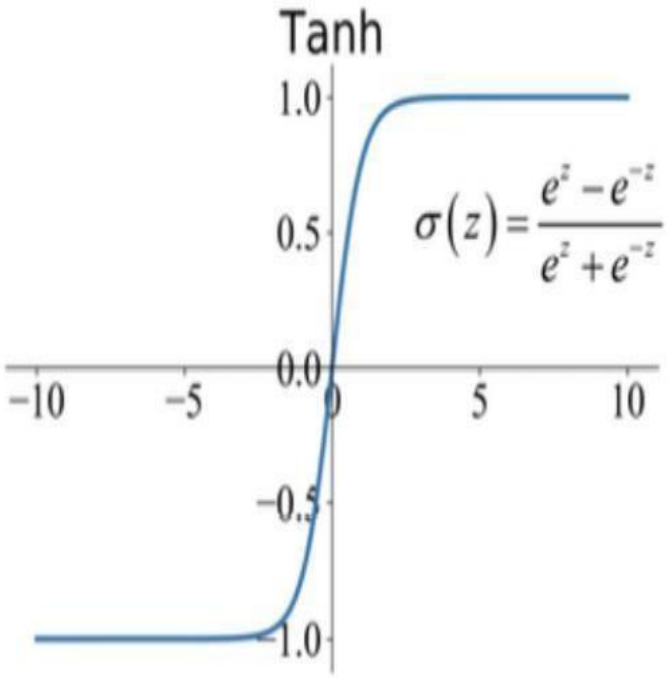



Figure 12: Tanh Graph

**Sigmoid:** It is curve look like an S-shape and its value range between 0 and 1. It is usually used for binary classification.

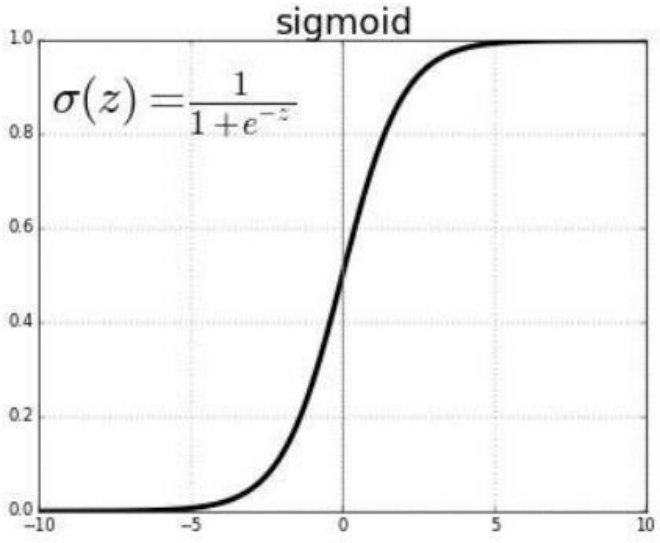



Figure 13: Sigmoid graph

**SoftMax:** It is used in an output layer of neural network which is used for multi-class classifications problems. It turns a vector of real values into vector of real values that sum to 1. The input variable can be negative, zero, or greater than one but the SoftMax function transform those values between 0 and 1. [11]

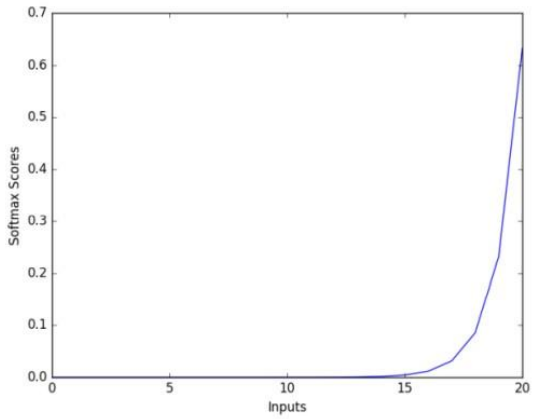



Figure 14: SoftMax graph

**ReLU:** We have used ReLu function in hidden layer which is used to increase the nonlinearity in our images. This function outputs the input values directly, it is positive values, otherwise, it will output zero. [10]

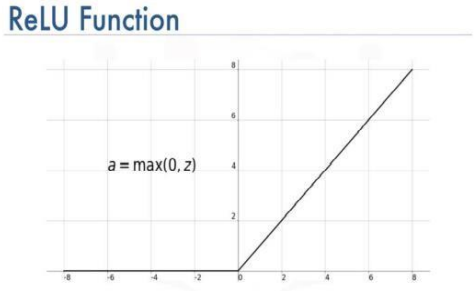



Figure 15: ReLU graph

# Chapter 5: Implementation and testing

Implementing and Testing is executing a system to identify and fixed any gaps, errors, or missing requirements in contrary to the actual requirements. Software system testing is a process of verifying and validating a software application or program.

## 5.1 Implementation

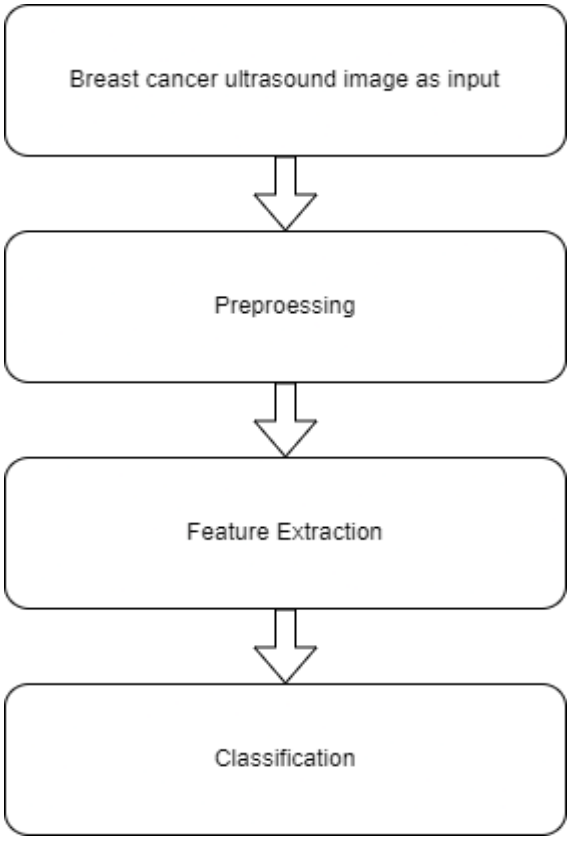



Figure 16: Implementation model for classification of breast cancer

First, input from user were taken as ultrasound breast cancer dataset as an image and then data preprocess was done by using preprocessing techniques. Then feature is extracted important from the given. Finally, CNN Algorithm was used to classify the image into corresponding class of Breast Cancer that is Benign and Malignant.

## 5.1. Tools Used

### i. HTML

HTML is the standard markup language for creating web pages. It was used to describe the structure of web pages using markup and to display the output of the machine learning model.

### ii. CSS

CSS was used in developing custom styles for the web application. It describes how HTML elements are to be displayed on a webpage.

### iii. Bootstrap

Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first frontend web development. It contains HTML, CSS and JavaScript-based design templates for typography, forms, buttons, navigation, and other interface components.

### iv. JavaScript

JavaScript was used for handling client-side user input and implementing AJAX features so that the web application does not have to reload for each user input.

### v. Jupyter Notebook

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

### vi. Spyder

Spyder is an open-source cross-platform integrated development environment for scientific programming in the Python language.

### vii. Flask

Flask is a micro web framework written in python. It provides the most flexibility in terms of customization and provides simplicity and fine-grained control.

### 5.1.1. Implementation Details of Modules

The implementation phase of the project has by far been the most challenging of the phases of this project.

