Faculty of Computer Science
and Information Technology

Breast Cancer Predication Model

Supervised by:

Dr. Mohamed Mamdouh Eissa

Assistant Professor Software Engineering Dept. Faculty of Computer Science and Information Technology

Breast Cancer Predication Model

Submitted by:

#	ID	Name
1	41810083	Ahmed Hamdy Abdallah
2	41810155	Ahmed Karem Farouk
3	41810038	Doaa Mamdouh Hussien
4	41810051	Yasmeen Ayman Mohamed
5	41810178	Maram Muhamed
6	41810283	Youssef Sayed Mohamed

A dissertation submitted in partial fulfillment of the requirements for the degree of Bachelor of computer science and information technology

Supervised by:

Dr. Mohamed Mamdouh Eissa

Egypt 2022

Committee Report

We certify we have read this graduation project report as examining committee, examine the student in its content and that in our opinion it its adequate as a project document for “ ”.

Chairman:	***Supervisor:***
Name:	Name:
Signature:	Signature:
Date: / /201	Date: / /201

Examiner:

Name:

Signature:

Date: / /201

Intellectual Property Right Declaration

This is to declare that the work under the supervision of **Dr. Mohamed Mamdouh** having title "**Breast cancer predication model**" carried out in partial fulfillment of the requirements of Bachelor of Science in Computer Science is the sole property of Ahram Canadian University and the respective supervisor. It is protected under the intellectual property right laws and conventions. It can only be considered/ used for purposes like extension for further enhancement, product development, adoption for commercial/organizational usage, etc. with the permission of the University and respective supervisor. This above statement applies to all students and faculty members.

Names:

#	ID	Name
1	**41810083**	**Ahmed Hamdy Abdallah**
2	**41810155**	**Ahmed Karem Farouk**
3	**41810038**	**Doaa Mamdouh Hussien**
4	**41810051**	**Yasmeen Ayman Mohamed**
5	**41810178**	**Maram Muhamed**
6	**41810283**	**Youssef Sayed Mohamed**

Supervisor: **Dr. Mohamed Mamdouh Eissa**

Acknowledgement

Thanks for ALLAH for his generous blesses and gifts to us and for helping me to accomplish this Project.

Many thanks to Dr. Mohammed Mamdouh Eissa for giving us the wonderful opportunity to complete our Bachelor of Science in Computer Science thesis under his supervision, it is truly an honor. Thank you for all the advice, ideas, moral support and patience in guiding us through this project. Thank you for your enthusiasm for the encourage us. Your wealth of knowledge in the field of Data Science in particular is inspiring.

Many thanks to Our parents and family for continuing to be part of our strong foundation, their constant support and never-ending encouragement.

Table of Contents

CHAPTER ONE
INTRODUCTION

CHAPTER 1 INTRODUCTION

Abstraction

Breast Cancer is the most frequent disease as a cancer type for women. Therefore, any development for diagnosis and prediction of cancer disease is capital important for a healthy life. In our project, we try to make an early predication to decrease number of patients over the world. We use 8 various algorithms on text data and Image data such as Artificial neural network algorithms: Multi-Layer Perceptron (MLP), Back-Propagation, Convolutional Neural Network (CNN), DenseNet and Machine learning algorithms such as Random Forest (RF), Logistic Regression (LR), K- nearest neighbor (KNN), Decision Tree (DT). Back-Propagation and Logistic Regression achieved high accuracy with 99.12% but Convolutional Neural Network (CNN) after augmentation achieved 75%. Using ensemble (voting) back-propagation is the best result without overfitting with 0.9912 accuracy. We use flutter framework to make web application and mobile application. We use Mongo DB For inserting and viewing the predicating data and for authenticating the registered data, we use Firebase.

1.1 Overview

Breast cancer is considered a severe danger threatening women's life and health. Breast cancer is observed to be one of the most prevalent types of cancer among women worldwide. In Egypt, all types of cancers are increasing rapidly, especially in the breast [1].

Breast cancer is the second leading reason for death among the women. Cancer starts from breast and spread to other part of the body. People are unable to identify their disease before it become dangerous. It can be cured if the disease identified at early stage.

Accurate classification of benign tumors can avoid patients undergoing unnecessary treatments. Data Analytics and machine learning methods provides framework for prognostic studies by errorless classification of data instances into relevant based on the cancer severity.

Breast cancer diagnoses can use various imaging techniques such as mammography, positron emission tomography (PET), magnetic resonance imaging (MRI), Computed tomography (CT), and single-photon emission computed tomography (SPECT). Early detection has set a necessity for the use of machine learning techniques. Breast cancer types and relative incidence (%) Ductal carcinoma in situ (23%) -Invasive lobular carcinoma (7%), Invasive ductal carcinoma (54%), Lobular carcinoma in situ (6%), Mucinous (colloid) carcinoma (1.5%), Tubular/cribriform carcinoma (4%), Medullary carcinoma (1.5%), Papillary carcinoma (0.5%), Met aplastic carcinoma (0.5%) and Other (2%) [2].

The algorithms used in that project are NN techniques, back propagation, multilayer perceptron and single layer perceptron, are adapted, the dataset of Wisconsin Breast Cancer (Diagnostic) has been used in this study, to predict the severity of breast cancer.

CHAPTER 1 INTRODUCTION

1.2Motivation

The identification method which needs if or not the cancer classification is benign or malignant in addition to need a big effort from a doctors and physicians. Once many tests are concerned within the identification of breast cancer, like clump thickness, uniformity of cell size, uniformity of cell form, etc., the ultimate result could also be difficult to get, even for doctors [3].

It is the second most common cause of death among women in the world, so that we decide to help the doctors to make an early detection by using an application which contain a modern machine learning and neural network algorithms to achieve a highest accuracy which leads to reduce the global death rate of the breast cancer.

1.3Objective

The aim of this project it to achieve high accuracy by collecting several data, following the previous project and develop those using suitable algorithms, the analysis aims to observe which features are most helpful in predicting malignant or benign cancer and to see general trends that may aid us in model selection and hyper parameter selection. The goal is to classify whether the breast cancer is benign or malignant. To achieve this i have used machine learning classification methods and neural network classification methods to fit a function that can predict the discrete class of new input and finally we aim to reduce the global death rate of the breast cancer [4].

1.4Project organization

At high level the project is organized as follows. Chapters 2 to 4, are for the first topic of the dissertation: Background and related work, planning and analysis, and design. User Interface, Proposed Model, work flow, Diagrams, Models Description, Results, conclusion and Future work. More specifically we highlight each chapter as follows.

Chapter 2 introduces the research on personalized search and reviews related prior work. We detail our approach of personalized search in Chapter 3. Designs, implementation, Results and comparative analysis, conclusion and future work are covered in chapter 4.Describe the most suitable model in details and talk about results and experiments in chapter 5 . we conclude the project and shown Future work in Chapter 6.

CHAPTER TWO

TECHNOLOGIES AND TOOLS

CHAPTER 2 TECHNOLOGIES AND TOOLS

2.1 History of machine learning.

In 1944, Logistic Regression is invented its event type is statistics. It is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist [38].

In 1957, Perceptron is invented its event type is discovery. The invention of the perceptron made a big eruption and had a big interest in mass media [38].

Multilayer perceptron used to solve nonlinear problems such as xor which add a hidden layer that helps to increase the accuracy [38].

In 1967, Nearest Neighbor is invented. It is the start of basic pattern recognition. The algorithm was used to map routes [38].

In 1972, Decision Tree made the first classification tree appeared in the THAID project (by Messenger and Mandell). It worked via splitting data to maximize the sum of cases in the modal category. The predicted class was a mode [38].

In 1986, Back-propagation is invented its event type is application. It works forward and backward to make a high accuracy which helps in solving complex problems [38].

In 1995, Random Forest is invented its event type is discovery. A forest is comprised of trees. Random forests create decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance [38].

2.2 Machine learning

Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence.

2.3 Artificial Neural Networks

Neural networks are known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Artificial neural networks (ANNs) are comprised of a node layer, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network [5].

Multilayer Perceptron (MLP): A multi-layer artificial neuron network is an integral part of deep learning. It is a combination of multiple perceptron models MLP Classifier implements a multi-layer perceptron (MLP) algorithm that trains using Back-propagation.

```
1: choose an initial weight vector ~w
2: initialize minimization approach
3: while error did not converge do
4: for all (~x, ~d) ∈ D do
5: apply ~x to network and calculate the
network output
6: calculate ∂e(~x)
7: end for
8: calculate ∂E(D)
9 for all weights summing over all
training patterns
10 perform one update step of the
minimization approach
11: end while
```

Figure 1 Pseudocode of MLP

The Multilayer Perceptron (MLP) learning procedure is as follows:

Starting with the input layer, propagate data forward to the output layer. This step is the forward propagation.

Based on the output, calculate the error (the difference between the predicted and known outcome). The error needs to be minimized.

Back-propagate the error. Find its derivative with respect to each weight in the network, and update the model.

Repeat the three steps given above over multiple epochs to learn ideal weights.

-Finally, the output is taken via a threshold function to obtain the predicted class labels.

*MLP trains using Stochastic Gradient Descent, Adam or L-BFGS. Stochastic Gradient Descent (SGD) updates parameters using gradient of loss function with respect to parameter that needs adaptation., i.e.

-Loss Function: $w \leftarrow w - \eta (\propto(\partial R(w))/(\partial w) + (\partial Loss)/(\partial w))$

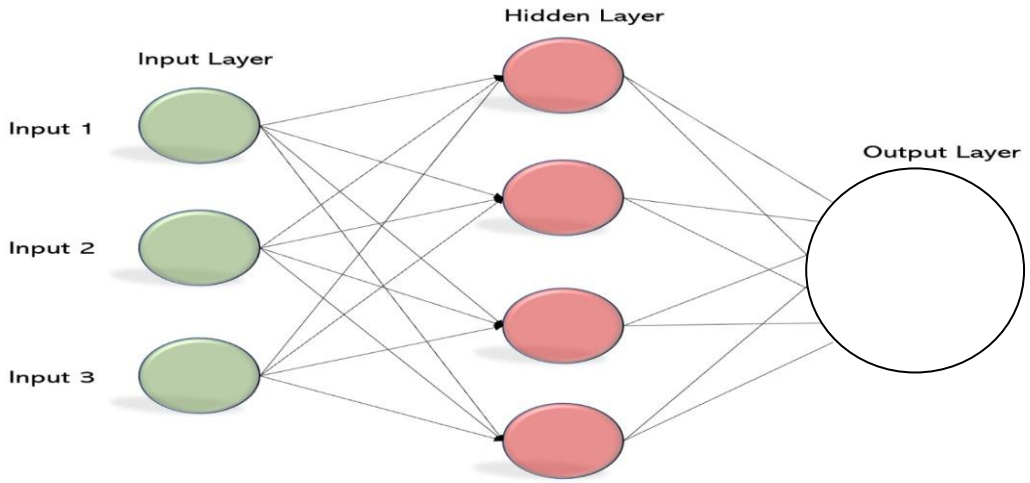


Figure 2: MLP

Back-propagation: is an algorithm for supervised learning of artificial neural networks that uses the gradient descent method to minimize the cost function. It searches for optimal weights that optimize the mean-squared distance between the predicted and actual labels.

```
Assign all network inputs and output
Initialize all weights with small random numbers, typically between -1 and 1
repeat
   for every pattern in the training set
      Present the pattern to the network
//       Propagated the input forward through the network:
    for each layer in the network
       for every node in the layer
          1. Calculate the weight sum of the inputs to the node
          2. Add the threshold to the sum
          3. Calculate the activation for the node
       end
    end
//       Propagate the errors backward through the network
     for every node in the output layer
       calculate the error signal
    end
        for all hidden layers
       for every node in the layer
          1. Calculate the node's signal error
          2. Update each node's weight in the network
       end
    end
//       Calculate Global Error
    Calculate the Error Function
   end
while ((maximum  number of iterations < than specified) AND
   (Error Function is > than specified))
```

Figure 3 Pseudocode of Backpropagation

The back-propagation learning algorithm can be divided into two phases:

1- propagation

2- weight update.

Each propagation involves the following steps:

Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations [8].

Backward propagation of the propagation's output activations through the neural network using the training pattern target in order to generate the deltas of all output and hidden neurons.

Loss Function = ∑ ½ (Desired Output - Predicted Output) ^2

Here, ½ is a constant used.

After got the loss function, now we should minimize the loss to have more accuracy (we have to decrease or minimize the value of the loss function to zero).

To update the weights, we should follow the following steps:

1-value of weights

2- Update W(init) in the direction of minimizing gradient(slope), when it reaches the local minima, it would have decreased its error.

3-Update the weights with this formula
Updated weight -> weight - ∝ * ∂ (error)/∂ (weight)

∝ = Learning rate
Repeat 3 until local minima.

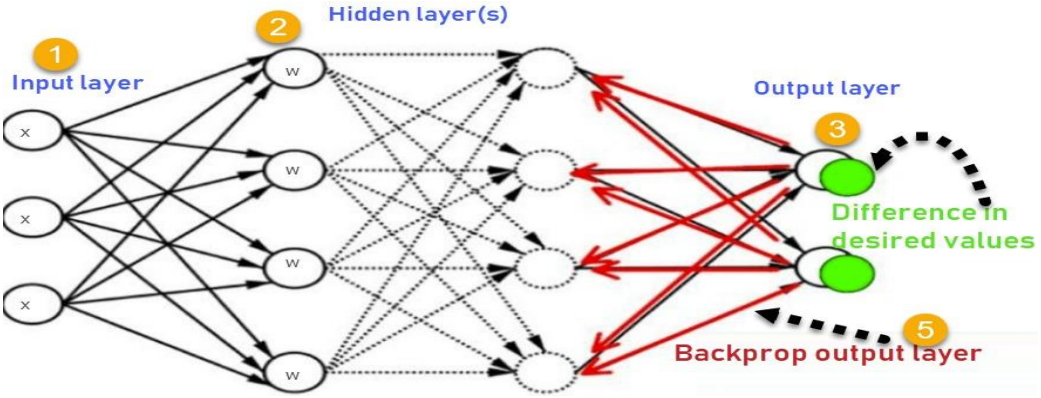


Figure 4: Back-Propagation Architecture

DenseNet is densely connected-convolutional networks. It is very similar to a ResNet with some-fundamental differences.

DenseNet is a network that portrays the importance of having dense connections in a network using dense blocks. This helps in feature-reuse, better gradient flow, reduced parameter counts and better transmission of features across the network.

Algorithm 2 : Extraction of Secret information

Input: Cover images $Pc = \{Pc_1, Pc_2, \ldots, Pc_{cg}\}$.

Output: Secret information S'.

1: for cg = 1:m
2: Extract DenseNet feature:F_{cg}=DenseNet(Pc_{cg})
3: Divide DenseNet feature into D blocks
4: for ib = 1:D
5: For each feature block B_{ib}, using Eq.(4) to calculate the feature coefficient Me_{ib}
6: end
7: Using Eq.(6) to obtain the hash sequence f_{cg} according to arithmetic scan
8: Connect the M: S' += f_{cg}
9: end
10: Return the$M$: S'

Figure 5 Pseudocode of DenseNet

The main key takeaway is the number of connections:

generally, in any architecture, the number of connections is typically the same as the number of layers, but here the number of connections is L(L+1)/2, here L = numbers of the layers.

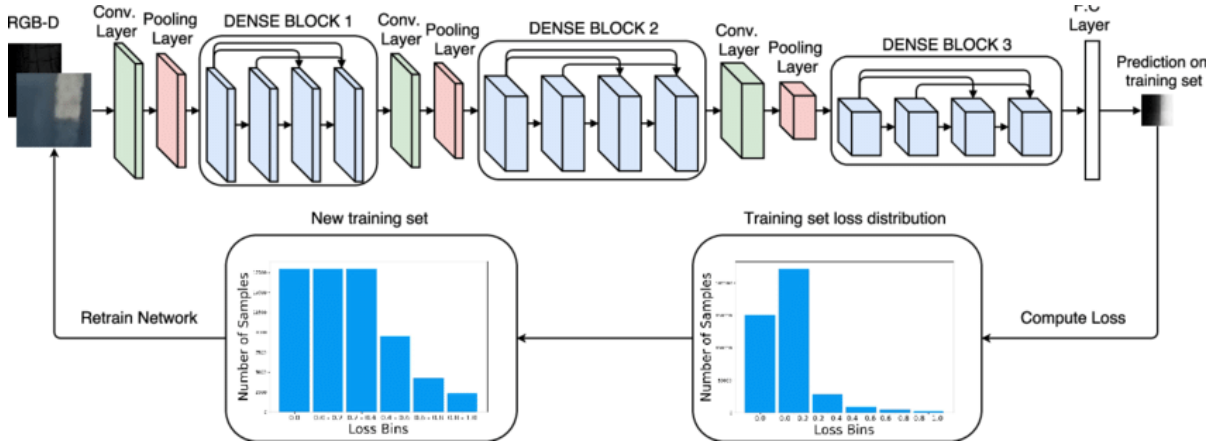


Figure 6 DenseNet architecture

CHAPTER 2 TECHNOLOGIES AND TOOLS

Convolutional Neural Network (CNN) is a Deep Learning algorithm specially designed for working with Images. It takes images as inputs, extracts and learns the features of the image, and classifies them based on the learned features [9].

CNN is used for images as:

-Some patterns are smaller than the whole image.

- The neuron does not have to see the whole image to discover the pattern.

- Connecting to small region with less parameters

- Subsampling the pixels will not change the object

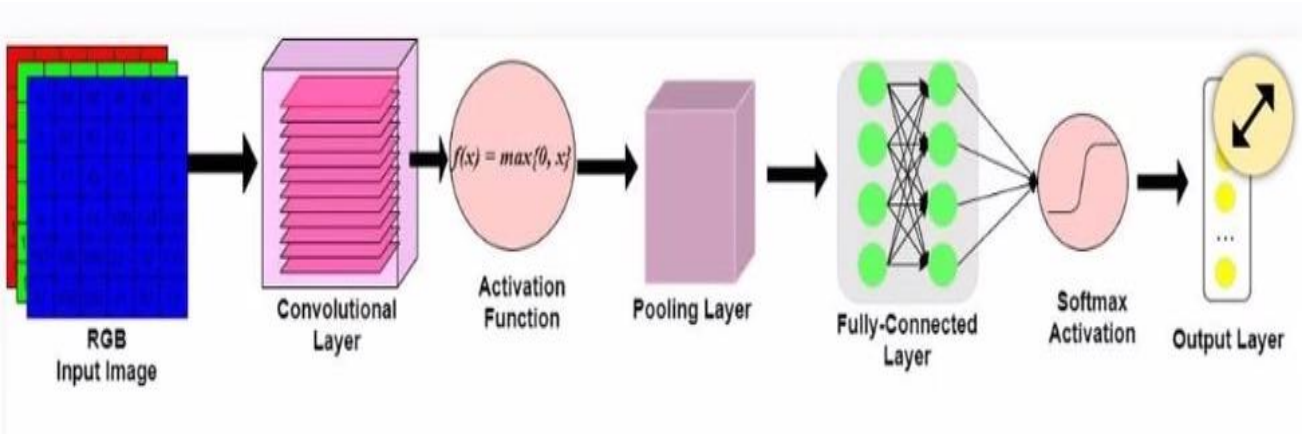


Figure 7 CNN architecture

A typical CNN model looks like this:

-Input layer: The input is image and may be Grayscale or RGB. Images consist of pixels that are in range from 0 to 255. Normalize them is essential to convert the range of every image between 0 to 1 before applying the model on it.

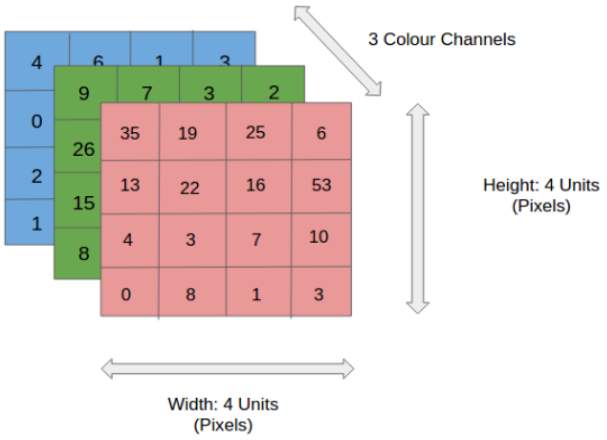


Figure 8 CNN Conv. Layer Matrix

Convolution layer + Activation Function

The convolution layer is the first step in the process of extracting valuable features from an image to make feature extraction. A filter is applied to the image multiple times and creates a feature map which helps in classifying the input image [10].