**Home Page for choosing file**

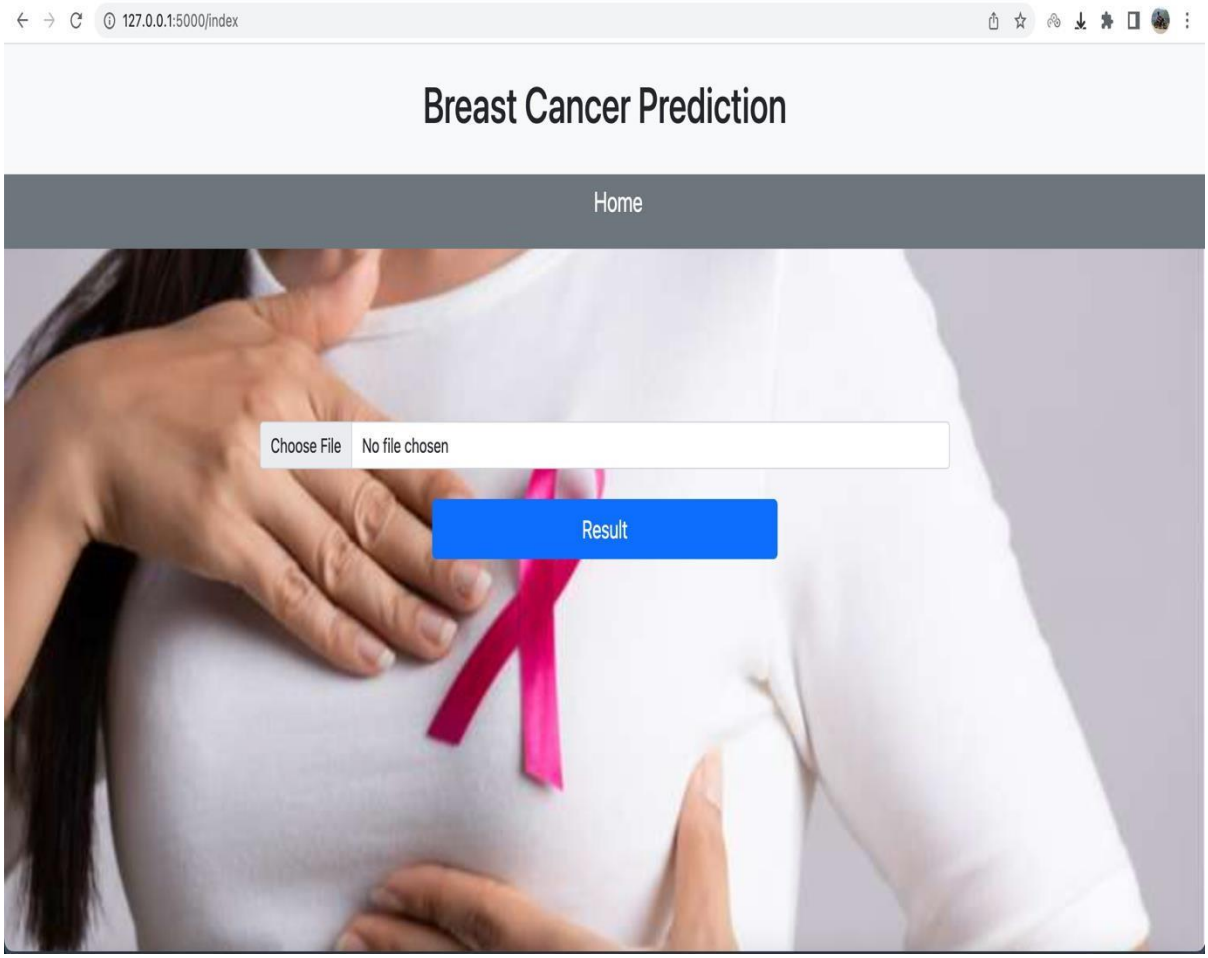



Figure 17: Home page for choosing file

**Choosing file**

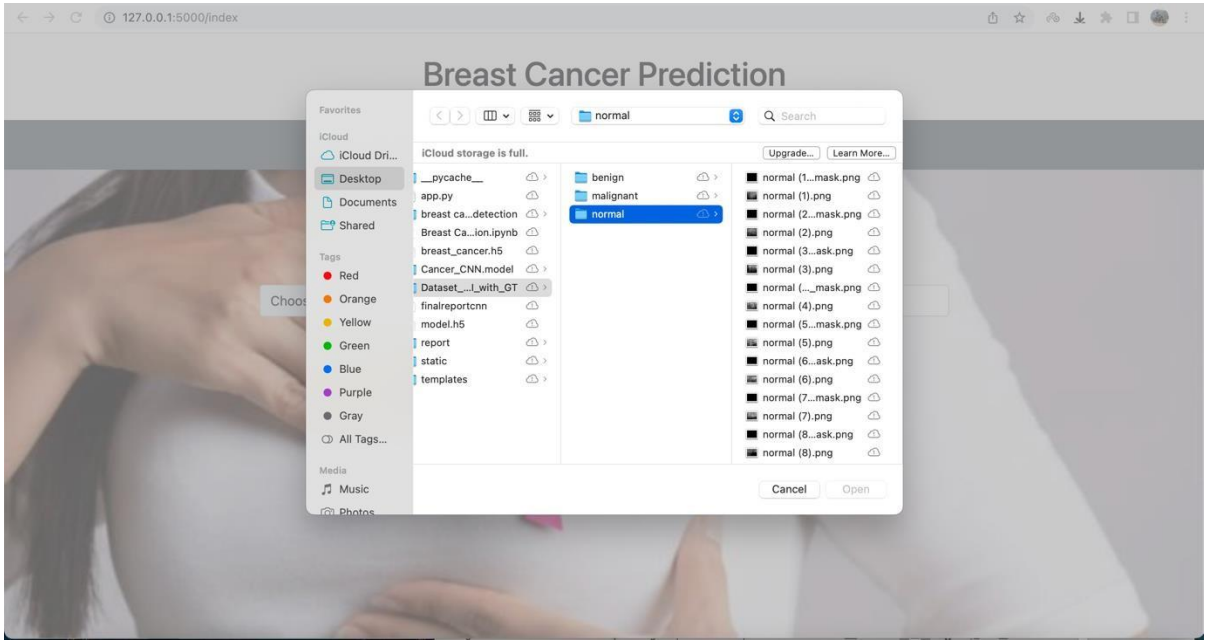



Figure 18: Choosing file **Files**

**choose**

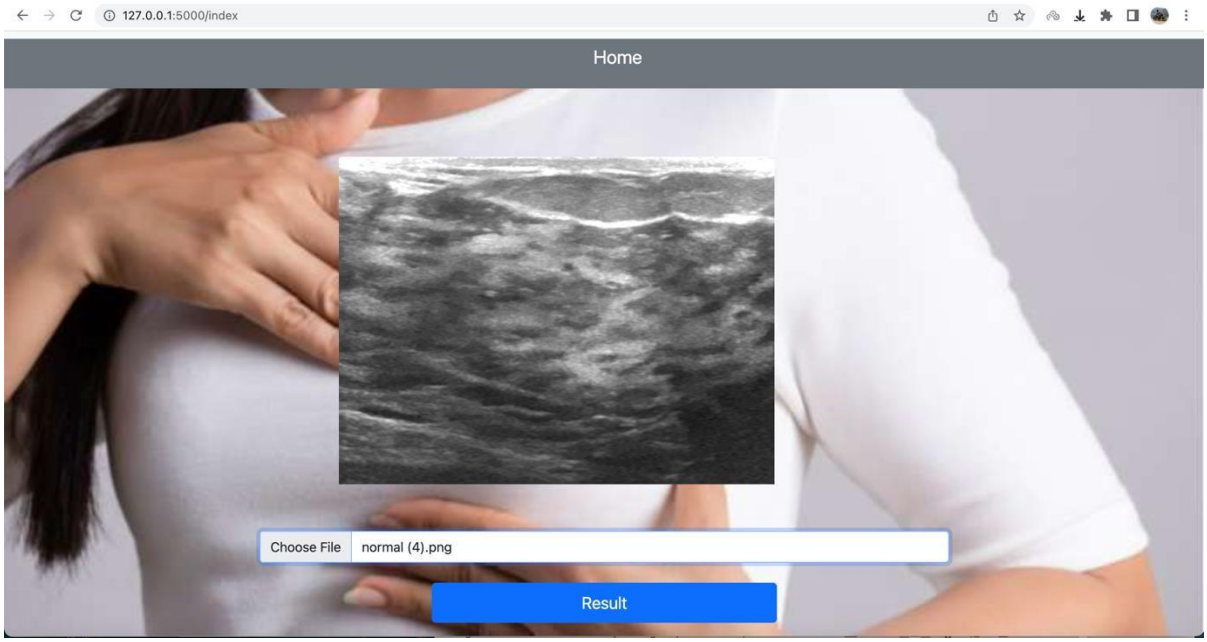



*Figure 19: File choose*

**Prediction**

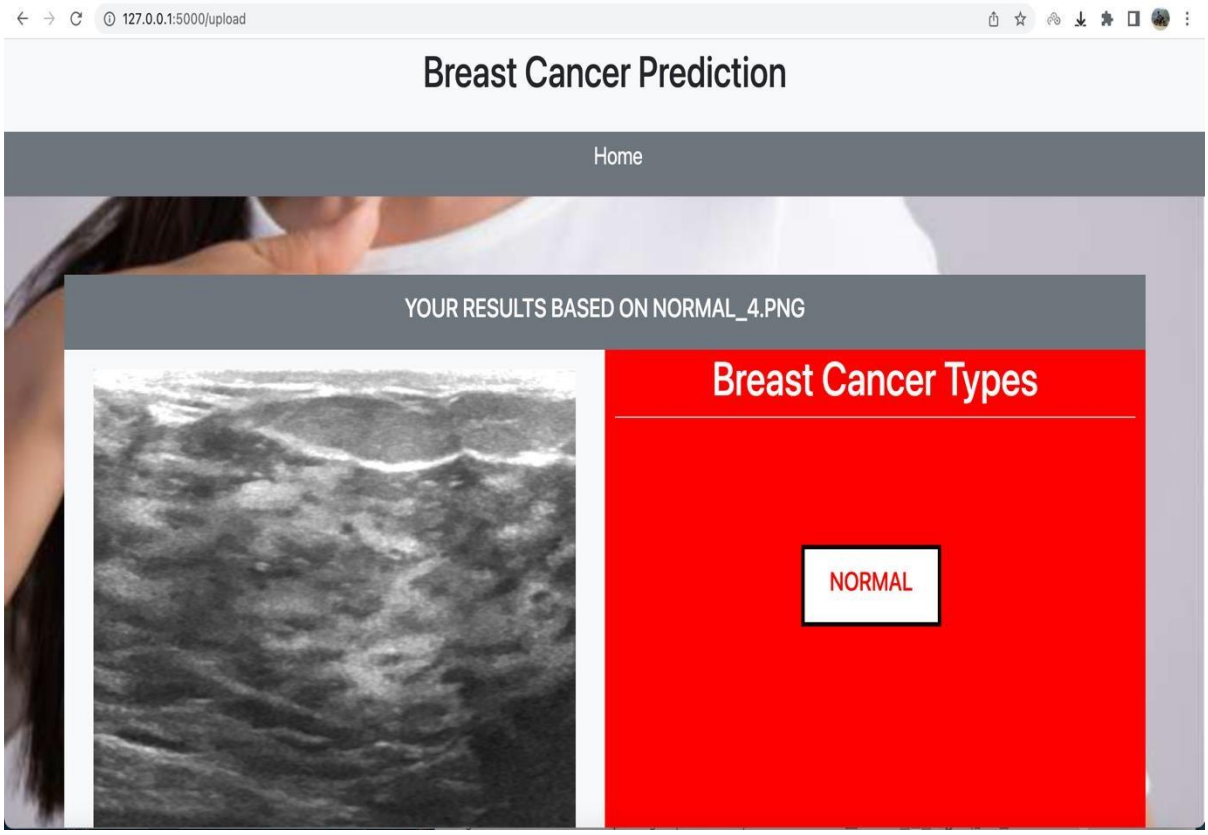



Figure 20: Prediction is done

## 5.2. Testing

The testing phase depicts the efficiency and inefficiency of the system.

### 5.2.1 Unit testing

Unit testing refers to the testing of every small modular component of the system, keeping them isolated from other modules. Here we mention testing result of the various part of the system. In unit testing, we design the whole system in modularized pattern and each module was tested. Till we get the accurate output from the individual module we have worked on the same module. We have checked for the outcome of each module. We use every unitary module of our system.

i. Test case 1: Upload image ii.
Test Case 2: Detection of Cancer

Table 1: Test case for uploading Image

| | |
|---|---|
| Case 1 | Successful upload of images in the system. |
| Expected result | User should be able to upload image in the file choose section |
| Actual result | Image Uploaded successfully. |
| Conclusion | Successfully |

Table 2: Test case for detecting cancer

| | |
|---|---|
| Case 2 | Detection of the cancer based on image provided. |
| Expected result | System should detect the cancer from the image provided and redirect to result page. |
| Actual result | Breast cancer can either be normal or benign or malignant is predicted and redirect to the result page. |
| Conclusion | Successfully |