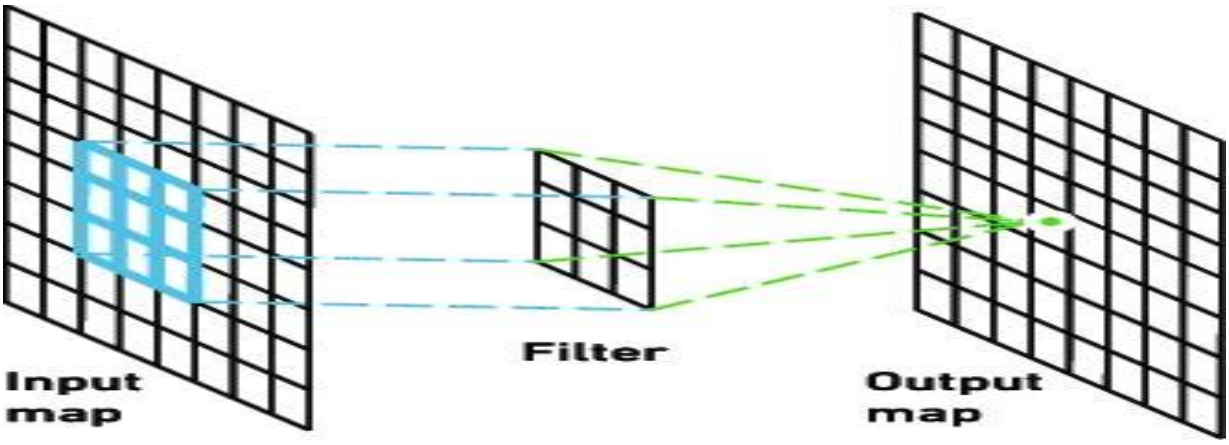


Figure 9 CNN Filters

The activation functions

SoftMax is numeric output of the last linear layer

of a multi-class classification neural network

Pooling layer

This layer is applied after Convolution layer it take image from the Convolution layer and downsizing dimension of the image. It keeps the important features and reduces the computation time.

It consists of Max pooling, Average pooling and Summation pooling.

Max pooling:

It separates the matrix into sub matrices and take the maximum value from each sub matrix.

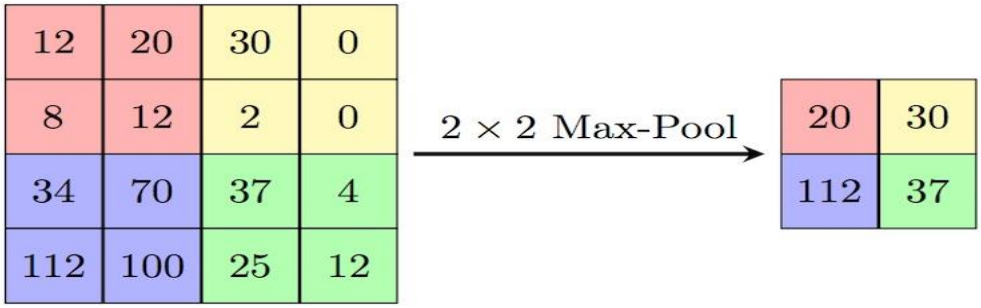


Figure 10 CNN Max Pooling

Average pooling:

It separates the matrix into sub matrices and take the average from each sub matrix.

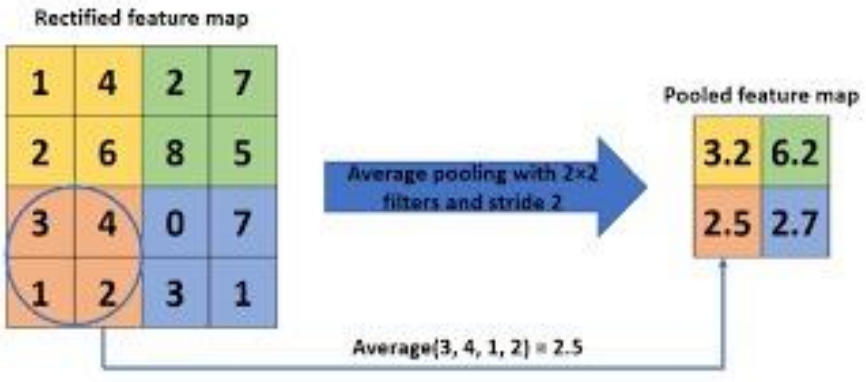


Figure 11 CNN Average Pooling

Summation pooling:

It separates the matrix into sub matrices and take the sum of each sub matrix.

Flatten layer:

It converts the 2D matrix into vector to use as an input in the ANN.

Fully Connected:

The Fully connected layer (as we have in ANN) is used for classifying the input image into a label. This layer connects the information extracted from the previous steps (i.e., Convolution layer and Pooling layers) to the output layer and eventually classifies the input into the desired label [11].

```
1: algorithm Parallel-CNN
2: input: d: dataset, l: dataset true labels, W:
Word2Vec matrix
3: output: score of Parallel-CNN trained model on
test dataset
4: let f be the featureset 3d matrix
5: for i in dataset do
6:   let f_i be the featureset matrix of sample i
7:   for j in i do
8:     v_j ← vectorize_(j, W)
9:     append v_j to f_i
10:    append f_i to f
11: f_train, f_test, l_train, l_test← split feature set and labels
into train subset and test subset
12: M ← Parallel-CNN (f_train, l_train)
13: score ← evaluate (i, l_test, M)
14: return score
```

Figure 12 pseudo code of CNN

Ensemble learning:

The idea of ensemble classification is to learn not just one classifier but a set of classifiers, and them to combine their predictions for the classification of unseen instances using some form of voting.

There are several methods to combine the outputs of the various classifiers.

Voting or averaging of predictions of multiple trained models.

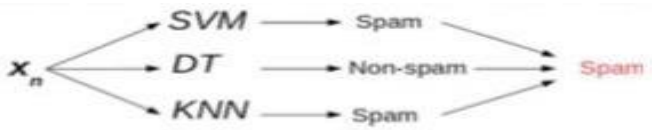


Figure 13 ensemble

Stacking: Use predictions of multiple models as “features” to train a new model and use the new model to make predictions on test data.

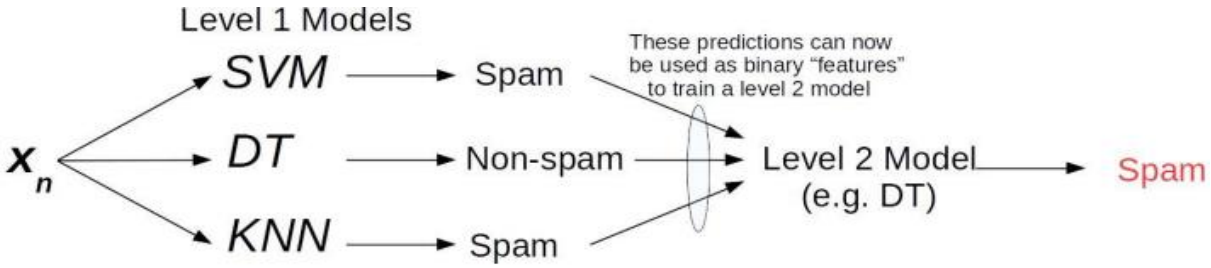


Figure 14 stacking

2.4 Machine learning algorithms

Logistic regression: is used for predicting the categorical dependent variable using a given set of independent variables and used for solving the classification problems [12].

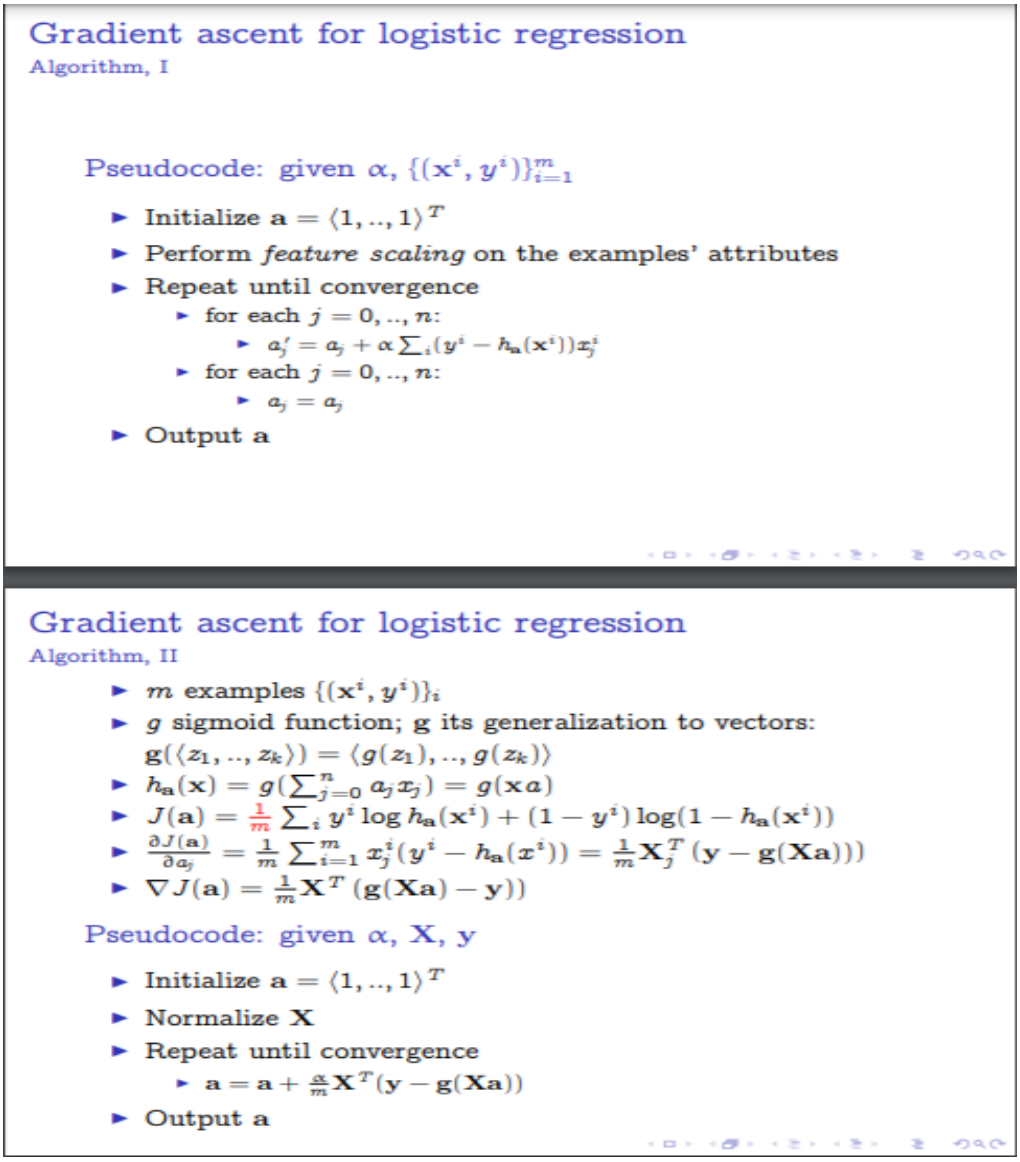


Figure 15 Pseudocode of logistic regression

This screenshot contains pseudocode of logistic regression as shown in the previous figure [41].

In logistic regression, we pass the weighted sum of inputs through an activation function that can map values in between 0 and 1. Such activation function is known as sigmoid function.

The equation of logistic regression:

Logit(p) = log(p/1-p)

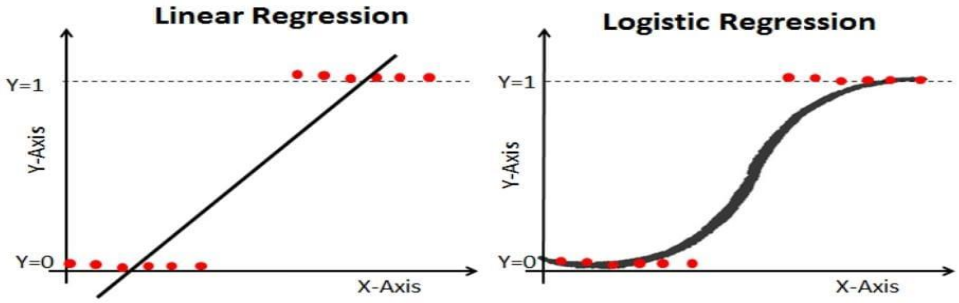


Figure 16 difference between logistic and linear regression

Decision Tree: C 5 algorithm is a tree-structured plan of a set of attributes to test in order to predict the output. To decide which attribute should be tested first, simply find the one with the highest information gain [13].

Information gain: is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.

Information Gain= 1-Entropy

Advantages of the Decision Tree:

It can be very useful for solving decision-related problems.

There is less requirement of data cleaning compared to other algorithms

Disadvantages of the Decision Tree

The decision tree contains lots of layers, which makes it complex.

It may have an overfitting issue, which can be resolved using the Random Forest algorithm.

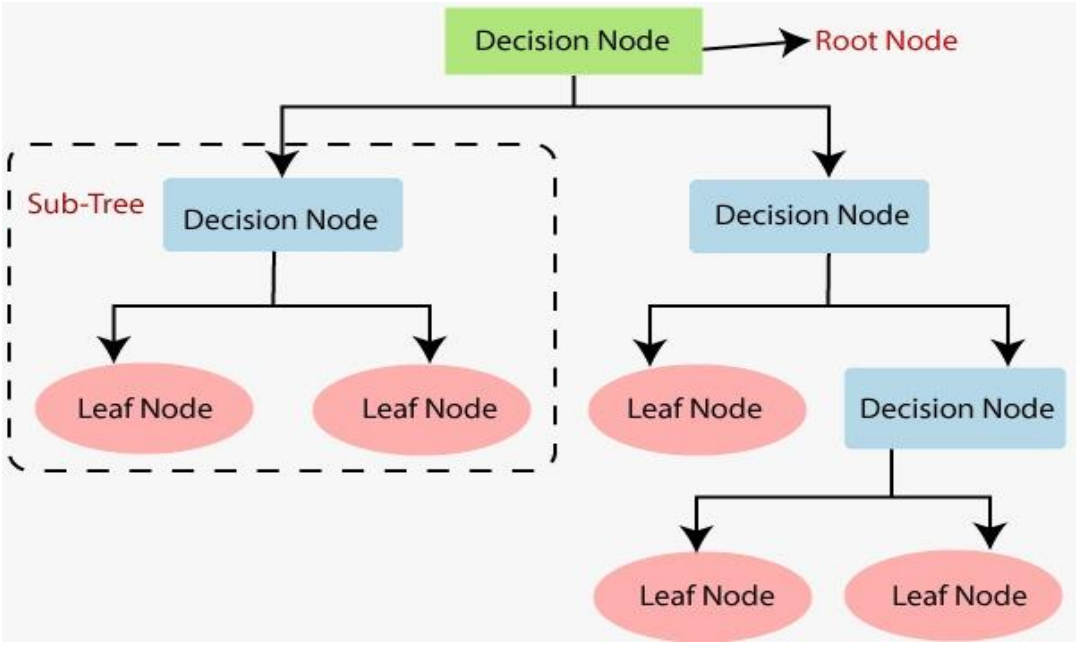


Figure 17 decision tree

```
Tree-Learning (TR, Target, Attr)
      TR: training examples
      Target: target attribute
      Attr: set of descriptive attributes
{
      Create a Root node for the tree.
      If TR have the same target attribute value t_i,
        Then Return the single-node tree, i.e. Root, with target attribute = t_i
      If Attr = empty (i.e. there is no descriptive attributes available),
        Then Return  the single-node tree, i.e. Root, with most common value of Target in TR
      Otherwise
      {
         Select attribute A from Attr that best classify TR based on an entropy-based measure
         Set A the attribute for Root
         For each legal value of A, v_i, do
         {
            Add a branch below Root, corresponding to A = v_i
            Let TR_vi be the subset of TR that have A = v_i
            If TR_vi is empty,
              Then add a leaf node below the branch with target value =  most common value of
                   Target in TR
              Else below the branch, add the subtree learned by
                   Tree-Learning(TR_vi, Target, Attr-{A})
         }
      }
      Return (Root)
}
```

Pseudocode of decision tree learning.

Figure 18 pseudocode of decision tree

This screenshot contains pseudocode of decision tree as shown in the previous figure [35].

KNN algorithm: is used for the Classification problems. it is a data classification algorithm that attempts to determine what group a data point is in by looking at the data points around it.

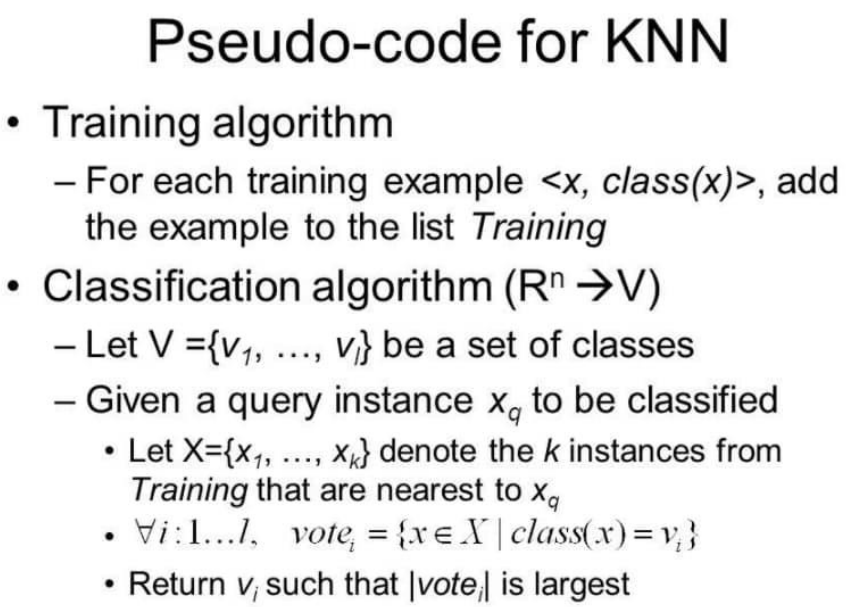


Figure 19 Pseudocode of KNN

The KNN working can be explained:

1- Select the number K of the neighbours

2-Calculate the distance of K number of neighbours

3- Take the K nearest neighbours as per the calculated distance.

4- Among these k neighbours, count the number of the data points in each category.

5- Assign the new data points to that category for which the number of the neighbour is maximum.

Advantages of KNN Algorithm:

It is simple to implement.

It can be more effective if the training data is large.

Disadvantages of KNN Algorithm:

Always needs to determine the value of K which may be complex some time.

The computation cost is high because of calculating the distance between the data points for all the training samples.

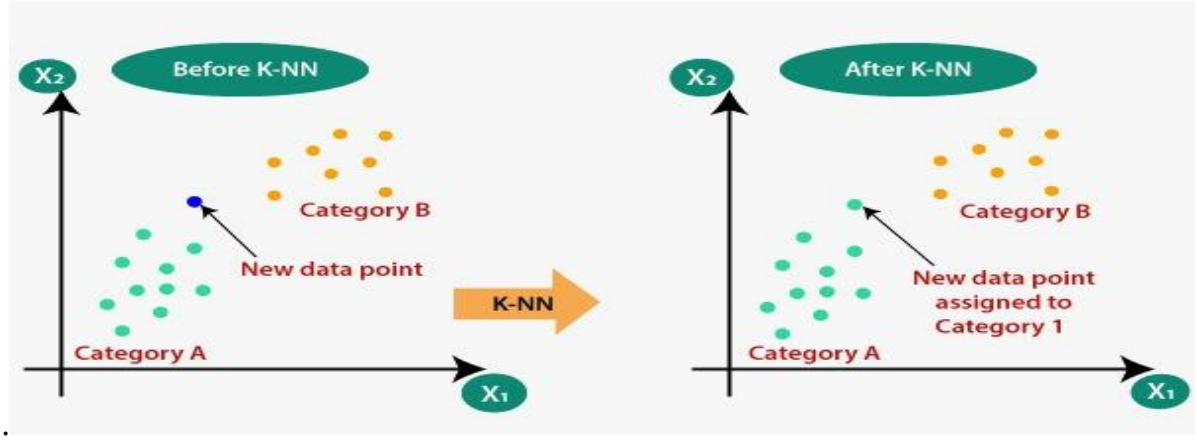


Figure 20 KNN

Random Forest: It is the most flexible and easy to use algorithm. A forest is comprised of trees. Random forests create decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance [14].

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble.

Algorithm 1: Pseudo code for the random forest algorithm

To generate c classifiers:
for i = 1 to c **do**
  Randomly sample the training data D with replacement to produce D_i
  Create a root node, N_i containing D_i
  Call BuildTree(N_i)
end for

BuildTree(N):
if N contains instances of only one class **then**
  return
else
  Randomly select x% of the possible splitting features in N
  Select the feature F with the highest information gain to split on
  Create f child nodes of N, N_1,..., N_f, where F has f possible values (F_1, ... , F_f)
  for i = 1 to f **do**
    Set the contents of N_i to D_i, where D_i is all instances in N that match F_i
    Call BuildTree(N_i)
  end for
end if

Figure 21 Pseudocode of Random Forest

Ensemble: It is a machine learning paradigm where multiple learners are trained to solve the same problem.

Advantages:

The algorithm can be used in both classification and regression problems.

Random forests can also handle missing values. There are two ways to handle these: using median values to replace continuous variables, and computing the proximity-weighted average of missing values.

It reduces overfitting in decision trees and helps to improve the accuracy [15].

Disadvantages:

Random forests are slow in generating predictions because it has multiple decision trees. Whenever it makes a prediction, all the trees in the forest have to make a prediction for the same given input and then perform voting on it. This whole process is time-consuming.

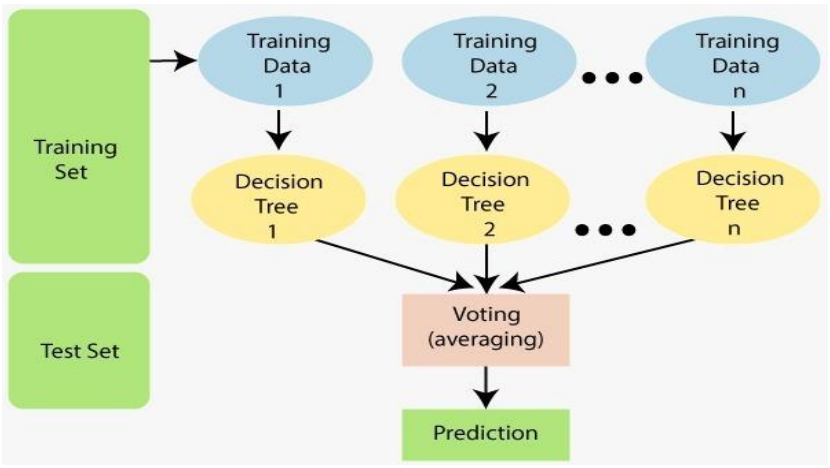


Figure 22 Random Forest

2.3 Flutter

Flutter is a free and open-source mobile UI framework created by Google and released in May 2017. In a few words, it allows you to create a native mobile application with only one codebase. This means that you can use one programming language and one codebase to create two different apps (for iOS and Android). Flutter consists of two important parts: An SDK (Software Development Kit): A collection of tools that are going to help you develop your applications. This includes tools to compile your code into native machine code (code for iOS and Android).

A Framework (UI Library based on widgets): A collection of reusable UI elements (buttons, text inputs, sliders, and so on) that you can personalize for your own needs. To develop with Flutter, you will use a programming language called Dart. The language was created by Google in October 2011, but it has improved a lot over these past years. Dart focuses on front-end development, and you can use it to create mobile and web applications. If you know a bit of programming, Dart is a typed object programming language. You can compare Dart's syntax to JavaScript. Supported by Android Studio and VS Code Flutter is available on different IDEs. The two main code editors for developing with this technology are Android Studio (IntelliJ) and VS Cod [34].

CHAPTER THREE
RELATED WORK

CHAPTER 3 RELATED WORK

In 2017, Lingraj Dora, Sanjay Agrawal, Rutuparna Panda b, Ajith Abraham, experiments on different datasets (WBCD and WDBC) demonstrate that GNRBA outperforms the existing models. It is observed that the proposed method achieved the highest classification accuracies of 98.54% for 50–50 partition, 99.27% for 60–40 partition and 100% for 70–30 partition for a subset (WBCD) that contains 74, 92 and 02 training samples, respectively. Results also show the superiority of the proposed method in terms of sensitivity, specificity, confusion matrix, statistical test and AUC. in the breast cancer classification problem, tumor samples are considered for two classes only i.e., benign or malignant. [1].

In 2017, Shajib Ghosh, Jubaer Hossain, Dr. Shaikh Anowarul Fattah, Dr. Celia Shahnaz, Asir Intisar Khan, in 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 Accuracy (%):higher accuracy is obtained which is up to 98% to 99%.Data Percentage is Test Data(15%),Train Data(85%) Naïve Bayes(92.9412),Random Forest(97.6471)KNN(96.4706) SVM(98.8235),Logistic Regression(95.2940)Test Data(20%),Train Data(80%) NaïveBayes(90.3509) Random Forest(93.8596) KNN(95.6140),Logistic Regression(95.6140),SVM(98.2456) Our results of CNN classifier (98.06% accuracy) [2] .

In 2017, Abdelkader Helwana, John Bush Idokob, Rahib H. Abiyevb the used networks in this work showed promising performances when tested on un seen cases; breast cancer other than those were used in the training phase. Best overall classification accuracies obtained from the BPNNs and RBFNs are 93.39% for BPNN3 and 94.33% for RBFN2[3].

In 2017 , Dada Emmanuel Gbenga, Ngene Christopher, Daramola Comfort Yetunde the paper investigate the performance of 8 machine learning algorithms that have been applied for timely detection of breast cancer .The experimental results indicated that support vector machine (SVM) have the best performance in term of classification accuracy (97.07%) and lowest error rate compared to Radial Based Function(96.49%) ,Simple logistic linear regression model (96.78%), Naïve Bayes(96.48%), K- Nearest Neighbor (96.34%) , Ada boost (96.19%), Fuzzy Unordered Role Induction algorithm (96.78%) and Decision Tree(96.48%)[4]**.**

In 2018 ,Omar Ibrahim Obaid , Mazin Abed Mohammed , Salama Mostafa, Mohd Khanapi Abd Ghani In this paper ,The authors used three machine-learning algorithms (Support Vector Machine, K-nearest neighbors, and Decision tree) have been used and the performance of these classifiers has been compared in order to detect which classifier works better in the classification of breast cancer .The outcomes of this study have revealed that quadratic support vector machine grants the largest accuracy of (98.1%) with lowest false discovery rates[5].

CHAPTER 3 RELATED WORK

In 2018, Muhammet Fatih Aslan, Yunus Celik, Kadir Sabanci, Akif Durdu the authors use ANN, ELM, k-NN and SVM methods which are determined using Hyperparameter optimization technique. the highest accuracy rate and the lowest training period are provided by standard Almac. Rate (%): ANN: 79.4304, ELM: 80, k-NN: 77.5, SVM: 73.5 the highest accuracy rate and the lowest training period are provided by standard ELM [6].

In 2018, DALAL BARDOU, KUN ZHANG, AND SAYED MOHAMMAD AHMAD it achieved accuracy between 96.15% and 98.33% for the binary classification and 83.31% and 88.23% for the multi-class classification. In this work, they compared the performance of convolutional neural networks with various configurations for the classification of breast cancer histology images into benign and malignant, and also into benign and malignant subclasses too. The designed CNN topology worked well on both binary and multi-class classification tasks Our proposed CNN topology has beaten the previous ones for the binary classification task, where we reached a performance of between 96.15% and 98.33 [7].

In 2018, Yixuan Li, Zixuan Chen in this study, the method of DT, SVM, RF, LR and NN models are employed as the classification to predict the nature of breast cancer with other attributes. Then we preprocess the raw data of WBCD dataset and obtained the data that contains 683 volunteers with 9 attributes and the index indicating whether the volunteer has the malignant tumor. After comparing the accuracy, F-measure metric and ROC curve of five classification models, the result has shown that RF is selected as the primary classification model in this study. Therefore, the results of this study provide a reference for experts to distinguish the nature of breast cancer. Currently, the factors of malignant breast cancer become more and more complex [8].

In 2018, Walid CHERIF the authors used K-Nearest Neighbors algorithm (KNN) Results of the proposed approach exceeded most known classification techniques with an average f-measure exceeding 94% on the considered breast-cancer Dataset and summarizes the results of classification of the proposed approach compared to four other algorithms. The proposed approach yielded an average f-measure: 94.1%, exceeding: SVM (89.7%), NB (92.2%) and KNN (91.1). the only algorithm which returned slightly better performance is ANN with an average f-measure: 95.6% [9].

CHAPTER 3 RELATED WORK

In 2018, Wenbin Yue, Zidong Wang, Hongwei Chen, Annette Payne, Xiaohui Firstly, we provide an overview of ML techniques including artificial neural networks (ANNs), support vector machines (SVMs), decision trees (DTs) and k-nearest neighbors (k-NNs). Then, we investigate their applications in BC. Compared with traditional PSO algorithms, the modified PSO algorithms have better performances in many aspects. For instance, using the switching particle swarm optimization (SPSO) algorithm based on Markov chains and the competitive penalized method can achieve faster local convergence speed with higher accuracy and reliability, and the switching delayed particle swarm optimization (SDPSO) algorithm has been validated showing it can improve global search and increase possibility of reaching the global best. Next, we will test these different novel PSO algorithms in combination with ANN via tuning weights to find out whether we can achieve a higher classification accuracy with WBCD data. These datasets have been analyzed using some traditional statistical approaches, such as cross-sectional and linear regression analysis to gain general statistical knowledge about the distribution and variation of the data [10].

In 2019, Rati Shukla, Vikash Yadav, Parashu Ram Pal, Pankaj Pathak the authors used Machine learning (ML) algorithms Classification of cancer sufferer using Machine Learning methodologies in different class of risk criterion such as high, low and medium has led many research dimensions of life science data. Therefore, Machine Learning is one of the very use full methodologies to study and design the different class of development and prognosis of cancerous situation. Machine learning methods are very powerful and effective tool for key feature extraction and classification form complex cancerous data set. In this study, we put forward applicability of different Machine Learning classification techniques employed in the prediction and prognosis of Breast Cancer [11].

In 2019, Ali Al Bataineh in this paper, a performance comparison between five nonlinear machine learning algorithms viz Multilayer Perceptron (MLP),K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Nave Bayes (NB) and Support Vector Machines (SVM) on the Wisconsin Breast Cancer Diagnostic (WBCD) dataset is conducted Results of this study confirm that the MLP model has the highest performance in terms of accuracy, precision, and recall of 99.12%, 99.00%, and 99.00% respectively [12].

In 2019, Bibhuprasad Sahu, Sachi Nandan Mohanty, Saroj Kumar Rout For automation of the diagnosis process data mining plays a significant role. This study suggests a hybrid feature selection method to be used with PCA (Principal Component Analysis) and Artificial Neural Network (ANN). Pre-processing of data and extracting the most relevant features done by PCA. The authors have done comparison between different hybrid methods like PCA+RF, PCA+KNN, PCA+NN.As expected this model performs better and provides a promising result of ANN for 70-30% partition spilt. From this comparison we can easily analyze KNN, ANN, PCA+ANN provide the accuracy of 97%, whereas PCA, PCA+RF and NB provide accuracy of 95% and 91% respectively.

if we comparison between KNN, ANN and PCA+KNN in term of all evaluation criteria such as Kappa statistics, Sensitivity, Specificity, Prevalence PCA+KNN performs better. So PCA+ANN considered as a better classifier as compared to others. The result reveals ANN (machine learning) plays measure factor for detection of cancer diagnosis to save the human life from the dangerous disease [13].

CHAPTER 3 RELATED WORK

In 2019, David A. Omondiagbe, Shanmugam Veeramani, Amandeep S. Sidhu the authors proposed a hybrid approach for breast cancer diagnosis by reducing the high dimensionality of features using linear discriminant analysis (LDA), and then applying the new reduced feature dataset to Support Vector Machine. The proposed approach obtained an accuracy of 98.82%, sensitivity of 98.41%, specificity of 99.07% and area under the receiver operating characteristic curve of 0.9994. Feature selection and feature extraction methods were performed on the data to reduce the dimension of features, there by producing reduced versions of the original dataset. The methods considered are CFS, RFE, PCA and LDA. SVM, ANN and NBC were employed to train the datasets.

This paper analyzed WDBC dataset. Simulation results showed that SVM-LDA and NN-LDA outperforms the other ML classifier models. Nevertheless, SVM-LDA is chosen over NN-LDA. This chosen approach showed good results over the validation dataset. It obtained a classification accuracy of 98.82%, sensitivity of 98.41%, specificity of 99.07% and area under the receiver operating characteristic curve of 0.9994[14].

In 2019, Ebru Aydındag Bayrak, Pınar Kırcı, Tolga Ensari The authors used machine learning techniques have been used for classification of Wisconsin Breast Cancer (Original) dataset and the classification performance of these techniques have been compared with each other using the values of accuracy, precision, recall and ROC Area. The best performance has been obtained by Support Vector Machine technique with the highest accuracy. Artificial Neural Network and Support Vector Machine are used as ML techniques for the classification of WBC (Original) dataset in WEKA tool. The author used 2 models in ANN: MLP with accuracy 95%, Voted perceptron with accuracy 90%, The author used 2 models in SVM: SMO with high accuracy 96%, Lib SVM with accuracy 95% [15].

CHAPTER 3 RELATED WORK

In 2019, Ratula Ray, Azian Azamimi Abdullah, Debasish Kumar Mallick, Satya Ranjan Dash we are focusing towards understanding the shortcomings of digital mammograms in detection of breast cancer and utilize Machine Learning classifiers for the classification of benign and malignant tumors using image analysis. we are also looking into implementing Supervised Machine Learning classifiers such as Decision Tree, (KNN), Random Forest and Gaussian Naïve Bayes classifiers for assessing the risks involved. In the numeric dataset, we find Decision Tree and Random Forest classifiers showing promising results with an accuracy score of 70%. Our result depicts that Random Forest Classifier shows the best result with F1-score of 92%. For the numeric dataset, we have seen that Random Forest and Decision Trees give comparable results 70% [16].

In 2020, Abderrahmane Eddaoudy, Khalil Maalmi the correct classification rate obtained with Support Vector Machine (SVM) model with AR shows the highest classification accuracy (98.00%) for eight attributes and 96.14% for 4 attributes. In this work, an approach for detecting breast cancer based on AR and SVM has been proposed. The proposed AR + SVM approach performance is evaluated and compared with SVM model and previous work in this context. SVM and AR feature selection gave the highest accuracy of 98.00%for eight inputs, while 96.14% is achieved using four inputs. This research demonstrated that the AR can be used for reducing the dimension of feature vector and proposed SVM model can be used to obtain efficient automatic diagnostic systems [17].

In 2020, Nikita Rane, Rucha Kanade, Jean Sunny, Prof. Sulochana Devi the authors present a comparison of six machine learning algorithms: Naive Bayes (NB), Random Forest (RT), Artificial Neural Networks (ANN), k Nearest Neighbor (KNN), Support Vector Machine (SVM) and Decision Tree (DT) on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. By using machine learning algorithms, we will be able to classify and predict the cancer into being or malignant. Machine learning algorithms can be used for medical oriented research, it advances the system, reduces human errors and lowers manual mistakes [18].

In 2020, Shallu Sharma, Rajesh Mehra used two machine learning approaches are thoroughly explored and compared for the task of automatic magnification-dependent multi-classification on a balanced Break His dataset for the detection of breast cancer. The first approach is based on handcrafted features which are extracted using Hu moment, color histogram, and Haralick textures. The second approach is based on transfer learning where the pre-existing networks (VGG16, VGG19, and ResNet50) are utilized as feature extractor and as a baseline model. The results reveal that the use of pre-trained networks as feature extractor exhibited superior performance in contrast to baseline approach and handcrafted approach for all the magnifications. In this context, the VGG16 network with linear SVM provides the highest accuracy that is computed in two forms, (a) patch-based accuracies (93.97% for 40×, 92.92% for 100×, 91.23% for 200×, and 91.79% for 400×); (b) patient-based accuracies (93.25% for 40×, 91.87% for 100×,91.5% for 200×, and 92.31% for 400×) for the classification of magnification-dependent histopathological images It has been observed throughout our study that the transfer learning approach as feature extractor provides a remarkable performance in contrast to other employed approaches. Among different combinations of classifiers, VGG16 + SVM (L, 1) provides the best result for all magnification factors (×40, ×100, ×200, and ×400) [19].

CHAPTER 3 RELATED WORK

In 2020, Sara Laghmati, Bouchaib Cherradi, Amal Tmiri, Othmane Daanouni and Soufiane Hamida used Machine learning (ML) algorithms offer alternative to breast cancer standard techniques of prediction. compares four machine learning algorithms (kNN), decision tree, Binary SVM, and Adaboost) to predict whether a patient has a malignant or a benign tumor. The predictive accuracy reached a 99.12% for the kNN model, the best predictive specificity obtained was 98.86% for the Binary SVM model and the highest predictive sensitivity obtained was up to one for both KNN and Adaboost models [20].

In 2020, Ghada Hamed(B), Mohammed Abd El-Rahman Marey, Safaa El-Sayed Amin, Mohamed Fahmy used DDSM, CBIS-DDSM and INbreast. Research indicates that most experienced physicians can diagnose cancer with 79% accuracy while 91% correct diagnosis is achieved using machine learning techniques. It is deduced that YOLO and RetinaNet is the new models that are recently used and considered more simpler than the conventional CNN networks and achieve better results and accurate performance for the mass detection and malignancy classification [21].

In 2020, Puja Guptaa, Shruti Garga used Support Vector Machine with radial basis function kernel. Deep learning using Adam Gradient Descent Learning was also applied because it combines the benefits of adaptive gradient algorithm and root mean square propagation and combine of six machine learning algorithms. The accuracy achieved by deep learning using Adam Gradient Descent Learning is 98.24%. six machine learning models: KNN 95.8%, LG 95.8%, DT 95.8%, RF 97.2%, SVM 97.2%, DL_ANN 98.24% [22].

In 2020, Md. Milon Islam, Md. Rezwanul Haque, Hasib Iqbal, Md. Munirul Hasan, Mahmudul Hasan, Muhammad Nomani used four different machine learning algorithms were used to support vector machine (SVM), K-nearest neighbors, random forests, artificial neural networks (ANNs) and logistic regression. The results reveal that the ANNs obtained the highest accuracy, precision, and F1 score of 98.57%, 97.82%, and 0.9890, respectively, whereas 97.14%, 95.65%, and 0.9777 accuracy, precision, and F1 score are obtained by SVM, respectively. Accuracy (%): SVM:97.14K-NN:97.14 RF:95.71 ANN:98.57 LR: 95.71 [23].

In 2020, Jiande Wu and Chindo Hicks used machine learning techniques. evaluated four different classification models including Support Vector Machines, K-nearest neighbor, Naïve Bayes and Decision tree using features selected at different threshold levels to train the models for classifying the two types of breast cancer. Results: Among the four ML algorithms evaluated, the Support Vector Machine algorithm was able to classify breast cancer more accurately. Accuracy: K-nearest neighbor (kNN) 87%, Naïve Bayes (NGB) 85%, Decision trees (DT) 87%, Support Vector Machines (SVM) 90% [24].

CHAPTER 3 RELATED WORK

In 2020, Yadavendra, Satish Chand use different machine learning methods such as logistic regression, random forest, support vector classifier (SVC), AdaBoost classifier, bagging classifier, voting classifier, and exception model to classify the breast cancer tumor and evaluate their performances. We used a standard dataset. We use 60% of the above-mentioned dataset for training, 20% for validation, and 20% testing to all above-mentioned classifiers. The logistic regression classifier provides the scores of each precision, recall, and F1 measure as 0.72. The random forest method provides the score of each precision, recall, and F1 score as 0.80. The bagging and voting classifiers both have the values of each precision, recalls, and F1 scores as 0.81. In this case, both SVC and AdaBoost classifiers have the score of each precision, recall, and F1 score as 0.82 [25].

In 2021, Srwa Hasan Abdulla, Ali Makki Sagheer use Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), and Random Forest (RF to classify medical images into malignant and benign. Moreover, deep learning has been employed recently for the same purpose, among them, Convolutional Neural Network (CNN) is one of the most popular techniques. The results showed that the SVM achieved high accuracy, about 97%, therefore, the researchers utilized various functions for this algorithm and added more features such as bagging and boosting to increase its efficacy. In addition, deep learning obtained high accuracy using CNN which is higher than 98% [26].

In 2021, Said Boumaraf , Xiabi Liu , Yuchai Wan 2, Zhongshu Zheng , Chokri Ferkous , Xiaohong Ma ,Zhuo Li ,Dalal compare the performance of conventional machine learning (CML) against deep learning (DL)-based methods. For CML-based methods, we extract a set of handcrafted features using three feature extractors and fuse them to get image representation that would act as an input to train five classical classifiers. For DL-based methods, we adopt the transfer learning approach to the well-known VGG-19 deep learning architecture, where its pre-trained version on the large-scale ImageNet, is block-wise fine-tuned on histopathological images. The highest classification results achieved using CML approaches were between 85.65% and 89.32% for binary classification and between 63.55% and 69.69% for eight-class classification, while the best obtained classification accuracies using the DL approach range from 94.05% to 98.13% for binary classification and between 76.77% and 88.95% for eight-class classification [27].

In 2021, Dina A. Ragab , Omneya Attallah , Maha Sharkas , Jinchang Ren , Stephen Marshall presents a new computer-aided diagnosis (CAD) system based on feature extraction and classification using DL techniques to help radiologists to classify breast cancer lesions in mammograms In the second one, the deep features of the DCNNs are extracted and fed to a support vector machine (SVM) classifier with different kernel functions. The third experiment performs deep features fusion to demonstrate that combining deep features will enhance the accuracy of the SVM classifiers. In the fourth experiment, principal component analysis (PCA) is introduced to reduce the large feature vector produced in feature fusion and to decrease the computational cost. For the CBIS-DDSM dataset the best result is: GoogleNet **,** DCNN Accuracy is 76.01% ,Training Time is 12 for hour the MIAS dataset the best result is: GoogleNet **,** DCNN Accuracy is76.01% [28].

CHAPTER 3 RELATED WORK

In 2021, V Chourasia , MK Pandey , S Pal discussed the implementation of data mining strategies to detection as well as prediction of breast malignant tumors, including random forest (RF), support vector classifier (SVC), k-nearest neighbors (KNN), linear discriminant analysis (LDA), Gradient Boosting Classifier (GBC), Decision Tree (DT) In addition, principal component analysis (PCA. The sequential feature selection (SFS) method is used for comparing the accuracy of a data set with all features and the accuracy of a classifier with selected features. The results show that RF_sfs, KNN_sfs, SVC_rbf and SVC_sfs have the highest and equal accuracy, which is 97.66%. They perform well and can predict the growth of harmful malignant tumors [29].

CHAPTER FOUR
Analysis Phase

4.1 UML Diagrams

A sequence diagram describes some objects work together and how they interact with each other in an organized sequence [37].