### 5.2.2 System Testing

In this testing phase our system was tested. Every individual component was integrated and tested against user and hardware compatibility. Sometime in this testing process dependency error was found due to different local server environment and dependency conflict. To overcome this problem virtual environment and docker container were used. Table 3: System Testing

| S. N | Test Description | Expected result | Actual Result | Remarks |
|---|---|---|---|---|
| 1. | Test image benign (1) | Benign | Benign | Pass |
| 2. | Test image malignant (1) | Malignant | Malignant | Pass |
| 3. | Test image normal (1) | Normal | Benign | Fail |
| 4. | Test image normal (4) | Normal | Normal | Pass |

## 5.3. Result Analysis

Studying several ICT in medical projects around the globe, it is seen that the public in those countries were facing the same problems as the people of Nepal were facing. With the use of

ICT tools, i.e., web-based technology and internet it was possible to overcome the problems of breast cancer especially in rural areas. The research papers, case studies and journals explained that the cure of patients could be enhanced with the use of ICT in the medical sector. In order to detect the presence or absence of breast cancer through mammographic image in both male and female, Convolutional Neural Network was used. While training the model, the train loss and validation loss, train accuracy and validation accuracy in graph was compared. Similarly, confusion matrix was also used to evaluate the performance of the model.

### i. Training and validation accuracy

```
print(history.history.keys())
# summarize history for accuracy
plt.plot(history.history['accuracy'])
plt.plot(history.history['val_accuracy'])
plt.title('model accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
# summarize history for loss
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```