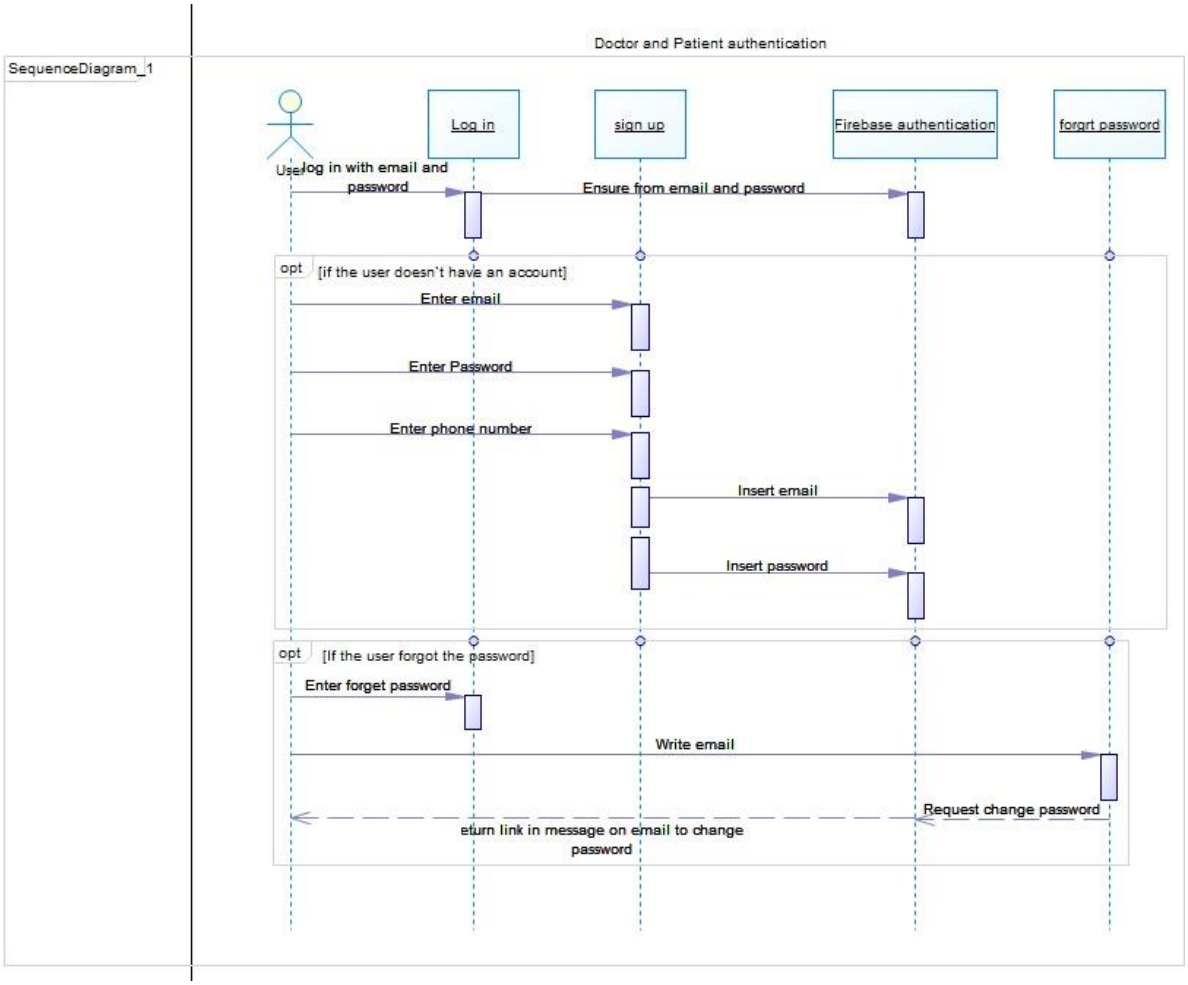


Figure 23 Sequence Diagram

User log in to his /her account with email and password then the email and password are checked by firebase authentication whether valid or invalid.

If the user doesn't have an account, the user should enter email, password and phone number and the email and password are checked by firebase authentication whether valid or invalid.

If the user forgot the password, user should click on forget password button and enter his/her email then the user will receive a message on his/her email to change the password.

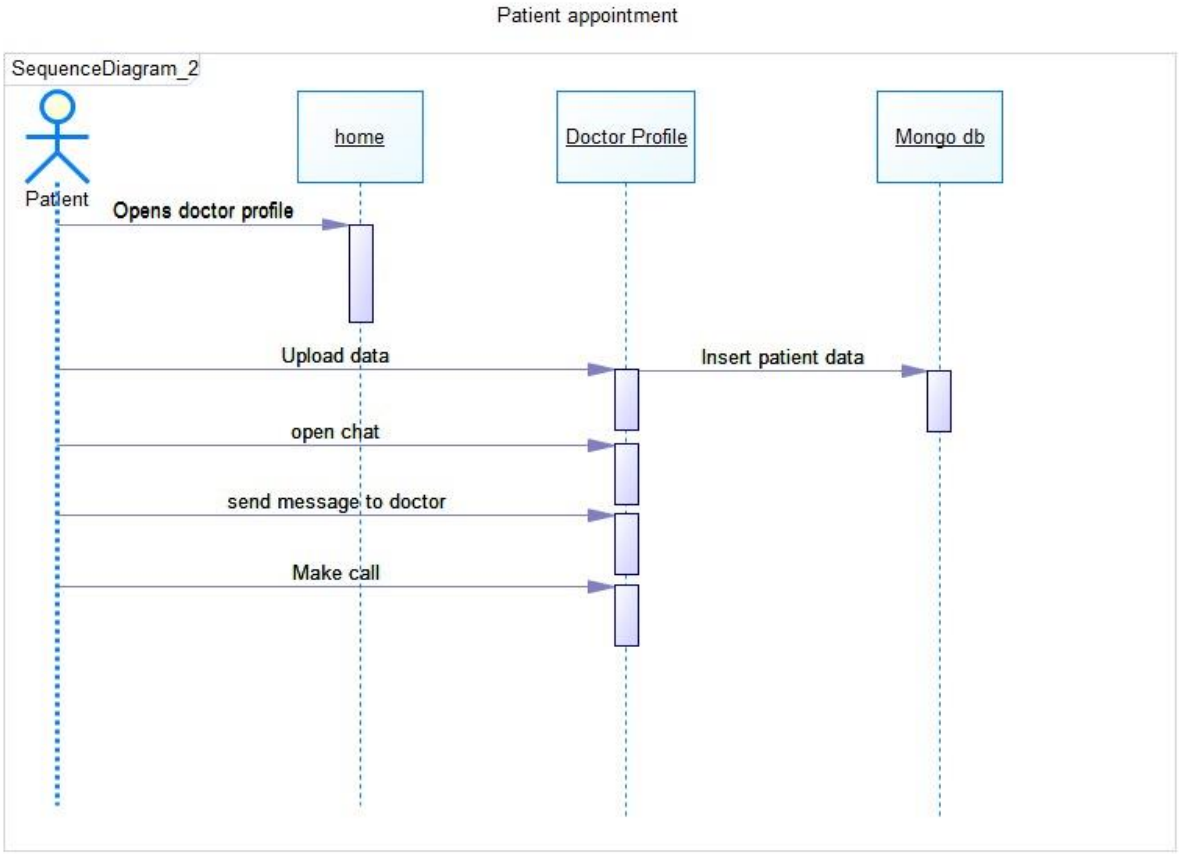


Figure 24 Sequence Diagram 2

Patient opens doctor profile from home page then upload data to enables doctor to check her case after that data is inserted in mongo DB.

Patient opens chat to communicate with the doctor .

Patient can send message to the doctor.

Patient can call the doctor.

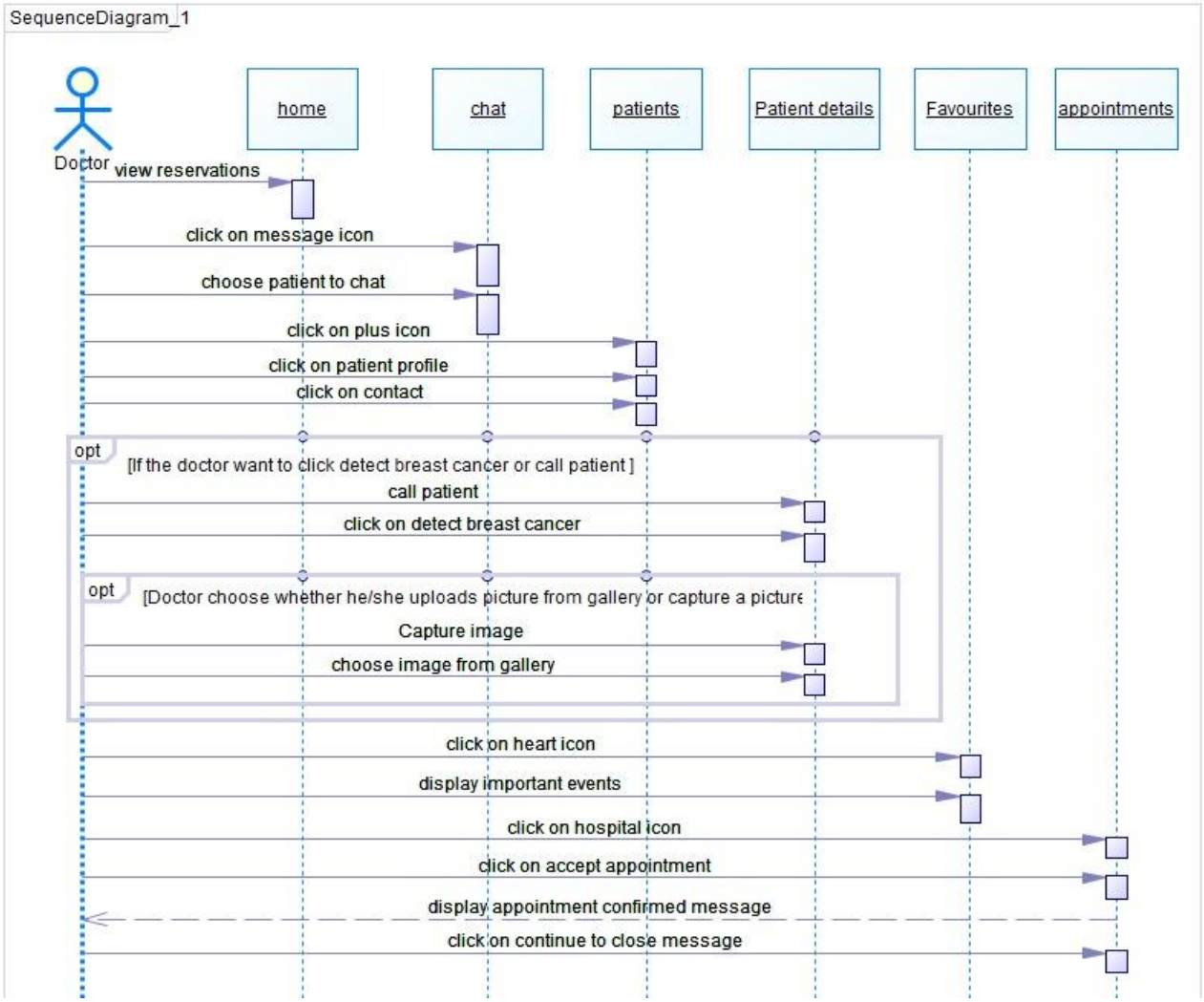


Figure 25 Sequence diagram 3

After doctor logged in his/her account home screen will appear which enable him to view reservations.

Doctor clicks on message icon and choose patient to chat with him or her .

Doctor clicks on plus icon and choose patient profile, Doctor clicks on contact button, if the doctor would like to capture image or choose image from gallery he/she should click on detect breast cancer button and if he/she would not, he/she can call the patient.

Doctor can see important events by clicking on heart icon.

Doctor can accept appointments from hospital icon then a confirmation message will appear after that he/she clicks on continue button to close the message.

CHAPTER 4 Analysis Phase

A use case diagram usually considered as a dynamic or behavior diagram in UML diagrams. It represents the functionality of any system with actors, use cases and arrows which describes the relationship between any use case and another use case whether it is include or extend [36].

It describes some actions (use cases).

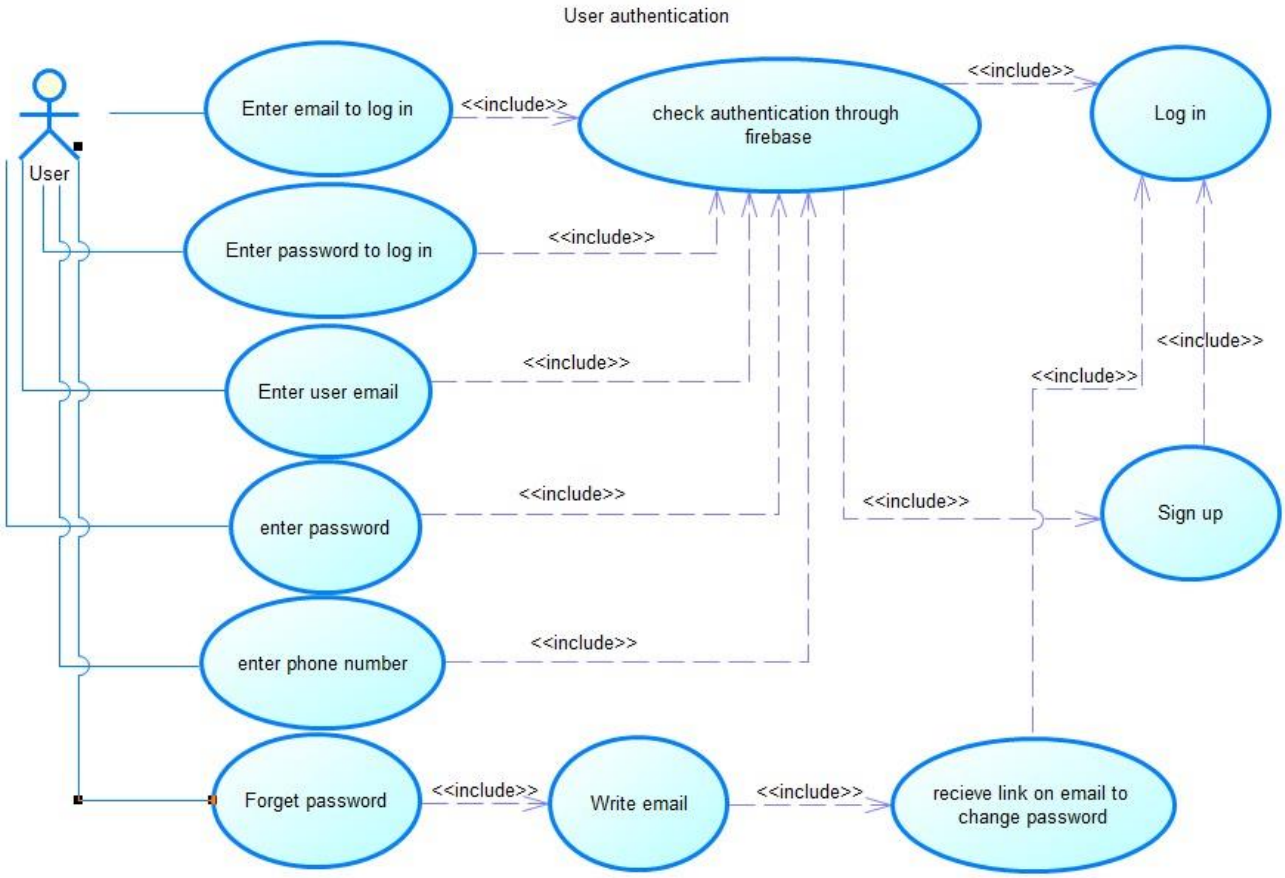


Figure 26 Use Case Diagram

First step, user enter email and password then the email and password are checked using fire-base authentication after that if the email and password are valid the user is logged in to his/her account.

If the user doesn't have an account he should enter email, password and phone number then email ,password and phone number are checked using fire-base authentication after that if the email, password and phone number are valid the user go to log in to account.

If the user forgot his/her password he /she should write his/her email then the user will receive a link on a message on his/her email to change the password and back to log in.

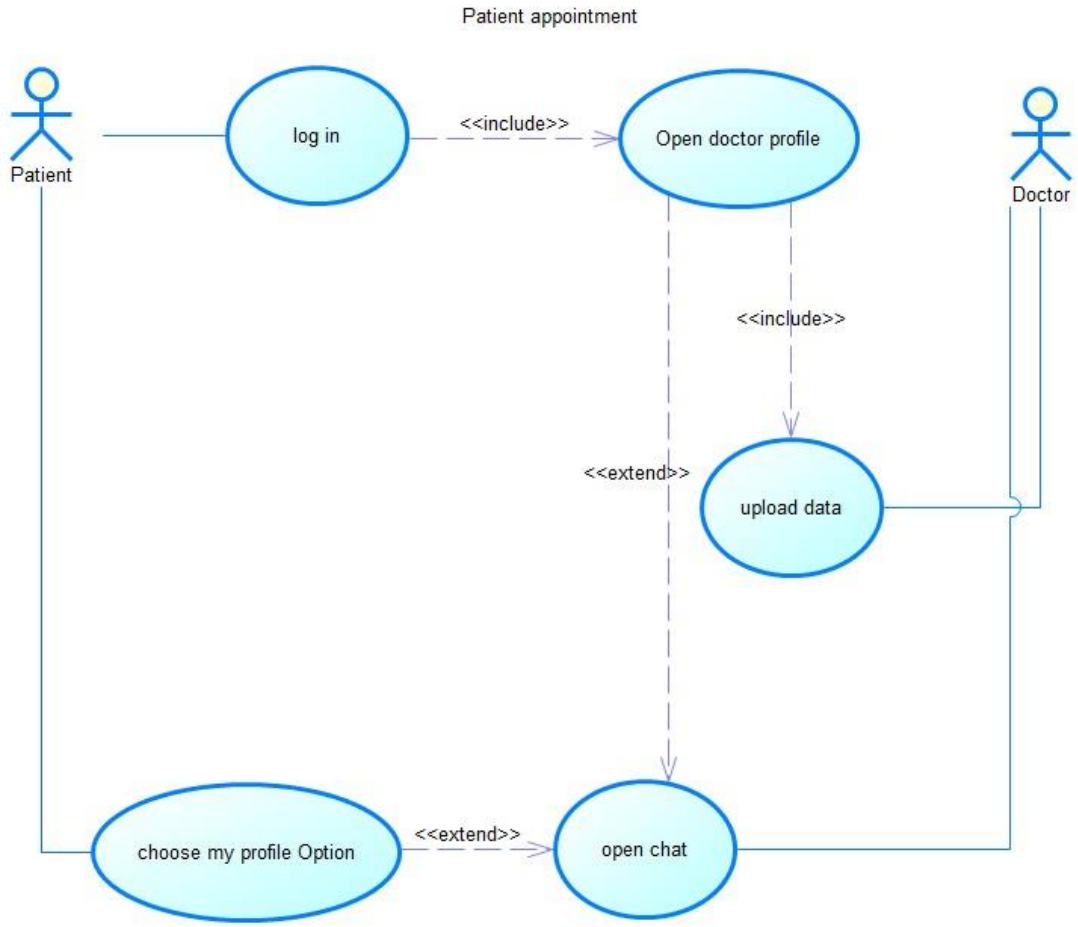


Figure 27 Use Case Diagram

After the patient logged in to his/her account, the patient opens the doctor profile patient upload data to enable the doctor to check her case or clicks on chat button to chat with the doctor and if the patient would like to communicate with the doctor ,patient choose from the options list my profile and clicks on chat option to communicate with the doctor.

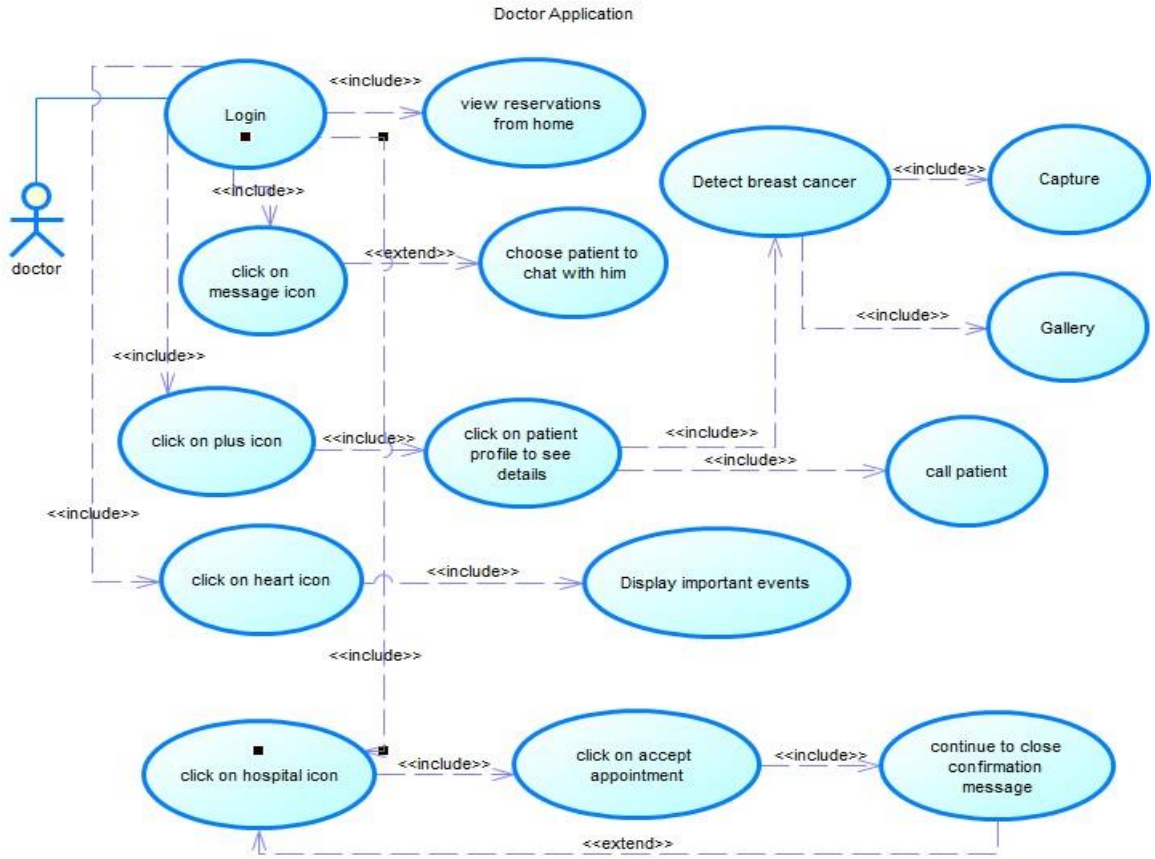


Figure 28 Use case

After doctor logged in to his/her account he/she can view reservations from home page.

After doctor logged in to his/her account he/she can choose patient to chat with him/her by clicking on message icon.

After doctor logged in to his/her account he/she can capture image or choose from gallery by clicking on detect breast cancer button all of this can be done after clicking on patient profile and also he/she can see patient details and call the patient all of this in the plus icon.

After doctor logged in to his/her account he/she can see important events by clicking on heart icon.

After doctor logged in to his/her account he/she can accept appointment from the hospital icon and click on continue to close the confirmation message.

Activity Diagram

The Unified Modeling Language includes several subsets of diagrams, including structure diagrams, interaction diagrams, and behavior diagrams. Activity diagrams, along with use case and state machine diagrams, are considered behavior diagrams because they describe what must happen in the system being modeled [39].

Benefits of activity diagrams

Activity diagrams present a number of benefits to users. Consider creating an activity diagram to [40]:

Demonstrate the logic of an algorithm.

Describe the steps performed in a UML use case.

Illustrate a business process or workflow between users and the system.

Simplify and improve any process by clarifying complicated use cases.

Model software architecture elements, such as method, function, and operation.

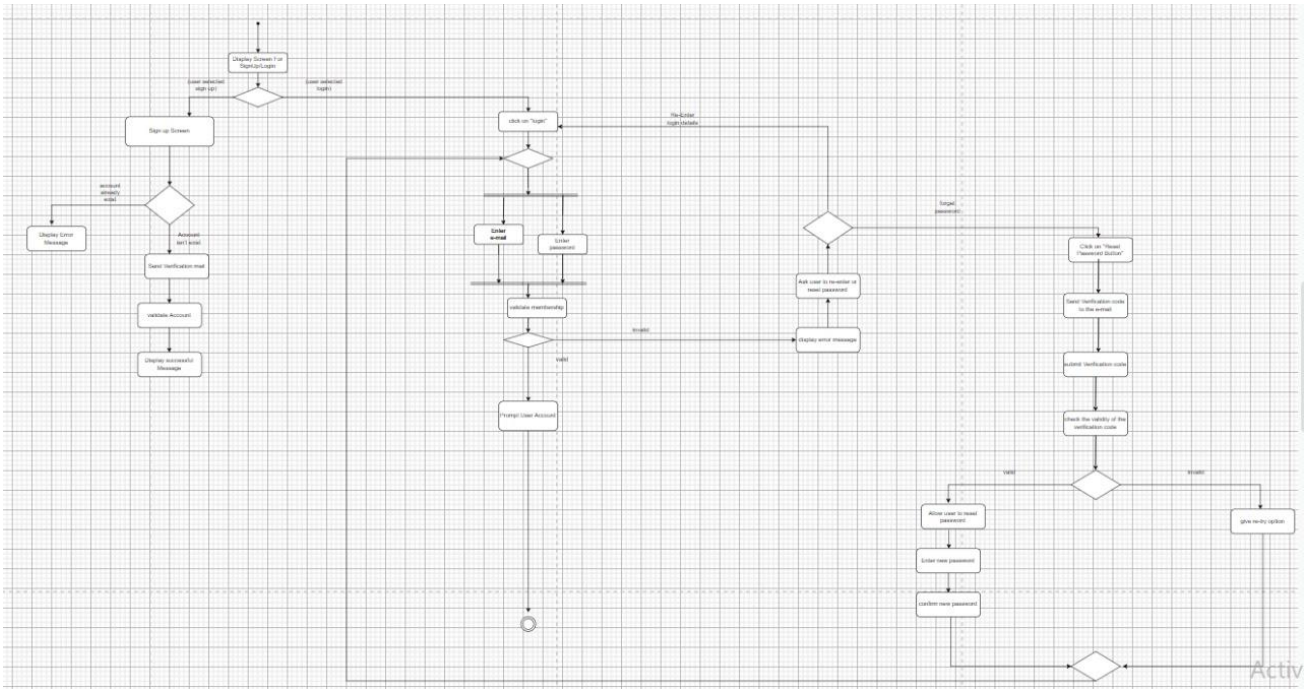


Figure 29 Activity diagram

First thing the user meet either the user is patient or a doctor a registration/login page the user selects on sign up screen in case the user hasn't account yet, the user starts to enter the email then the validation process starts. the email is checked if it is false Error Message will be displayed and if it is true verification mail is sent and validated account will be created then login screen appears for user to enter. The username, password and the phone number then the user account is prompt (this in case of registration)

Second case if the user already has an account on application the user will select login on the Sign up/Login Screen. The email, password and phone number are required. Then the user account gets through validation process to validate membership if it's valid then user account is prompt. In case user forgot the password. In case the account is invalid, display error message and Ask user to reenter or reset the password return back to login page to reenter the login details.

And in case the user forgot the account password, the user clicks on reset password button, send verification code to the email, submit the verification code and then check the validity of the verification code if it is valid Allow the user to reset password (enter new password and confirm new password) but if it is invalid give retry option and, in both cases, valid or invalid, we return to the login page

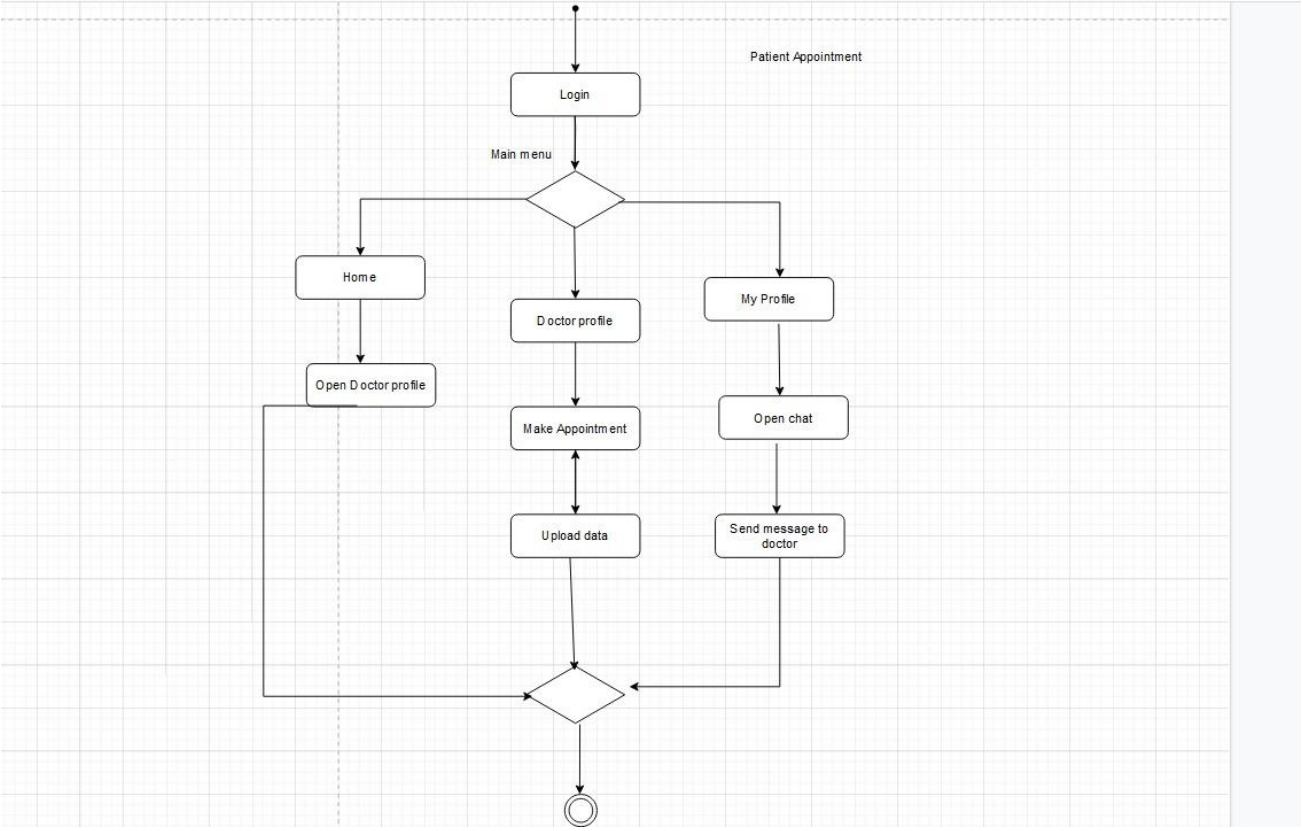


Figure 30 Activity diagram

Once the patient opens the application homepage using the account the user click on main menu list and many options appears to the user (Doctor profile, my profile (user profile), home). incase user want to Make an appointment the user click on doctor profile and make appointment with the doctor and then upload the data. And in case the user wants to talk to the doctor the patient clicks on my profile, open chat and then send message to doctor Once the patient open the application homepage using the account the user click on main menu list and many options appears to the user (Doctor profile, my profile (user profile), home). in case user want to Make an appointment the user click on doctor profile and make appointment with the doctor and then upload the data. And in case the user wants to talk to the doctor the patient clicks on my profile, open chat and then send message to doctor.

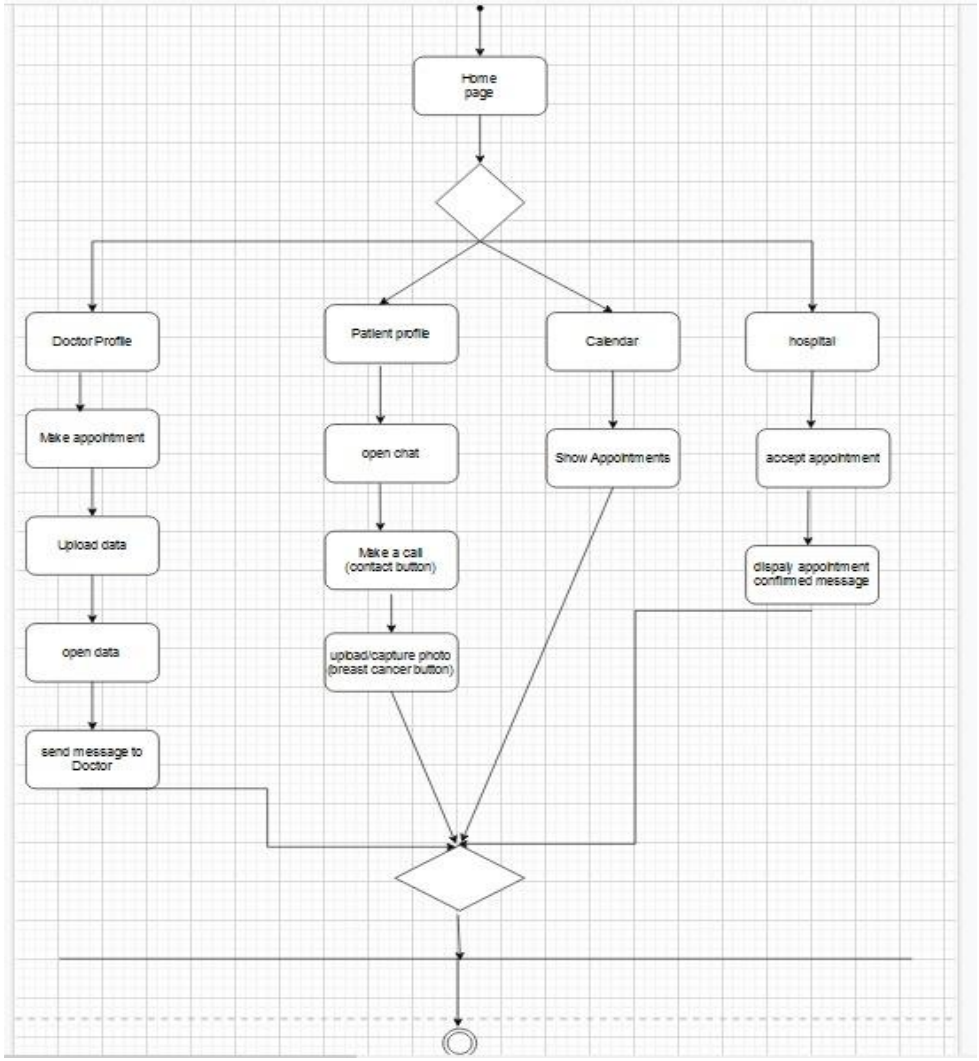


Figure 31 Activity diagram

Once the user (doctor or patient) open the application the user on homepage Click on main menu if the user is patient. The patient click on Doctor profile to make appointment and upload data also patient could upload or capture photo with the upload data . There is option to make chat with the doctor.while the user is a doctor , doctor can see the appointments from calendar and can contact with the patient, when patient upload images with upload data help doctor to detect if the cancer is malignant or begin images is showed and uploaded by clicking on breast cancer button. Also when patient choose the appointment . Hospital check it to see if this appointment is available or not if it's available confirmation message Is displayed.finally in the application there is a button called heart button to see favorite patients.

CHAPTER 5
Proposed breast cancer predication model

CHAPTER 5 Proposed breast cancer predication model

5.1 work flow

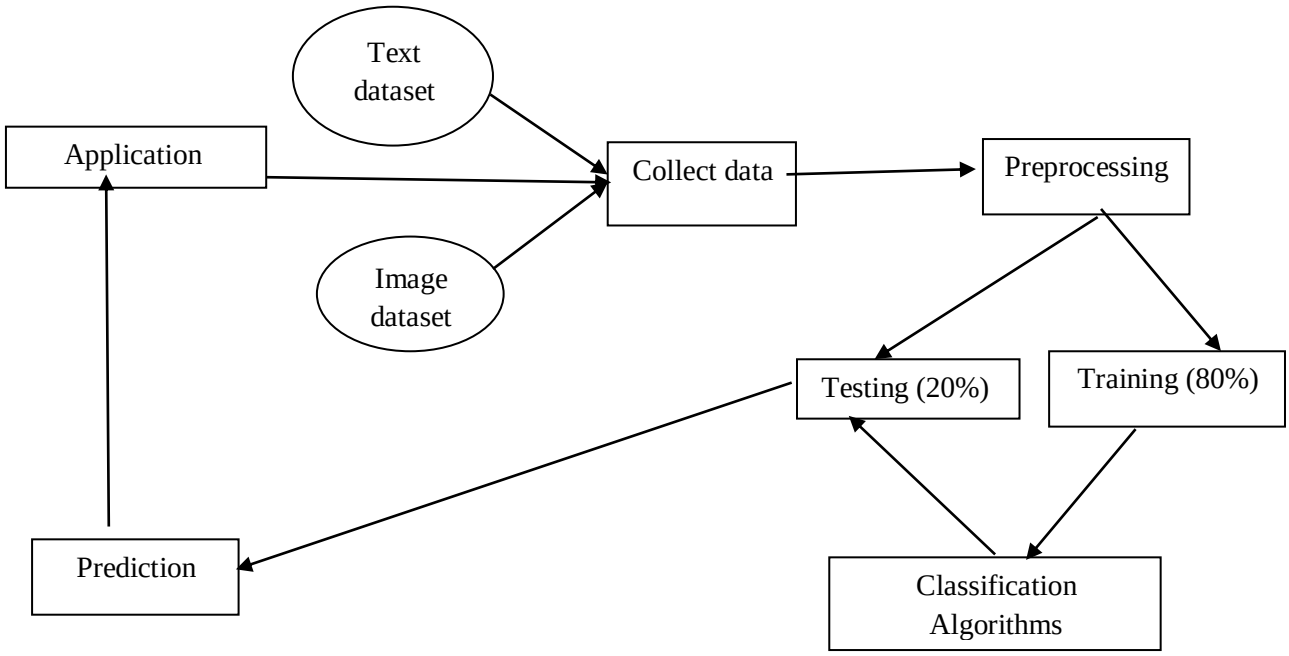


Figure 32 Work Flow

We collect data whether it is text or image then applying preprocessing such as checking null values , discretization and Normalization then splitting data into 80% Training and 20% Testing then using training data classify the data with classification algorithms (MLP, Back-Propagation, Logistic Regression, Decision Tree, Random Forest, KNN, CNN) then generate classification accuracy .Using classification accuracy to test the data then using classification accuracy and testing data to predict the result whether it is benign or malignant after that send the result to the doctor.

CHAPTER 6
Experimental design and result

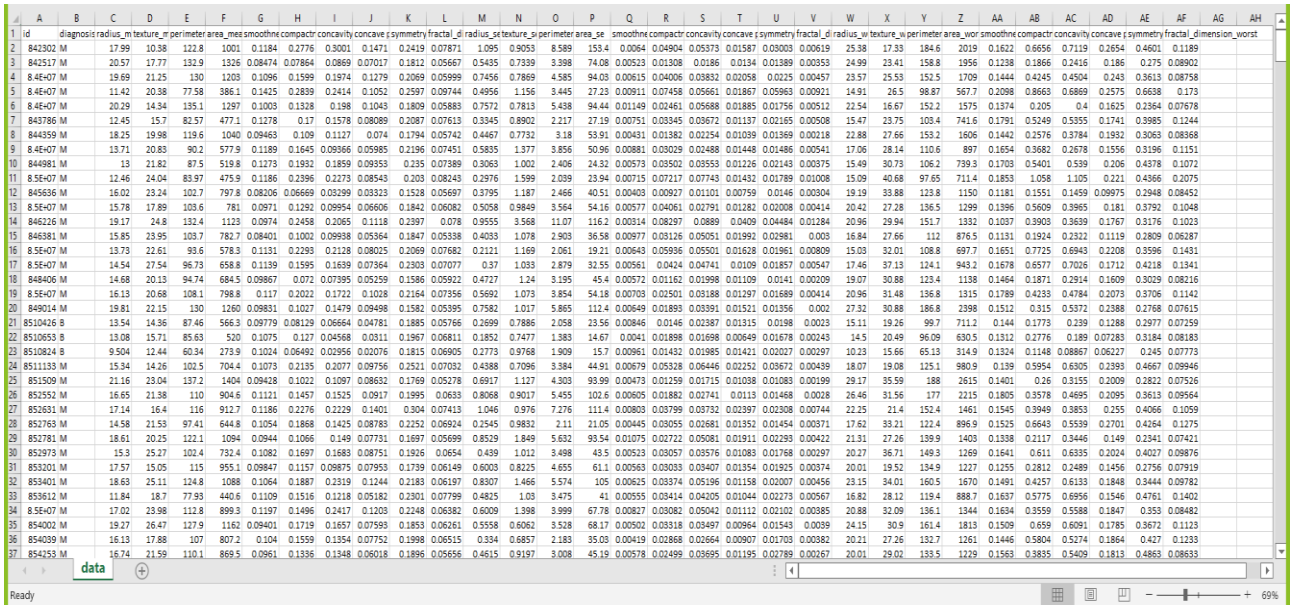
CHAPTER 6 Experimental design and result