```
dict_keys(['loss', 'accuracy', 'val_loss', 'val_accuracy'])
```

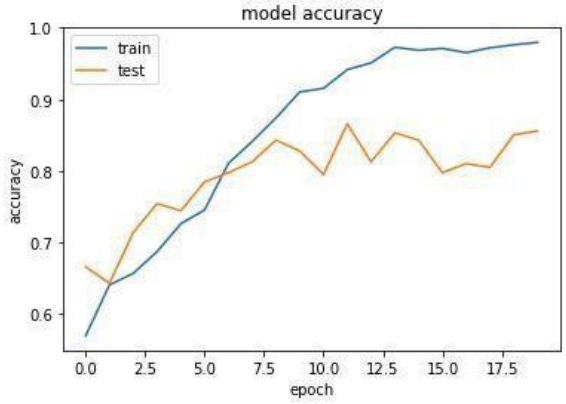


Figure 21: Graph of train and val accuracy

## ii. Training and validation loss

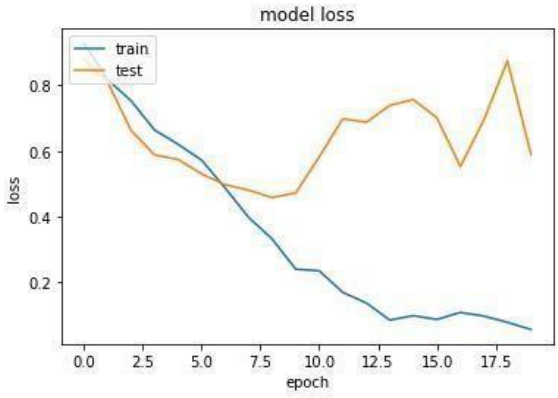



Figure 22: Graph of train and val loss

### iii. Confusion Matrix

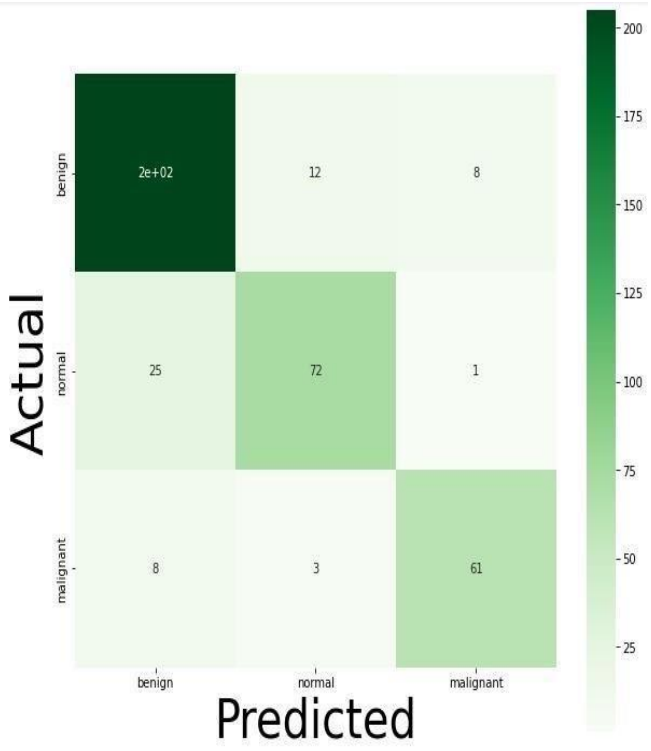



Figure 23: Confusion Matrix

# Chapter 6: Conclusion and Recommendation

## 6.1. Conclusion

The objective of building this project was to detect the ultrasound images or the mammographic images capable of automatically classifying categories based on cancerous tissue or not, implementing the Convolutional Neutral Network (CNN). At the end of developing phase, the system was able to meet the set objectives with satisfying accuracy and speed.

Thus, using various concepts of Deep Learning and with related tools, complex tasks of detecting Breast Cancer was invented.

## 6.2. Future Recommendations

Presently, this system is limited to identify either the women have breast cancer or not. If yes, system classify the cancer into malignant or benign. This being our initial phase for the development of machine learning model and implementing in the health sector specially focused in rural areas.

The application performs satisfactorily, however, there is always a place for future enhancement. Various improvements in both components, as well as performance, can be met. Some of them can be given as:

i. Refine the Training Datasets ii.

Increase in the number of Datasets iii.

User can directly input their symptoms.

Also, all the points as given in the limitations are intended to be implemented as a future enhancement for this project.

# REFERENCES


[1] M. G. Kanojia and S. Abraham, “Breast cancer detection using RBF neural network,” in 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Dec. 2016, pp. 363–368. doi: 10.1109/IC3I.2016.7917990.

[2] S. Das, “Detecting Breast Cancer using Neural Networks,” Int. J. Res. Appl. Sci. Eng.Technol., vol. 8, no. 5, pp. 1157–1163, May 2020, doi: 10.22214/ijraset.2020.5183.

[3] V.Dheeba, N.Albert Singh, and J.Amar Pratap Singh, ”Breast cancer Diagnosis: An Intelligent Detection System Using Wavelet Neural Network,” 2014, pp. 111–118. doi: 10.1007/978-3-319-02931-3_14.

[4] H. Zhou, Y. Zaninovich, and C. Gregory, “Mammogram Classification Using Convolutional Neural Networks”, [Online]. Available: http://ehnree.github.io/documents/papers/mammogram_conv_net.pdf

[5] C.-M. Lo et al., “Multi-Dimensional Tumor Detection in Automated Whole Breast Ultrasound Using Topographic Watershed,” IEEE Trans. Med. Imaging, vol. 33, no. 7, pp. 1503–1511, Jul. 2014, doi: 10.1109/TMI.2014.2315206.

[6] Chuyang Ye, V. Vaidya, and Fei Zhao, “Improved mass detection in 3D automated breast ultrasound using region based features and multi-view information,” in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 2014, pp. 2865–2868. doi: 10.1109/EMBC.2014.6944221.

[7] C. Muramatsu, Y. Hiramatsu, H. Fujita, and H. Kobayashi, “Mass detection on automated breast ultrasound volume scans using convolutional neural network,” in 2018 International Workshop on Advanced Image Technology (IWAIT), Jan. 2018, pp. 1–2. doi: 10.1109/IWAIT.2018.8369795.

[8] D. Dumitru, “Prediction of recurrent events in breast cancer using the Naive Bayesian classification”, [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.818.204&rep=rep1&type=pdf

[9] S. Aruna and D. S. P. R. and L. V. Nandakishore, “KNOWLEDGE BASED ANALYSIS OF VARIOUS STATISTICAL TOOLS IN DETECTING BREAST CANCER”, doi: 10.5121/csit.2011.1205.

[10] S. SHARMA, “Activation Functions in Neural Networks”, [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

[13] J. Brownlee, “Softmax Activation Function with Python”, [Online]. Available: https://machinelearningmastery.com/softmax-activation-function-with-python/

# Annex I: Snapshots

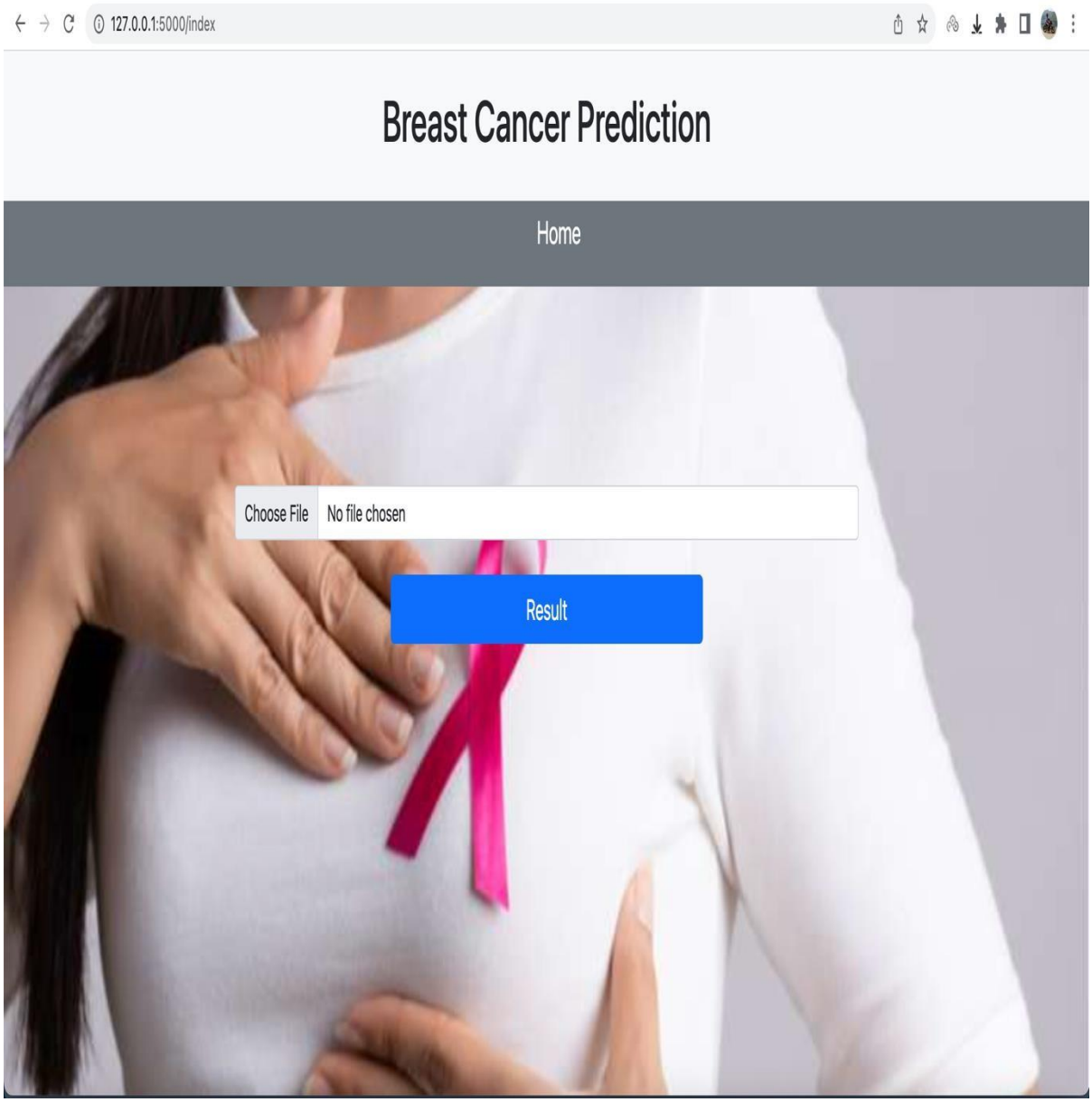



*Figure 24: Homepage*

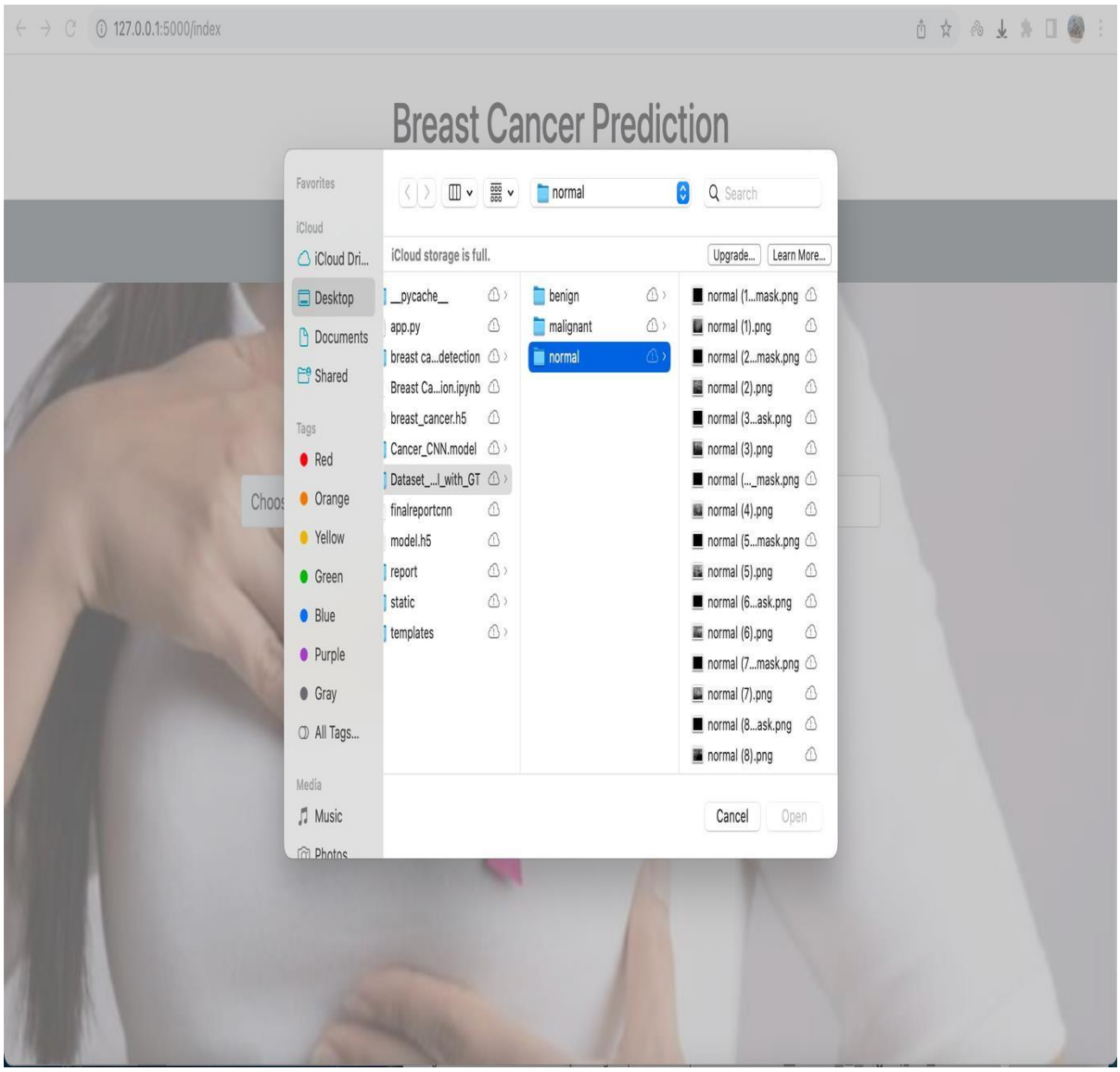



Figure 25: Image Upload Page

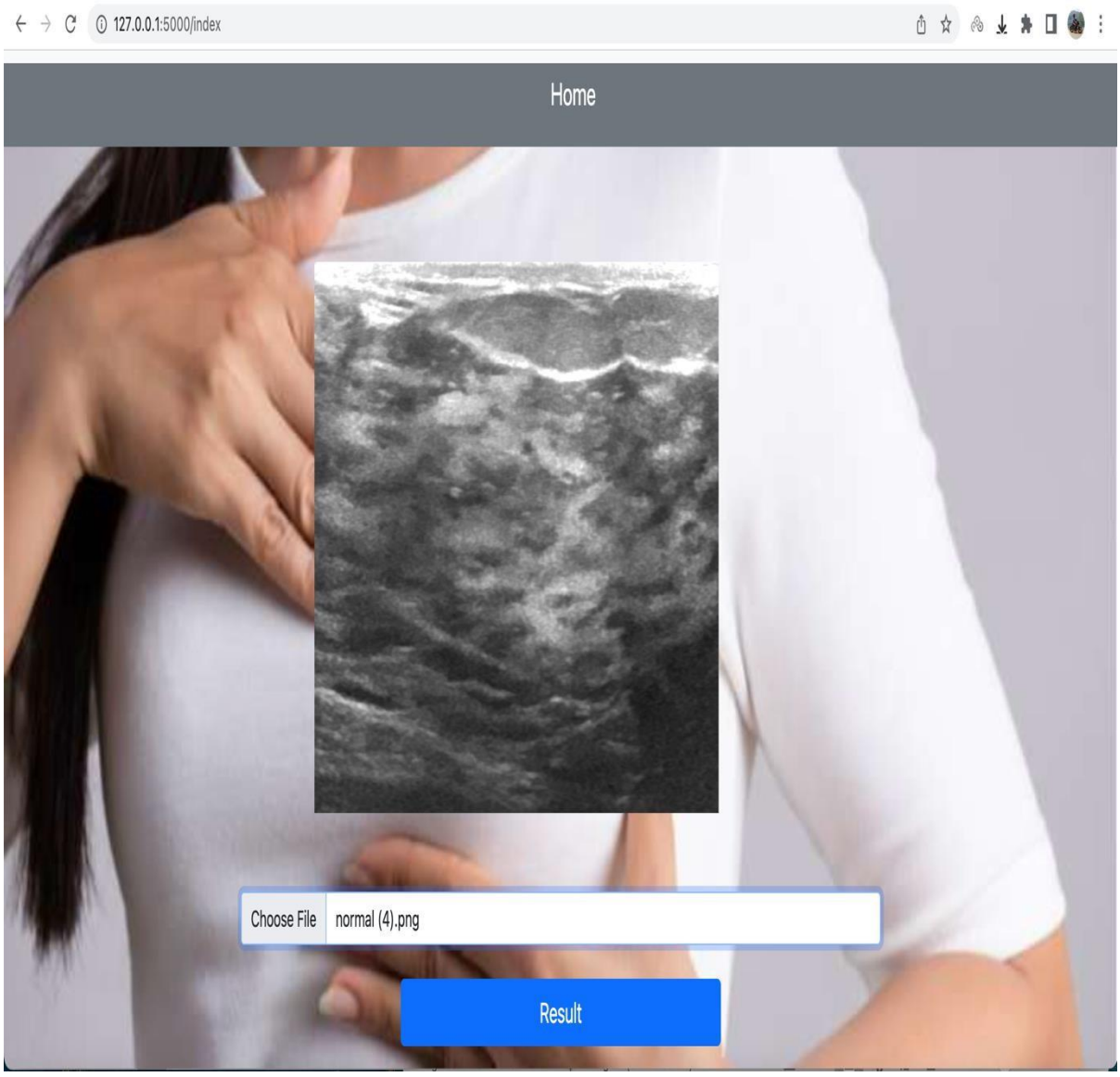



Figure 26: Image Uploaded Page

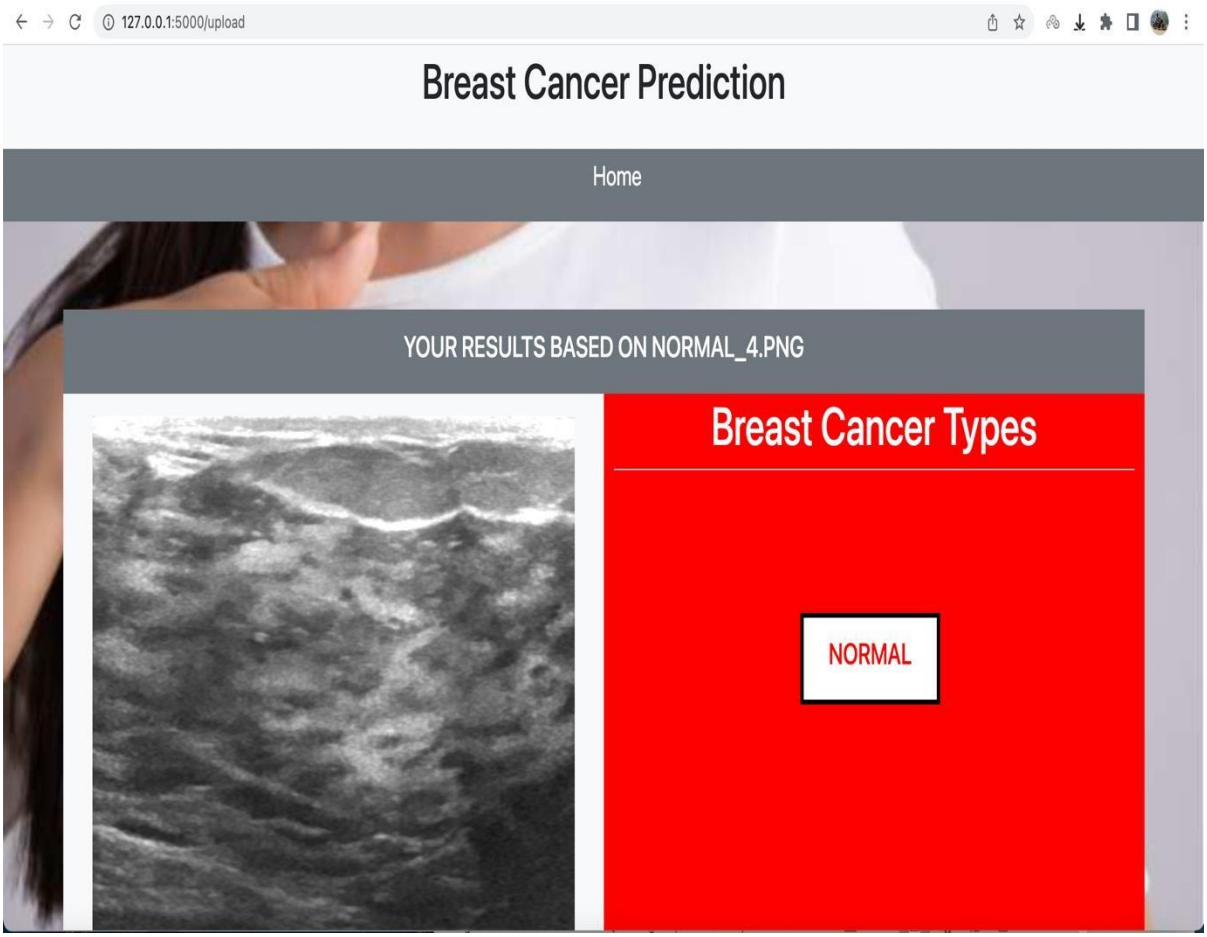



*Figure 27: Prediction*

# Annex II. Source Code

**APP.PY Code**

app.py

C: > Users > ASHMITA > Downloads > Breast Cancer Detection > app.py