6.1 Models Description

Dataset

Number of attributes: 32

Number of Rows: 569



id	diagnosis	radius_m	texture_m	perimeter	area_mea	smoothne	compactr	concavity	concave p	symmetry	fractal_di	radius_se	texture_se	perimeter	area_se	smoothne	compactr	concavity	concave p	symmetry	fractal_di	radius_wo	texture_wo	perimeter	area_wor	smoothne	compactr	concavity	concave p	symmetry	fractal_dimension_worst
842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871	1.095	0.9053	8.589	153.4	0.0064	0.04904	0.05373	0.01587	0.03003	0.00619	25.38	17.33	184.6	2019	0.1622	0.6656	0.7119	0.2654	0.4601	0.1189
842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.00523	0.01308	0.0186	0.0134	0.01389	0.00353	24.99	23.41	158.8	1956	0.1238	0.1866	0.2416	0.186	0.275	0.08902
8.4E+07	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.7869	4.585	94.03	0.00615	0.04006	0.03832	0.02058	0.0225	0.00457	23.57	25.53	152.5	1709	0.1444	0.4245	0.4504	0.243	0.3613	0.08758
8.4E+07	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.156	3.445	27.23	0.00911	0.07458	0.05661	0.01867	0.05963	0.00921	14.91	26.5	98.87	567.7	0.2098	0.8663	0.6869	0.2575	0.6638	0.173
8.4E+07	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.7813	5.438	94.44	0.01149	0.02461	0.05688	0.01885	0.01756	0.00512	22.54	16.67	152.2	1575	0.1374	0.205	0.4	0.1625	0.2364	0.07678
843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.8902	2.217	27.19	0.00751	0.03345	0.03672	0.01137	0.02165	0.00508	15.47	23.75	103.4	741.6	0.1791	0.5249	0.5355	0.1741	0.3985	0.1244
844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.7732	3.18	53.91	0.00431	0.01382	0.02254	0.01039	0.01369	0.00218	22.88	27.66	153.2	1606	0.1442	0.1576	0.3784	0.1932	0.3063	0.08368
8.4E+07	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451	0.5835	1.377	3.856	50.96	0.00881	0.03029	0.02488	0.01448	0.01486	0.00541	17.06	28.14	110.6	897	0.1654	0.3682	0.2678	0.1556	0.3196	0.1151
844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.002	2.406	24.32	0.00573	0.03502	0.03553	0.01226	0.02143	0.00375	15.49	30.73	106.2	739.3	0.1703	0.5401	0.539	0.206	0.4378	0.1072
8.5E+07	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243	0.2976	1.599	2.039	23.94	0.00715	0.07217	0.07743	0.01432	0.01789	0.01008	15.09	40.68	97.65	711.4	0.1853	1.058	1.105	0.221	0.4366	0.2075
845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697	0.3795	1.187	2.466	40.51	0.00403	0.00927	0.01101	0.00759	0.0146	0.00304	19.19	33.88	123.8	1150	0.1181	0.1551	0.1459	0.09975	0.2948	0.08452
8.5E+07	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842	0.06082	0.5058	0.9849	3.564	54.16	0.00577	0.04061	0.02791	0.01282	0.02008	0.00414	20.42	27.28	136.5	1299	0.1396	0.5609	0.3965	0.181	0.3792	0.1048
846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078	0.9555	3.568	11.07	116.2	0.00314	0.08297	0.0889	0.0409	0.04484	0.01284	20.96	29.94	151.7	1332	0.1037	0.3903	0.3639	0.1767	0.3176	0.1023
846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338	0.4033	1.078	2.903	36.58	0.00977	0.03126	0.05051	0.01992	0.02981	0.003	16.84	27.66	112	876.5	0.1131	0.1924	0.2322	0.1119	0.2809	0.06287
8.5E+07	M	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069	0.07682	0.2121	1.169	2.061	19.21	0.00643	0.05936	0.05501	0.01628	0.01961	0.00809	15.03	32.01	108.8	697.7	0.1651	0.7725	0.6943	0.2208	0.3596	0.1431
8.5E+07	M	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077	0.37	1.033	2.879	32.55	0.00561	0.0424	0.04741	0.0109	0.01857	0.00547	17.46	37.13	124.1	943.2	0.1678	0.6577	0.7026	0.1712	0.4218	0.1341
848406	M	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586	0.05922	0.4727	1.24	3.195	45.4	0.00572	0.01162	0.01998	0.01109	0.0141	0.00209	19.07	30.88	123.4	1138	0.1464	0.1871	0.2914	0.1609	0.3029	0.08216
8.5E+07	M	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	0.5692	1.073	3.854	54.18	0.00703	0.02501	0.03188	0.01297	0.01689	0.00414	20.96	31.48	136.8	1315	0.1789	0.4233	0.4784	0.2073	0.3706	0.1142
849014	M	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498	0.1582	0.05395	0.7582	1.017	5.865	112.4	0.00649	0.01893	0.03391	0.01521	0.01356	0.002	27.32	30.88	186.8	2398	0.1512	0.315	0.5372	0.2388	0.2768	0.07615
8510426	B	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05766	0.2699	0.7886	2.058	23.56	0.00846	0.0146	0.02387	0.01315	0.0198	0.0023	15.11	19.26	99.7	711.2	0.144	0.1773	0.239	0.1288	0.2977	0.07259
8510653	B	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967	0.06811	0.1852	0.7477	1.383	14.67	0.0041	0.01898	0.01698	0.00649	0.01678	0.00243	14.5	20.49	96.09	630.5	0.1312	0.2776	0.189	0.07283	0.3184	0.08183
8510824	B	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076	0.1815	0.06905	0.2773	0.9768	1.909	15.7	0.00961	0.01432	0.01985	0.01421	0.02027	0.00297	10.23	15.66	65.13	314.9	0.1324	0.1148	0.08867	0.06227	0.245	0.07773
8511133	M	15.34	14.26	102.5	704.4	0.1073	0.2135	0.2077	0.09756	0.2521	0.07032	0.4388	0.7096	3.384	44.91	0.00679	0.05328	0.06446	0.02252	0.03672	0.00439	18.07	19.08	125.1	980.9	0.139	0.5954	0.6305	0.2393	0.4667	0.09946
851509	M	21.16	23.04	137.2	1404	0.09428	0.1022	0.1097	0.08632	0.1769	0.05278	0.6917	1.127	4.303	93.99	0.00473	0.01259	0.01715	0.01038	0.01083	0.00199	29.17	35.59	188	2615	0.1401	0.26	0.3155	0.2009	0.2822	0.07526
852552	M	16.65	21.38	110	904.6	0.1121	0.1457	0.1525	0.0917	0.1995	0.0633	0.8068	0.9017	5.455	102.6	0.00605	0.01882	0.02741	0.0113	0.01468	0.0028	26.46	31.56	177	2215	0.1805	0.3578	0.4695	0.2095	0.3613	0.09564
852631	M	17.14	16.4	116	912.7	0.1186	0.2276	0.2229	0.1401	0.304	0.07413	1.046	0.976	7.276	111.4	0.00803	0.03799	0.03732	0.02397	0.02308	0.00744	22.25	21.4	152.4	1461	0.1545	0.3949	0.3853	0.255	0.4066	0.1059
852763	M	14.58	21.53	97.41	644.8	0.1054	0.1868	0.1425	0.08783	0.2252	0.06924	0.2545	0.9832	2.11	21.05	0.00445	0.03055	0.02681	0.01352	0.01454	0.00371	17.62	33.21	122.4	896.9	0.1525	0.6643	0.5539	0.2701	0.4264	0.1275
852781	M	18.61	20.25	122.1	1094	0.0944	0.1066	0.149	0.07731	0.1697	0.05699	0.8529	1.849	5.632	93.54	0.01075	0.02722	0.05081	0.01911	0.02293	0.00422	21.31	27.26	139.9	1403	0.1338	0.2117	0.3446	0.149	0.2341	0.07421
852973	M	15.3	25.27	102.4	732.4	0.1082	0.1697	0.1683	0.08751	0.1926	0.0654	0.439	1.012	3.498	43.5	0.00523	0.03057	0.03576	0.01083	0.01768	0.00297	20.27	36.71	149.3	1269	0.1641	0.611	0.6335	0.2024	0.4027	0.09876
853201	M	17.57	15.05	115	955.1	0.09847	0.1157	0.09875	0.07953	0.1739	0.06149	0.6003	0.8225	4.655	61.1	0.00563	0.03033	0.03407	0.01354	0.01925	0.00374	20.01	19.52	134.9	1227	0.1255	0.2812	0.2489	0.1456	0.2756	0.07919
853401	M	18.63	25.11	124.8	1088	0.1064	0.1887	0.2319	0.1244	0.2183	0.06197	0.8307	1.466	5.574	105	0.00625	0.03374	0.05196	0.01158	0.02007	0.00456	23.15	34.01	160.5	1670	0.1491	0.4257	0.6133	0.1848	0.3444	0.09782
853612	M	11.84	18.7	77.93	440.6	0.1109	0.1516	0.1218	0.05182	0.2301	0.07799	0.4825	1.03	3.475	41	0.00555	0.03414	0.04205	0.01044	0.02273	0.00567	16.82	28.12	119.4	888.7	0.1637	0.5775	0.6956	0.1546	0.4761	0.1402
8.5E+07	M	17.02	23.98	112.8	899.3	0.1197	0.1496	0.2417	0.1203	0.2248	0.06382	0.6009	1.398	3.999	67.78	0.00827	0.03082	0.05042	0.01112	0.02102	0.00385	20.88	32.09	136.1	1344	0.1634	0.3559	0.5588	0.1847	0.353	0.08482
854002	M	19.27	26.47	127.9	1162	0.09401	0.1719	0.1657	0.07593	0.1853	0.06261	0.5558	0.6062	3.528	68.17	0.00502	0.03318	0.03497	0.00964	0.01543	0.0039	24.15	30.9	161.4	1813	0.1509	0.659	0.6091	0.1785	0.3672	0.1123
854039	M	16.13	17.88	107	807.2	0.104	0.1559	0.1354	0.07752	0.1998	0.06515	0.334	0.6857	2.183	35.03	0.00419	0.02868	0.02664	0.00907	0.01703	0.00382	20.21	27.26	132.7	1261	0.1446	0.5804	0.5274	0.1864	0.427	0.1233
854253	M	16.74	21.59	110.1	869.5	0.0961	0.1336	0.1348	0.06018	0.1896	0.05656	0.4615	0.9197	3.008	45.19	0.00578	0.02499	0.03695	0.01195	0.02789	0.00267	20.01	29.02	133.5	1229	0.1563	0.3835	0.5409	0.1813	0.4863	0.08633

Figure 33 Text dataset

Discretized dataset

Updated dataset after applying discretization on permitter mean column

Number of attributes: 32

Number of Rows: 569

CHAPTER 6 Experimental design and result

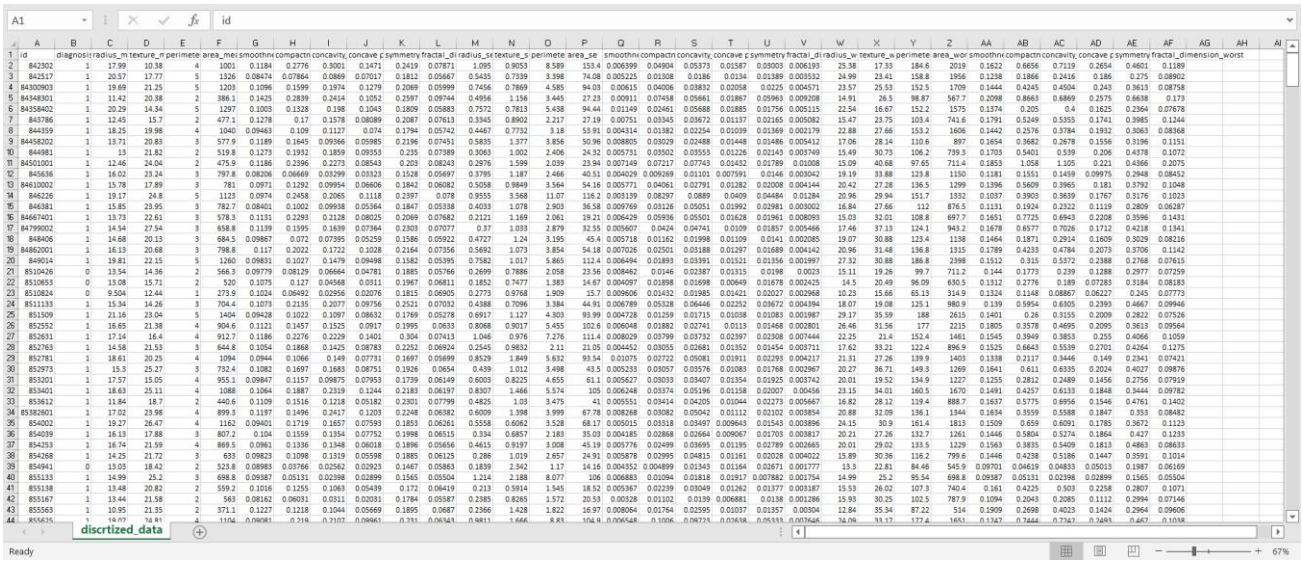


Figure 34 Discretized dataset

Normalized dataset

Updated dataset after applying normalization on it

Number of attributes: 30

Number of Rows: 569

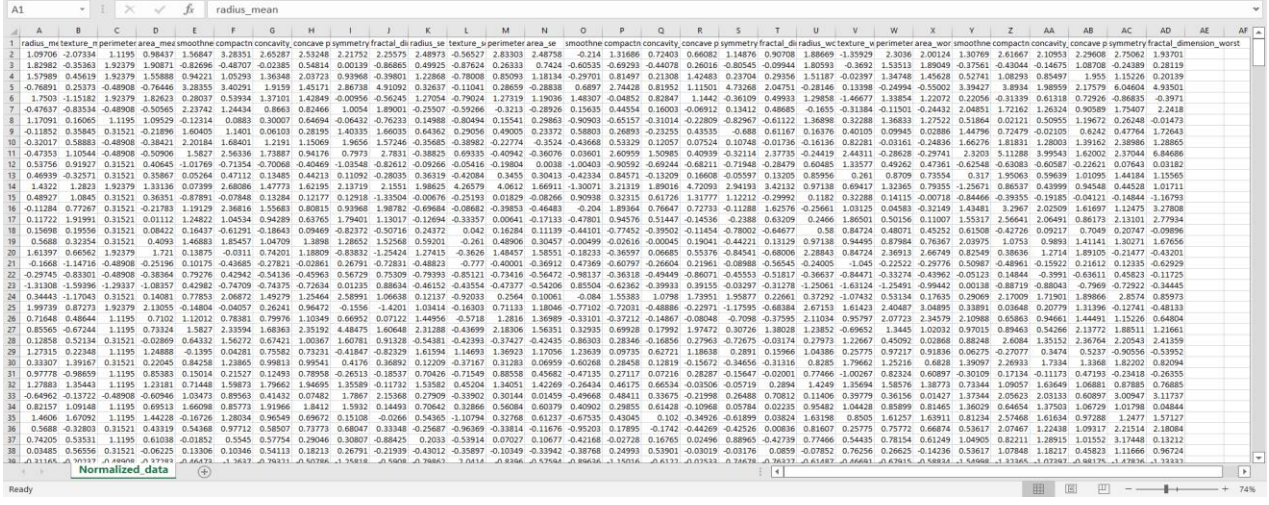


Figure 35 Normalized dataset

CHAPTER 6 Experimental design and result

Image Dataset

First Image dataset the data collected at baseline include breast ultrasound images among women in ages between 25 and 75 years old. This data was collected in 2018. The number of patients is 600 female patients. The dataset consists of 780 images with an average image size of 500*500 pixels. The images are in PNG format. The ground truth images are presented with original images. The images are categorized into three classes, which are normal, benign, and malignant.

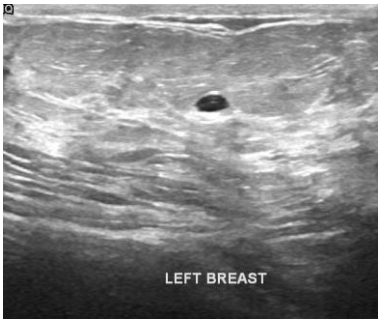


Figure 36 **Benign**

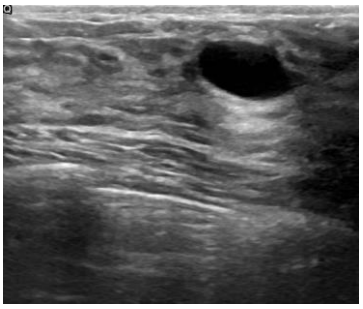


Figure 37 **Malignant**

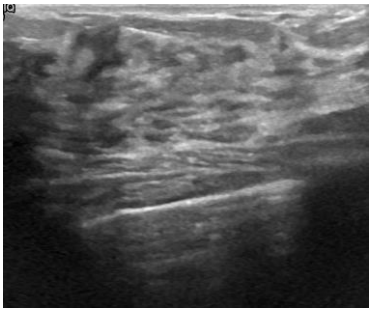


Figure 38 **Normal**

6.2Machine Learning

Neural Network algorithms

-The Multilayer Perceptron (MLP) learning procedure is as follows:

-Starting with the input layer, propagate data forward to the output layer. This step is the forward propagation.

-Based on the output, calculate the error (the difference between the predicted and known outcome). The error needs to be minimized.

-Back-propagate the error. Find its derivative with respect to each weight in the network, and update the model.

-Repeat the three steps given above over multiple epochs to learn ideal weights.

-Finally, the output is taken via a threshold function to obtain the predicted class labels.

*MLP trains using Stochastic Gradient Descent, Adam or L-BFGS. Stochastic Gradient Descent (SGD) updates parameters using gradient of loss function with respect to parameter that needs adaptation., i.e.

-Loss Function: $w \leftarrow w - \eta (\propto(\partial R(w))/(\partial w) + (\partial Loss)/(\partial w))$

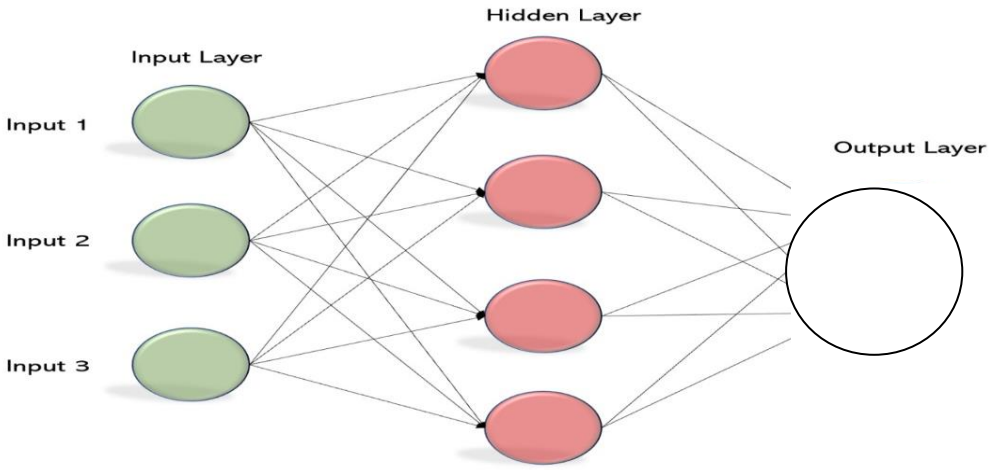


Figure 39 MLP

CHAPTER 6 Experimental design and result

The back-propagation learning algorithm can be divided into two phases:

1-propagation

2-weight update.

Each propagation involves the following steps:

Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.

Backward propagation of the propagation's output activations through the neural network using the training pattern target in order to generate the deltas of all output and hidden neurons.

Loss Function = ∑ ½ (Desired Output - Predicted Output) ^2

Here, ½ is a constant used.

After got the loss function, now we should minimize the loss to have more accuracy (we have to decrease or minimize the value of the loss function to zero).

To update the weights, we should follow the following steps:

1-value of weights

2- Update W(init) in the direction of minimizing gradient(slope), when it reaches the local minima, it would have decreased its error.

3-Update the weights with this formula
Updated weight -> weight - ∝ * ∂ (error)/∂ (weight)

∝ = Learning rate
Repeat 3 until local minima.

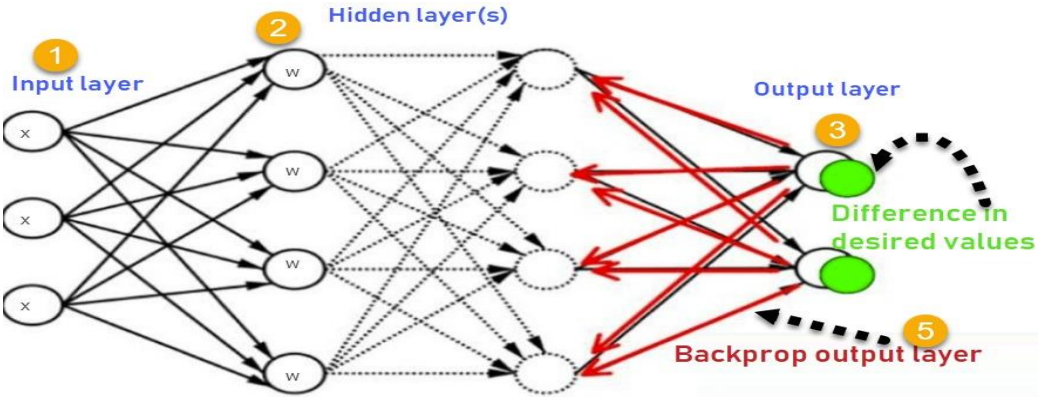


Figure 40 Back propagation

Advantages of the dense net:

Parameter efficiency: Every layer adds only a limited number of parameters.

Implicit deep supervision: Improved flow of gradient through the network.

Feature maps in all layers have direct access to the loss function and its gradient.

The important concept in DenseNet:

Growth rate: This determines the number of feature maps output into individual layers inside dense blocks.

Dense connectivity: By dense connectivity, we mean that within a dense block each layer gets us input feature maps from the previous layer

Composite functions: So, the sequence of operations inside a layer goes as follows. So, we have batch normalization, followed by an application of Relu, and then a convolution layer (that will be one convolution layer)

Transition layers: The transition layers aggregate the feature maps from a dense block and reduce its dimensions.

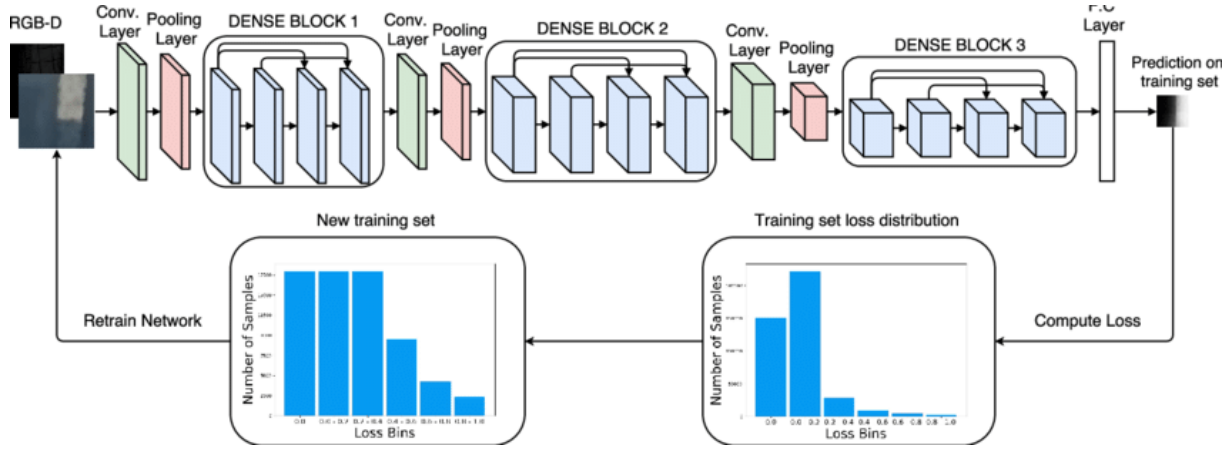


Figure 41 Dense net architecture

CHAPTER 6 Experimental design and result

CNN is used for images as:

Some patterns are smaller than the whole image.

The neuron does not have to see the whole image to discover the pattern.

Connecting to small region with less parameters

Subsampling the pixels will not change the object

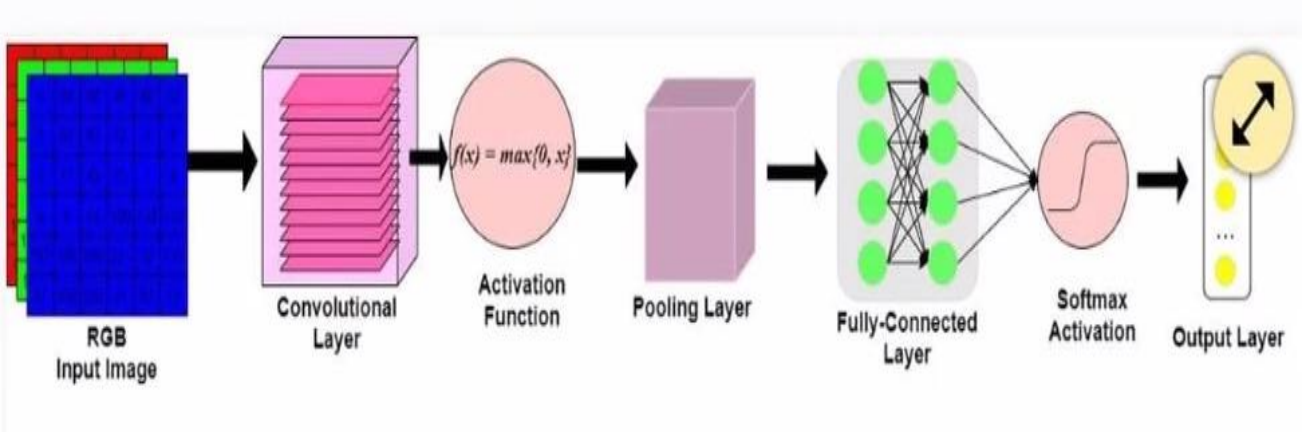


Figure 42 CNN architecture

A typical CNN model looks like this:

Input layer:

The input is image and may be Grayscale or RGB. Images are consisting of pixels that are in range from 0 to 255. Normalize them is essential to convert the range of every image between 0 to 1 before applying the model on it.

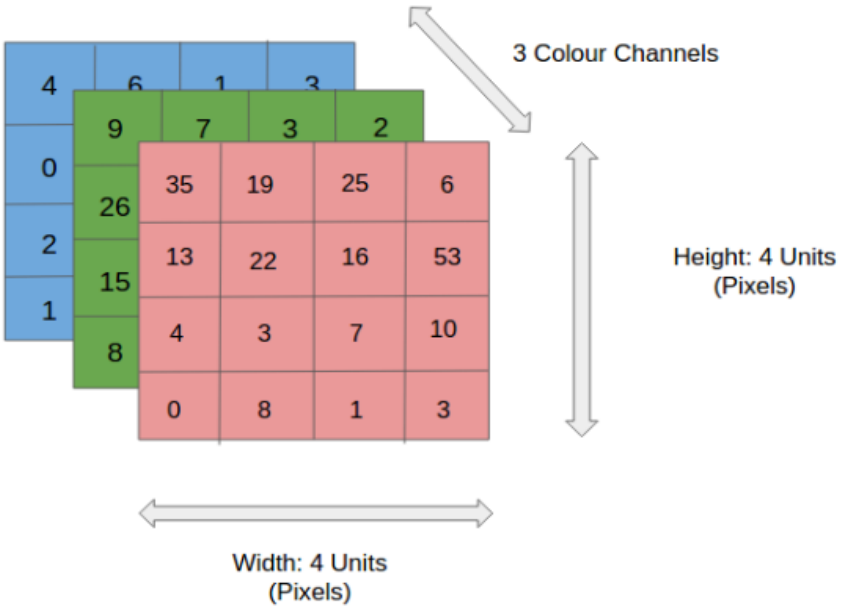


Figure 43 Conv matrix

Convolution layer + Activation Function

The convolution layer the first step in the process of extracting valuable features from an image to make feature extraction. A filter is applied to the image multiple times and creates a feature map which helps in classifying the input image.

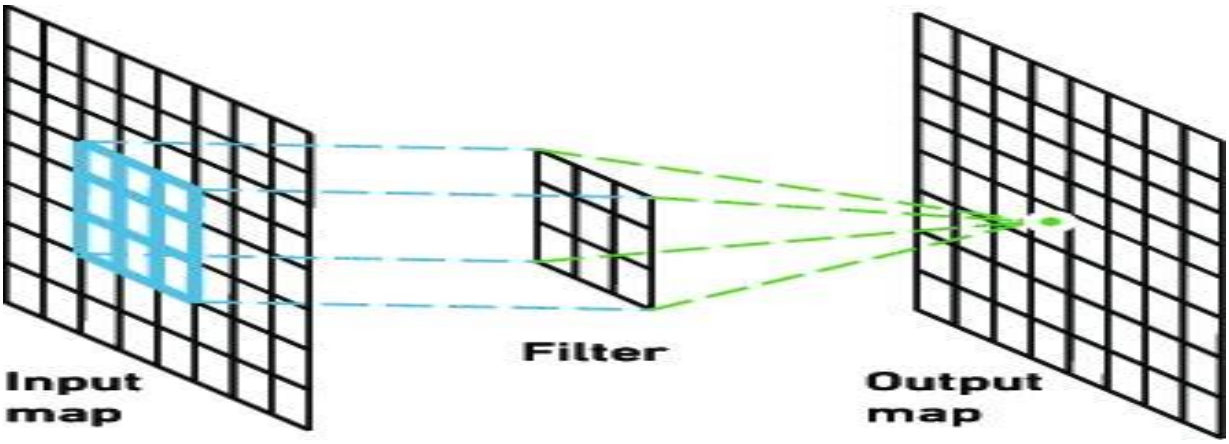


Figure 44 Convolution layer

The activation functions

SoftMax is numeric output of the last linear layer of a multi-class classification neural network

Pooling layer

This layer is applied after Convolution layer it take image from the Convolution layer and downsizing dimension of the image. It keeps the important features and reduces the computation time.

It consists of Max pooling, Average pooling and Summation pooling.

Max pooling:

It separates the matrix into sub matrices and take the maximum value from each sub matrix.

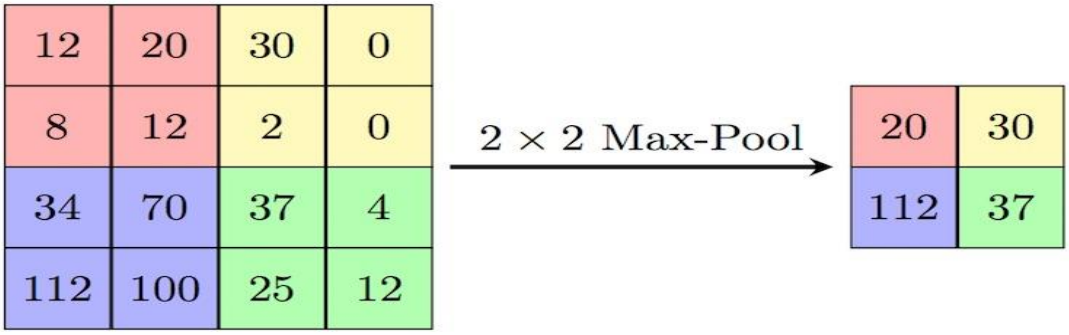


Figure 45 Max pooling

Average pooling:

It separates the matrix into sub matrices and take the average from each sub matrix.

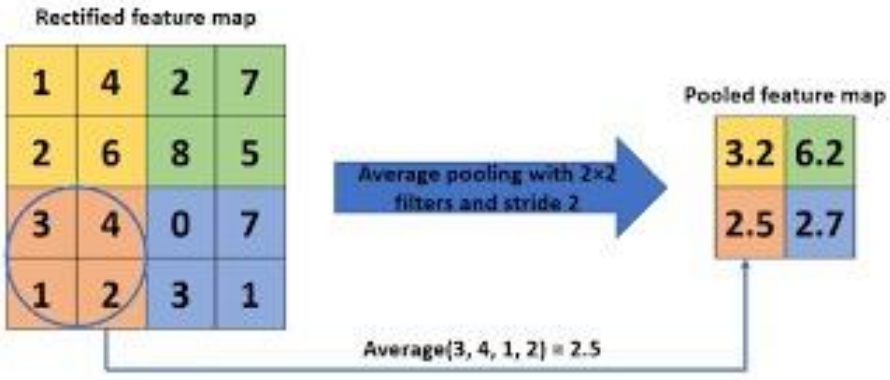


Figure 46 Average pooling

Summation pooling:

It separates the matrix into sub matrices and take the sum of each sub matrix.

Flatten layer:

It converts the 2D matrix into vector to use as an input in the ANN.

Fully Connected: The Fully connected layer (as we have in ANN) is used for classifying the input image into a label. This layer connects the information extracted from the previous steps (i.e., Convolution layer and Pooling layers) to the output layer and eventually classifies the input into the desired label.

Ensemble learning:

The idea of ensemble classification is to learn not just one classifier but a set of classifiers, and them to combine their predictions for the classification of unseen instances using some form of voting.

There are several methods to combine the outputs of the various classifiers.

Voting or averaging of predictions of multiple trained models.

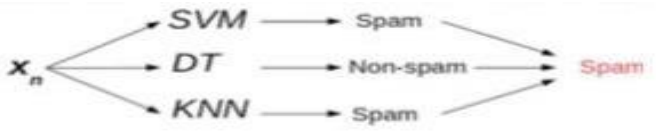


Figure 47 Ensemble

Stacking: Use predictions of multiple models as “features” to train a new model and use the new model to make predictions on test data.

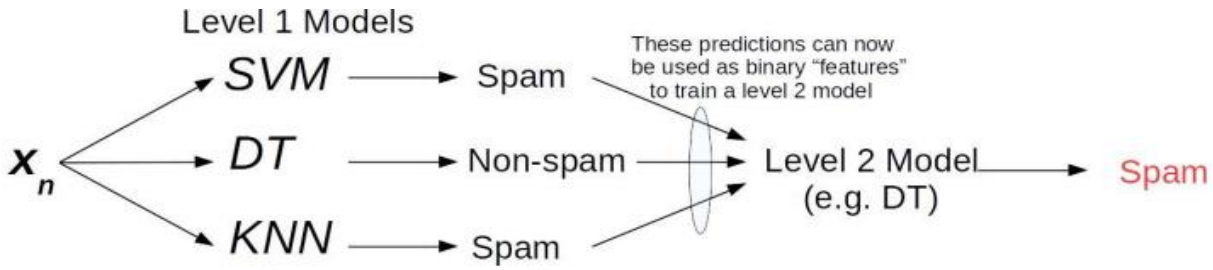


Figure 48 Stacking

Machine learning algorithms

In logistic regression, we pass the weighted sum of inputs through an activation function that can map values in between 0 and 1. Such activation function is known as sigmoid function.

The equation of logistic regression:

Logit(p) = log(p/1-p)

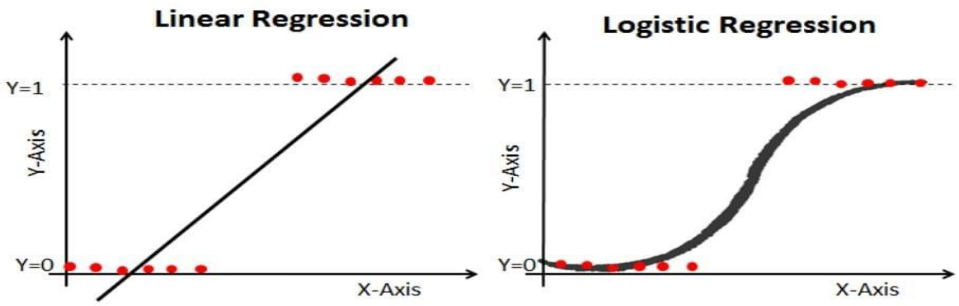


Figure 49 difference between logistic and linear regression

Advantages of the Decision Tree:

It can be very useful for solving decision-related problems.

There is less requirement of data cleaning compared to other algorithms

Disadvantages of the Decision Tree

The decision tree contains lots of layers, which makes it complex.

It may have an overfitting issue, which can be resolved using the Random Forest algorithm.

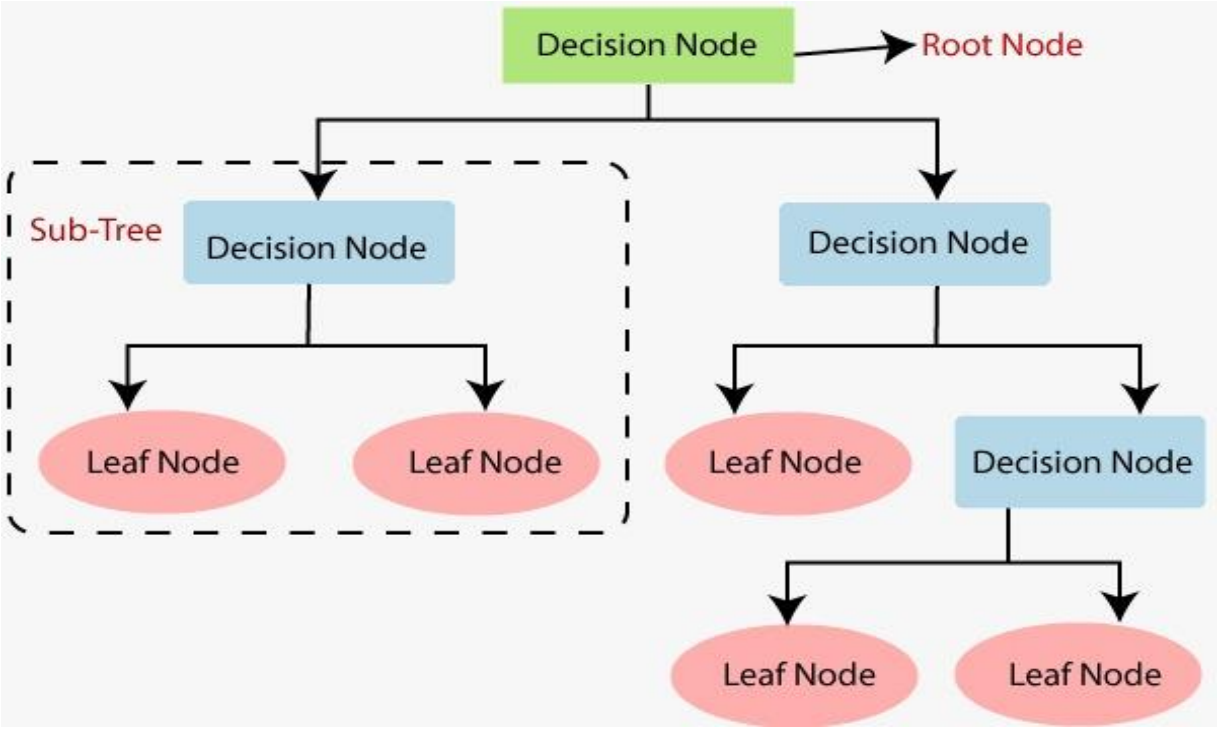


Figure 50 decision tree

The KNN working can be explained:

1- Select the number K of the neighbours

2-Calculate the distance of K number of neighbours

3- Take the K nearest neighbours as per the calculated distance.

4- Among these k neighbours, count the number of the data points in each category.

5- Assign the new data points to that category for which the number of the neighbour is maximum.

Advantages of KNN Algorithm:

It is simple to implement.

It can be more effective if the training data is large.

Disadvantages of KNN Algorithm:

Always needs to determine the value of K which may be complex some time.

The computation cost is high because of calculating the distance between the data points for all the training samples.

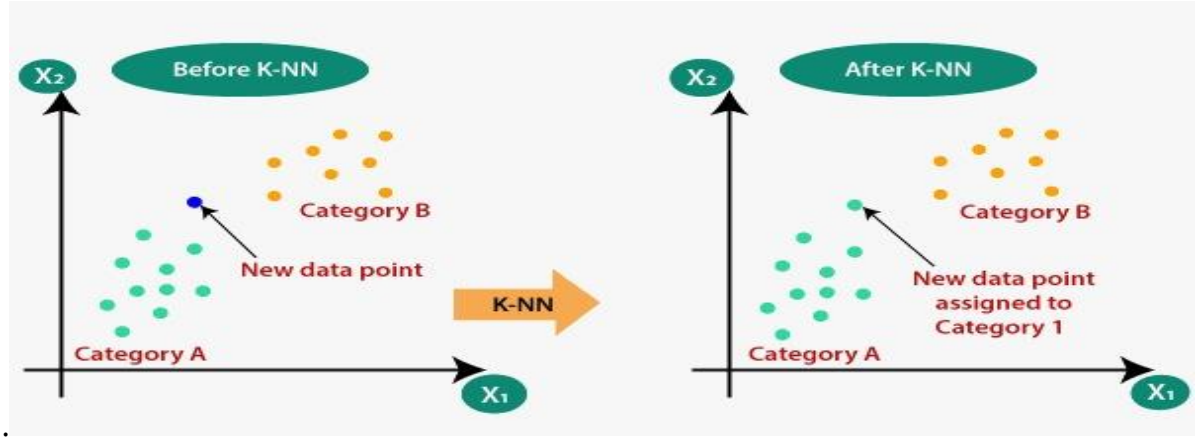


Figure 51 KNN

Advantages:

The algorithm can be used in both classification and regression problems.

Random forests can also handle missing values. There are two ways to handle these: using median values to replace continuous variables, and computing the proximity-weighted average of missing values.

It reduces overfitting in decision trees and helps to improve the accuracy.

Disadvantages:

Random forests are slow in generating predictions because it has multiple decision trees. Whenever it makes a prediction, all the trees in the forest have to make a prediction for the same given input and then perform voting on it. This whole process is time-consuming.

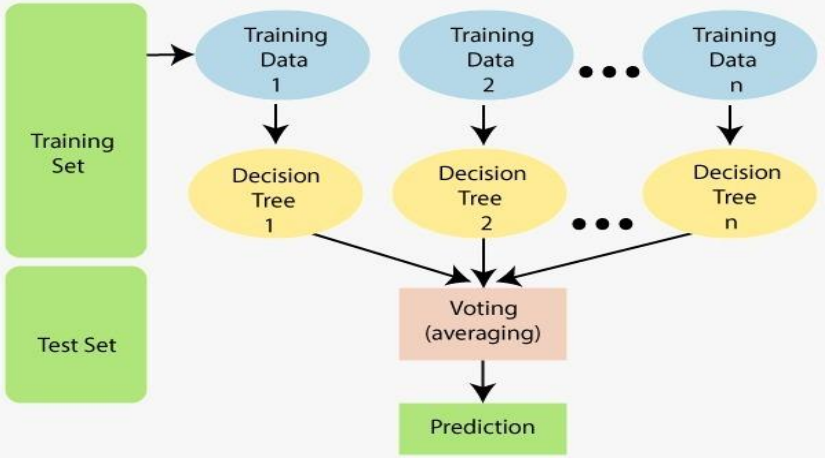


Figure 52 Random Forest

6.3Flutter

Our GUI contains web and mobile applications, we have made a simple web application for phase 1. First screen is the welcome page then the sign-in page which contains two types of registration (1-sign-in as patient) (2-Sign-in as doctor), the user should register his name, mail and password. But if the user already has an account, he writes his mail and password in the Login page.

After the user registers his information the first screen which displayed inside the application is the main page which contains a several features, the first feature is the home in the task bar which contains the breast cancer information button which takes the user to another page named Breast cancer information contains all the information about Breast cancer as overview, types and risk factors …etc.

Another feature in the application is the timetable which displays all the appointments of the doctor, it allows doctors to put all appointments. Then the patients profile page displays all the profiles of the patients to see the profile of patient, you should click on View profile button which take the user to patient profile details screen, it displays all information about the patient and his sick condition as his X-Ray, in the same page the doctor can comment and start chatting with the

patient by clicking on chat button, it takes him to the chat directly which can communicate with the patients.