```python
from flask import Flask, render_template, request, jsonify

import tensorflow as tf
from keras.models import model_from_json
from keras.applications.mobilenet_v2 import preprocess_input

from PIL import Image
from io import BytesIO
from tensorflow.keras.utils import img_to_array
import numpy as np

app = Flask(__name__)
model = None
graph = tf.compat.v1.get_default_graph()

def load_request_image(image):
    image = Image.open(BytesIO(image))
    if image.mode != "RGB":
        image = image.convert("RGB")
    image = image.resize((48, 48))
    image = img_to_array(image)
    image = preprocess_input(image)
    image = np.expand_dims(image, axis=0)

    return image

def load_model():
    json_file = open('./model/model.json', 'r')
    model_json = json_file.read()
```

Figure 28: App.py

**Frontend Source Code Base.html**

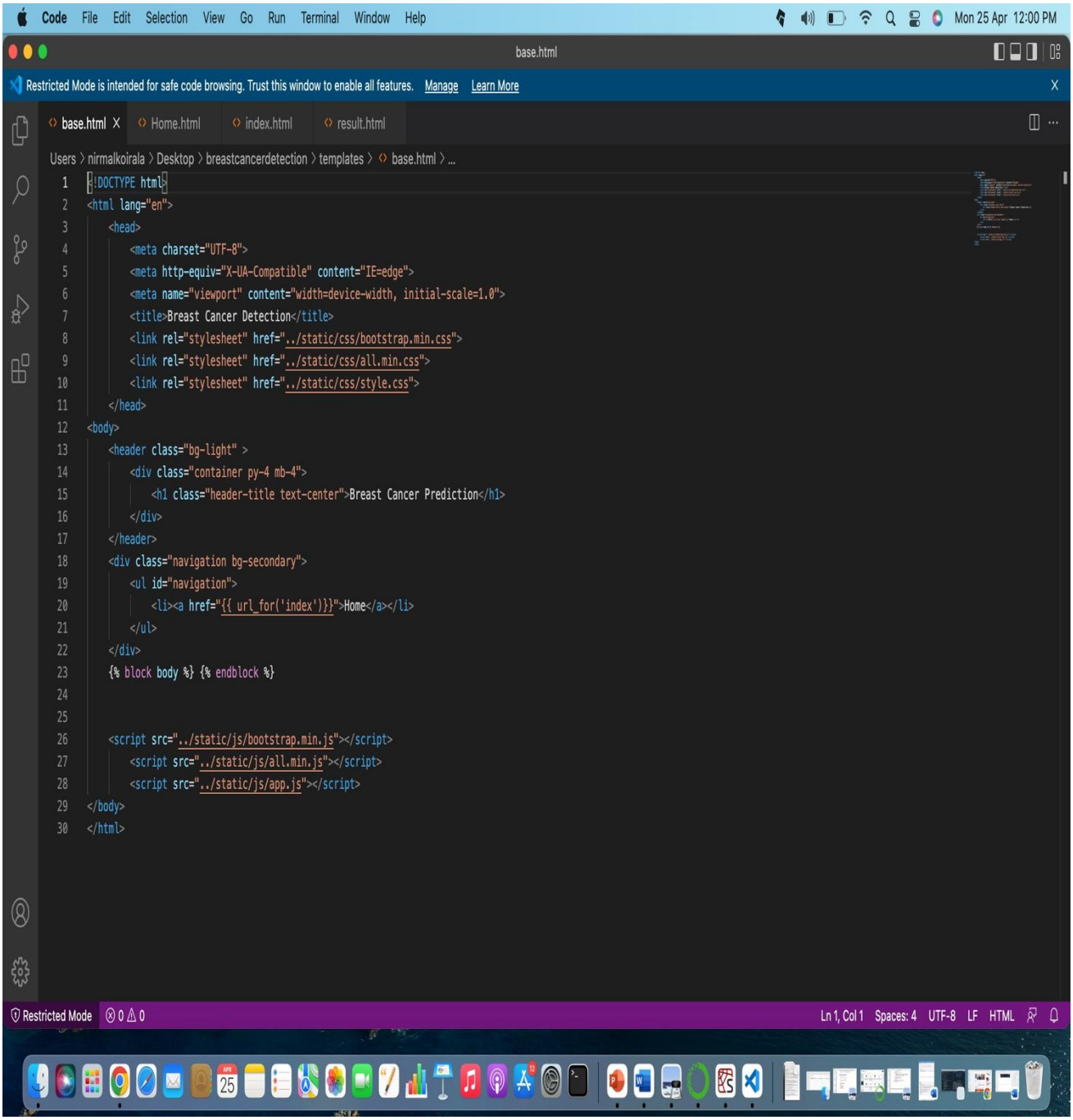



Figure 29: Base.html

**Home.html**

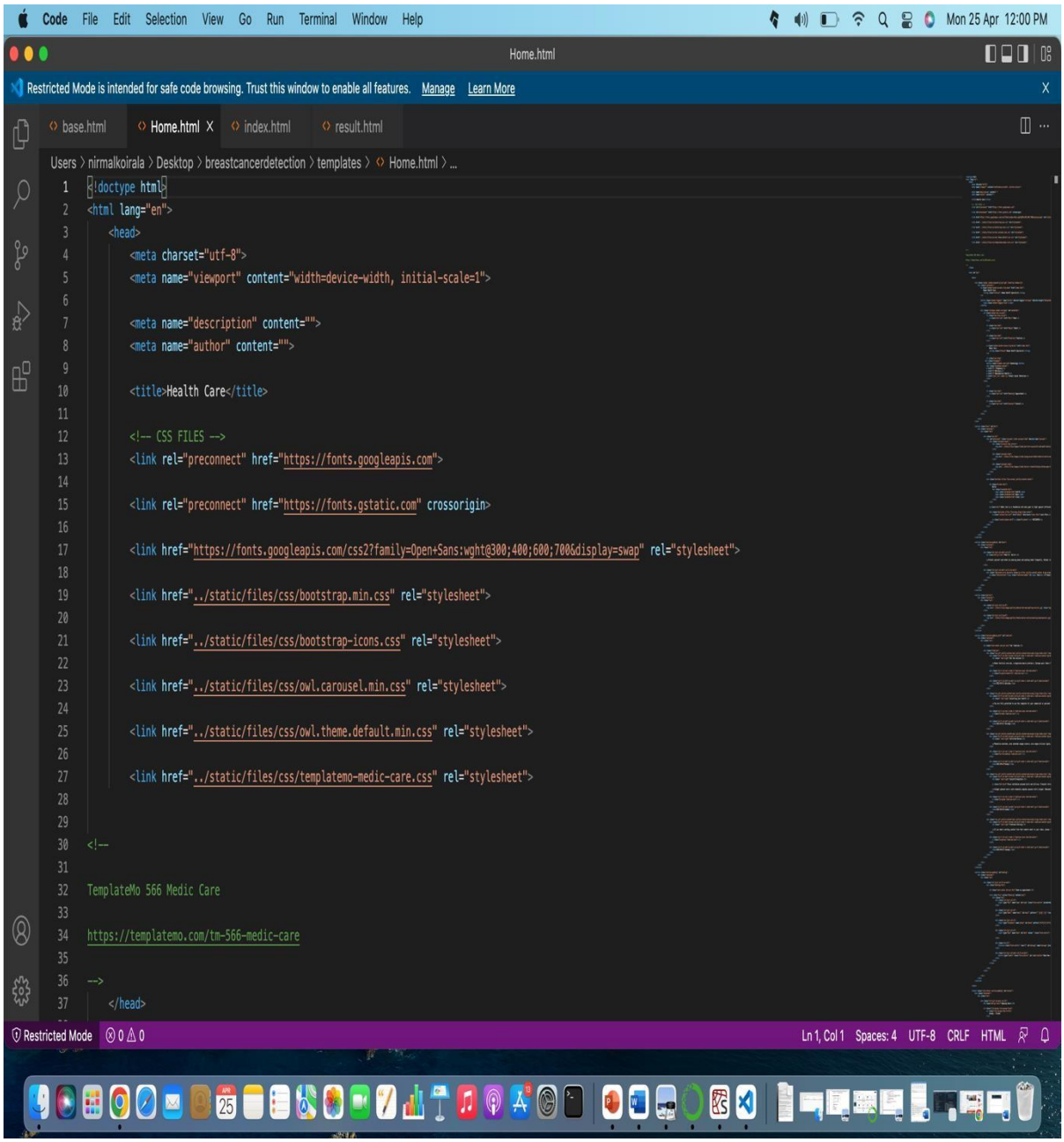



Figure 30: Home.html

**index.html**

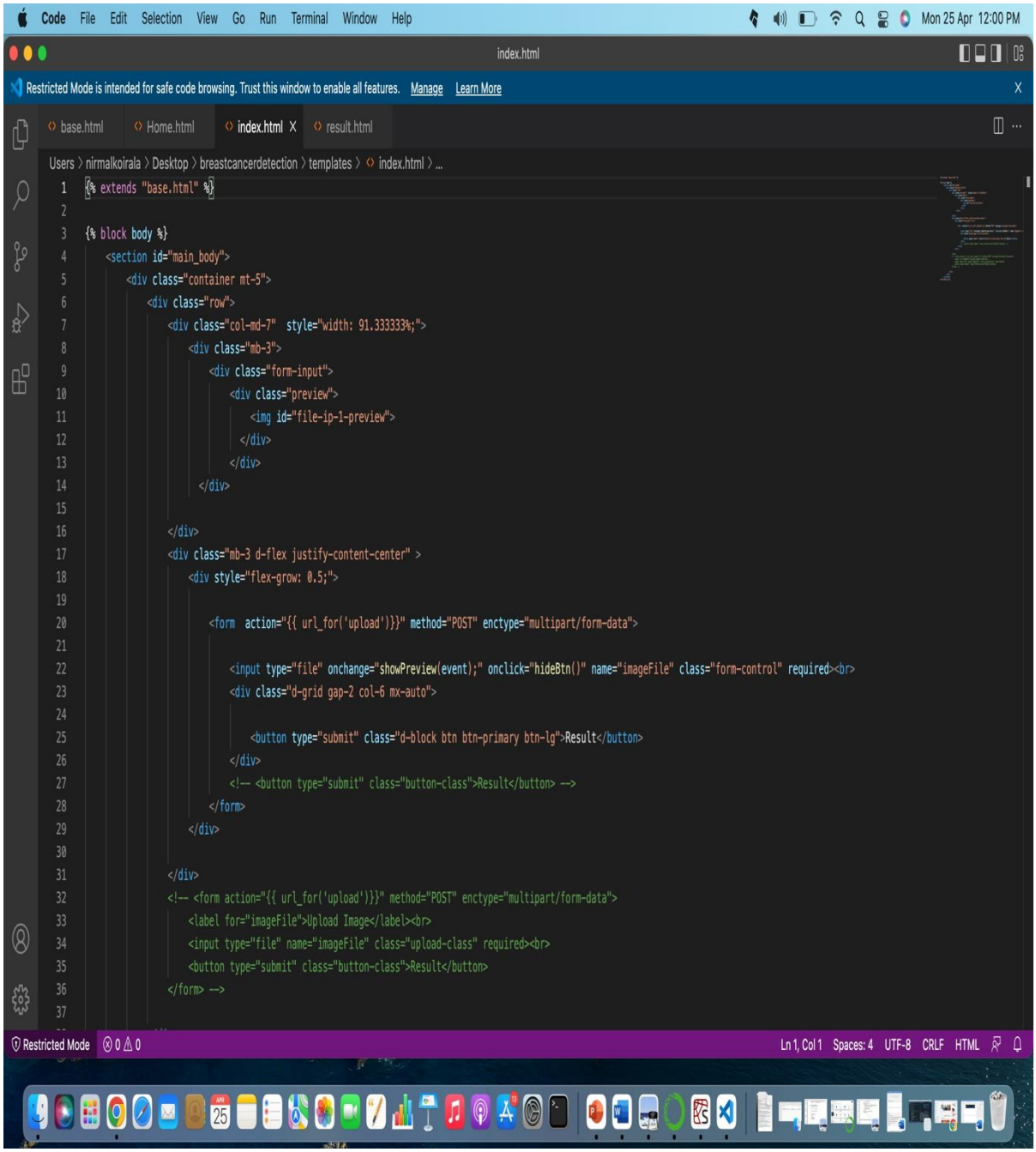



Figure 31: Index.html

**result.html**

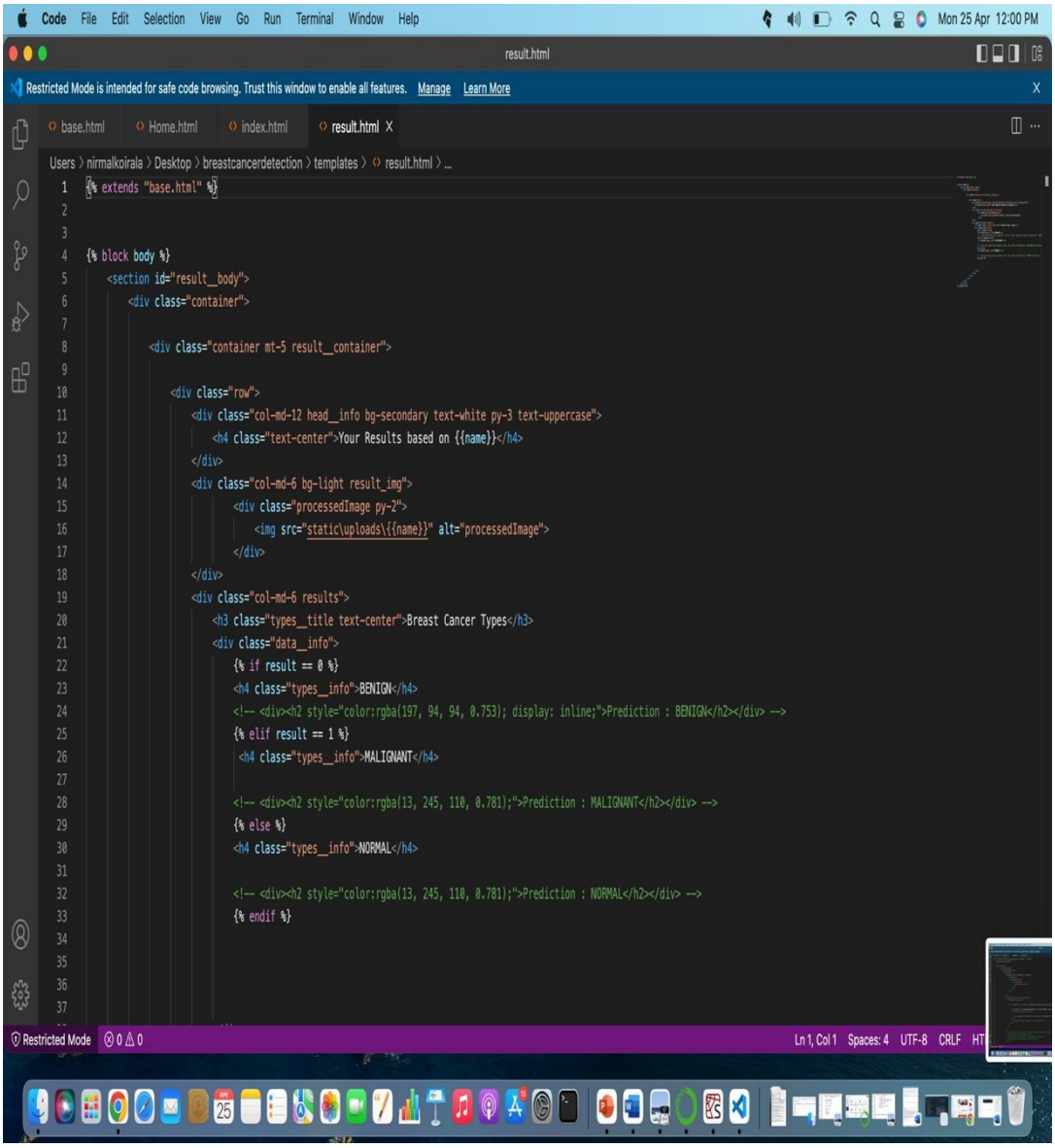



Figure 32 : result.html

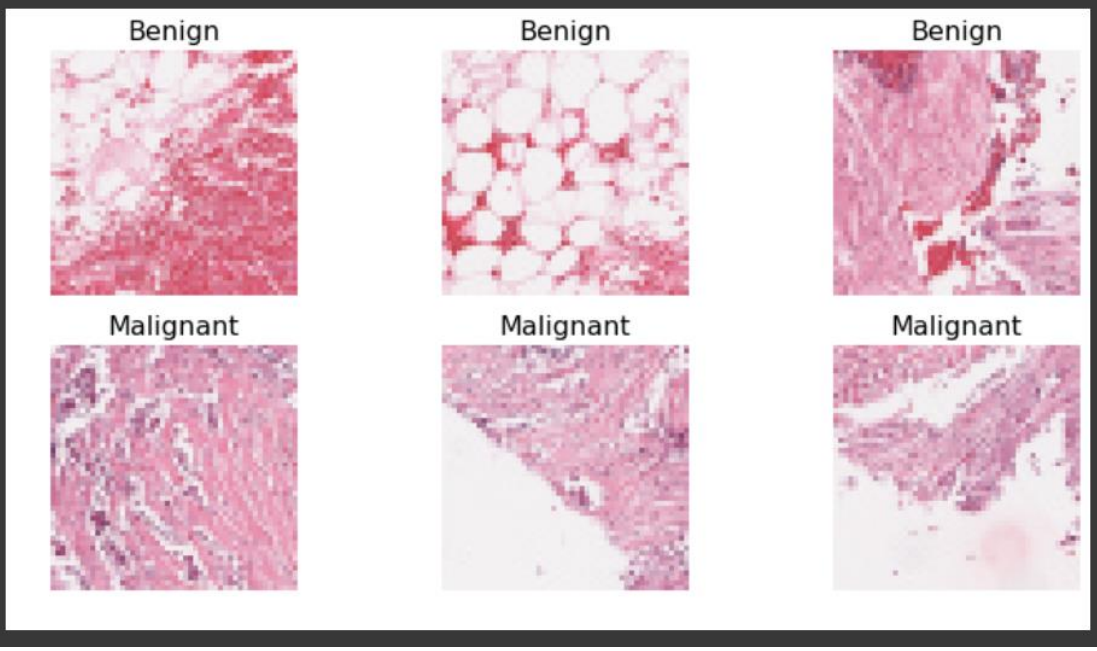
**Data Train Code**

```
+ Code  + Text                                                    T4  Disk
    sp.set_title(title, fontsize=16)
    plt.imshow(image_loaded, interpolation=None)

figure = plt.figure(figsize=(12, 6))

for image_index, image_name in enumerate(benign_images):
    plot_images(image_index, benign_images_first_folder, image_name, "Benign")

for image_index, image_name in enumerate(malignant_images):
    plot_images(image_index + 3, malignant_images_first_folder, image_name, "Malignant")
```



```
plot_validation_training("loss", trained_model)
```

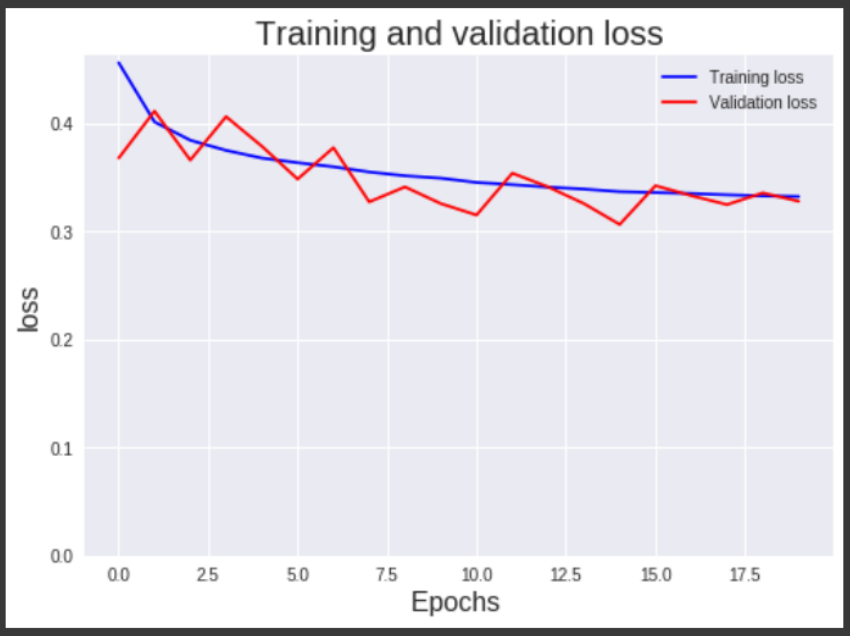



Figure 33: Data train code