In the mobile application, first screen the splash screen which displays an animation "welcome to your app" after a several seconds. The second screen is a splash screen contains a continue button which moves the user to sign in screen. In sign screen, the patient fills his mail and password. If the patient do not have an account ,he will move to sign up screen which he have to fill his name , phone number , mail and password . After that the user have to fill the sign in data in case of forgetting password the user will data in the forgetting password screen and will receive a support mail including the new password.

After registration, the home screen includes a category which displays some application features, the top-rated doctors and the search of the application. Each doctor has his own profile, the user can choose. After choosing the doctor and opening his profile. The doctor's profile displays the doctor's review, total score and satisfaction and more information about the doctor. The patient can communicate and make an appointment with the doctor. The patient can upload his data in csv to list screen by choosing the file and clicking on insert data. By clicking back to home, the patient returns back to the home screen. Patients can open them profile on the home screen. The profile screen includes the patient profile picture which he can change it at any time, Notifications, Settings, Help Center, Log out and the chat. The chat screen includes the conversations with the doctors.

In doctor's application, the first screen is the sign in screen which the doctor has to fill his mail and password. If the doctor does not have an account, he has to fill the data in the sign up which is his name, phone number and password. In case of forgetting password, the doctor has to go to forget screen and fill his mail and wait for the support mail including his old password. After the registration process, the first screen will face the doctor is the home screen which includes the features of the application, the doctor can open the chat and communicate with the patients, he can open the patient profile screen which contains the profiles of the patients. Each patient has a profile includes information about her and her medical case and the incoming appointment. The doctor also can receive and upload images. The application includes a calendar which contains all the appointments of the dr. and after accepting and appointment, the doctor receives an approved message.

CHAPTER 6 Experimental design and result

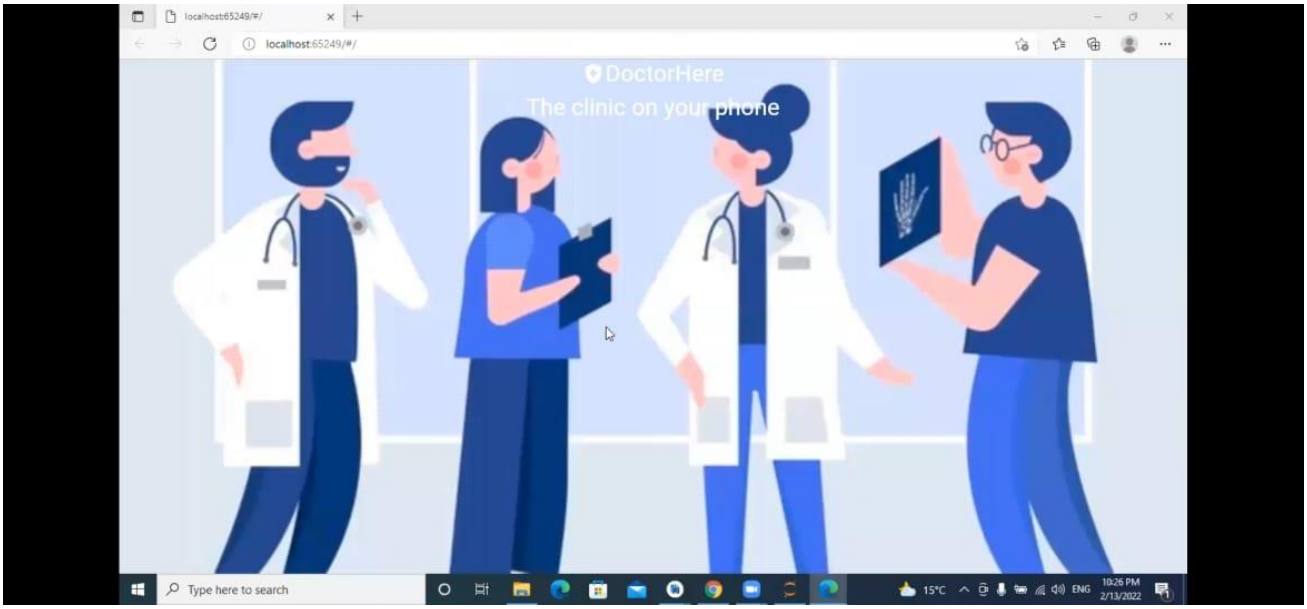


Figure 53 Welcome page

-It includes the welcome page which is the first screen as shown in the figure

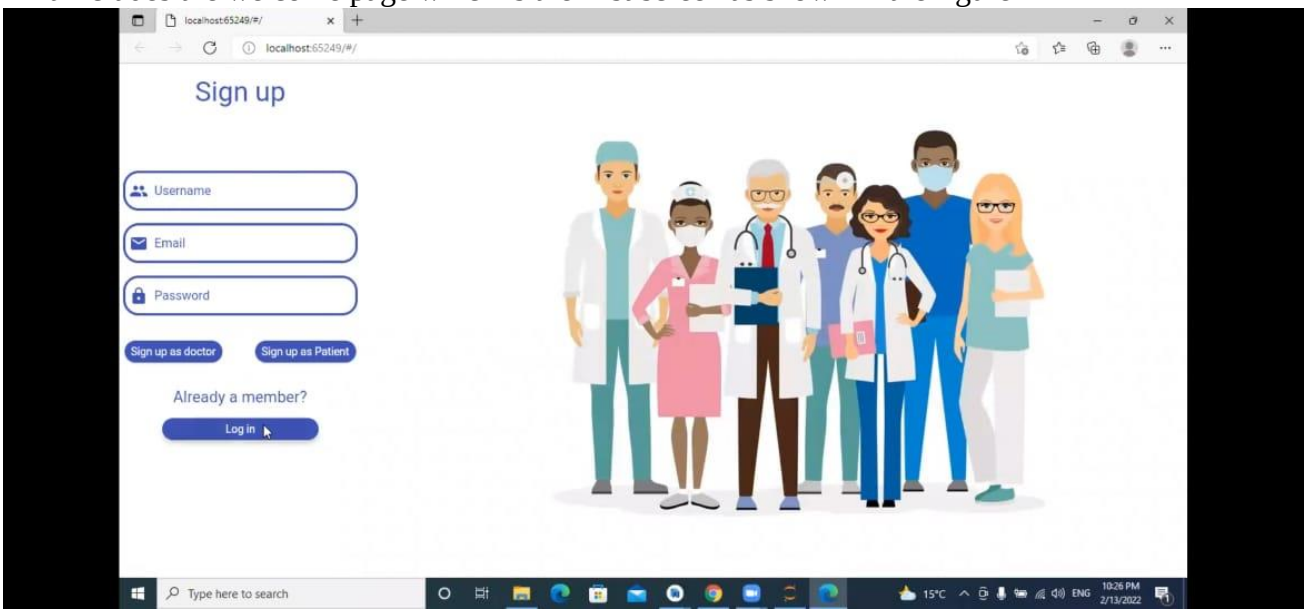


Figure 54 Sign up page

-The second screen is the sign-up page as shown in the figure

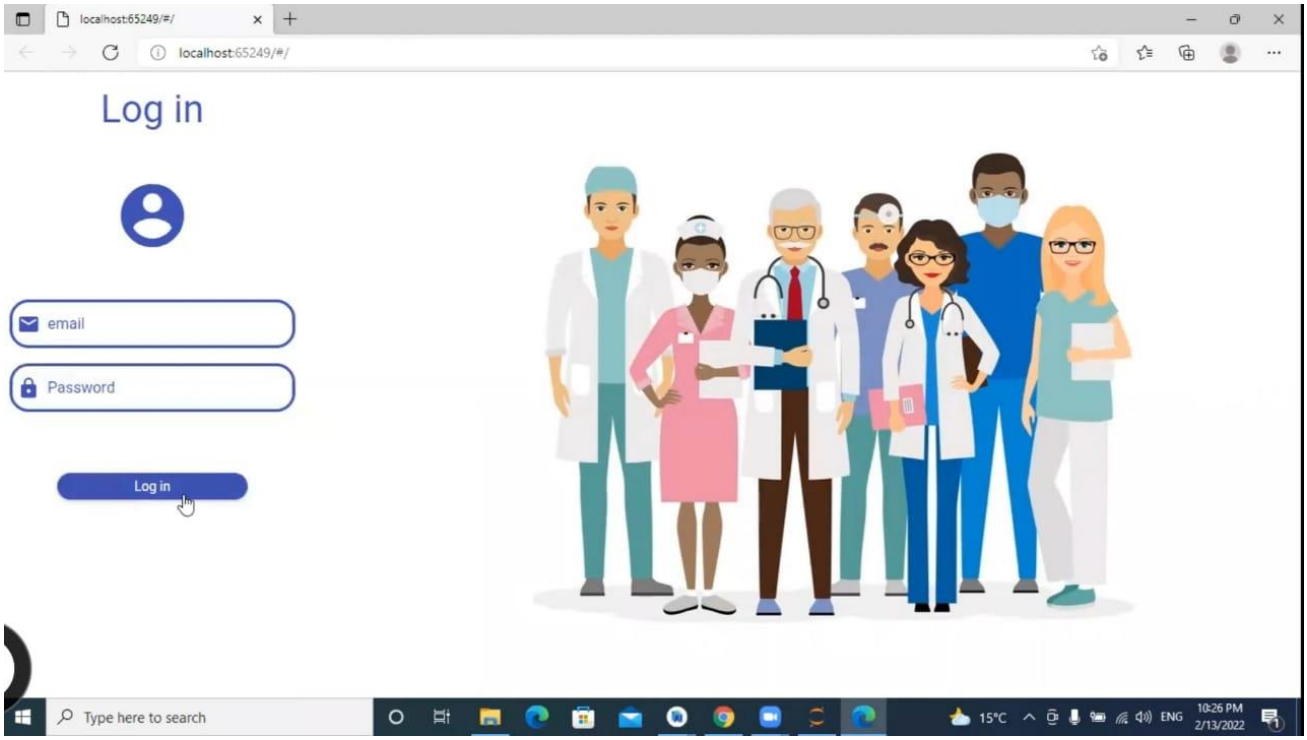


Figure 55 Log in page

-The log in page comes after the sign up page, the doctor uses it if he is already having an email as shown in the figure

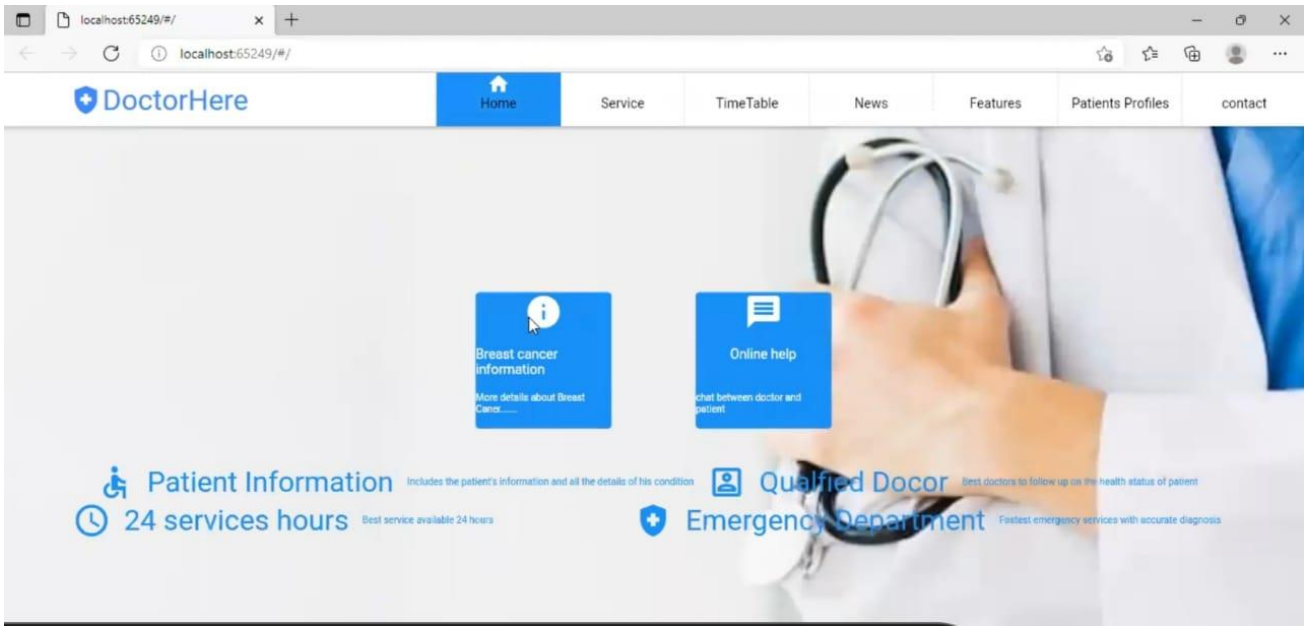


Figure 56 Home page

-Home page contains a several features which can the doctor use as shown at the previous figure.

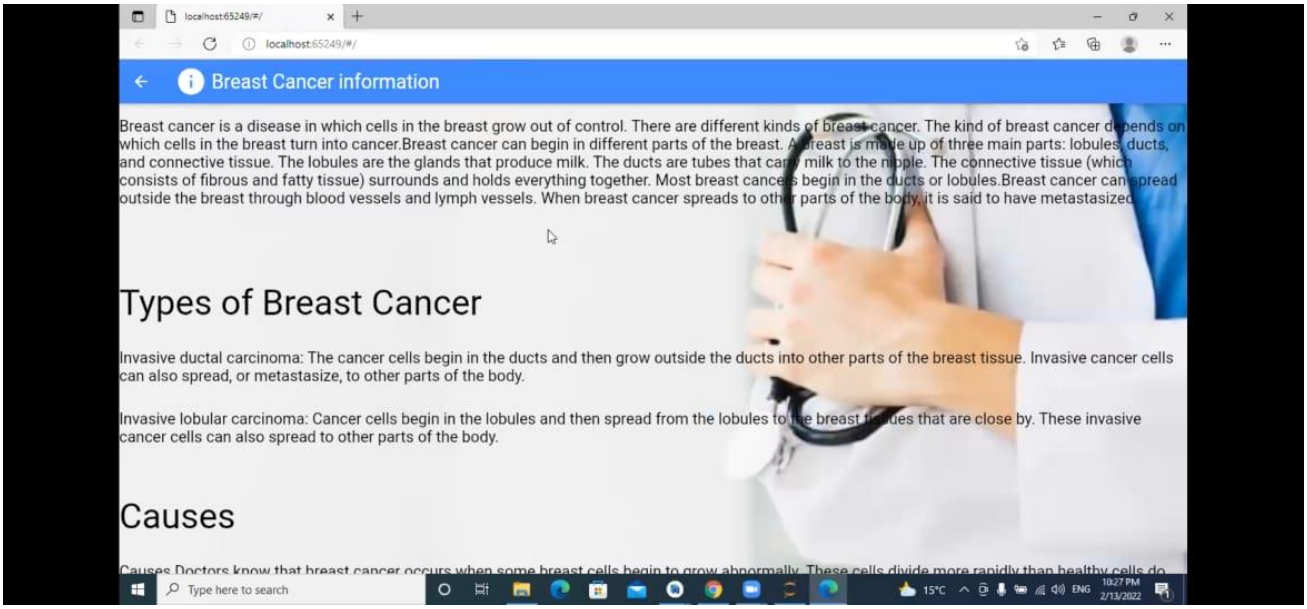


Figure 57 Breast cancer information

-Breast Cancer information screen contains all the information about the breast cancer as shown in the figure.

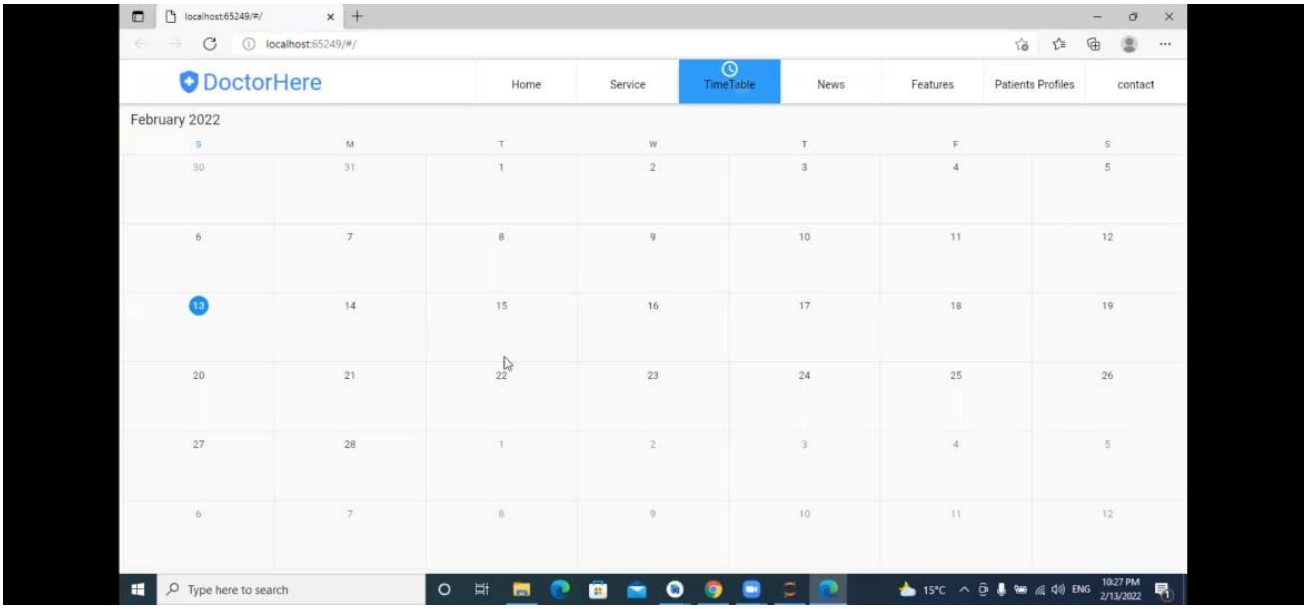


Figure 58 Timetable

- Timetable screen displays the appointments of the doctor.

CHAPTER 6 Experimental design and result

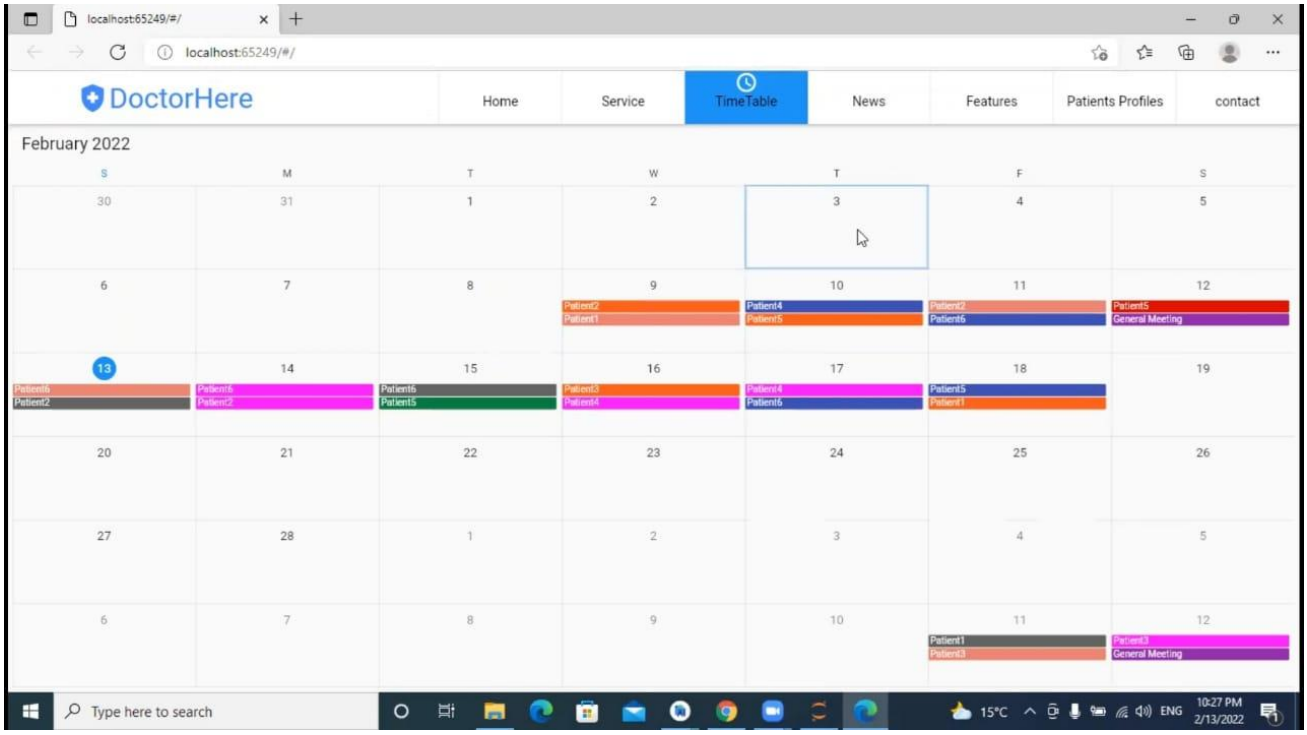


Figure 59 Timetable

- Timetable screen displays all the doctor appointments as shown in Figure 58,59.

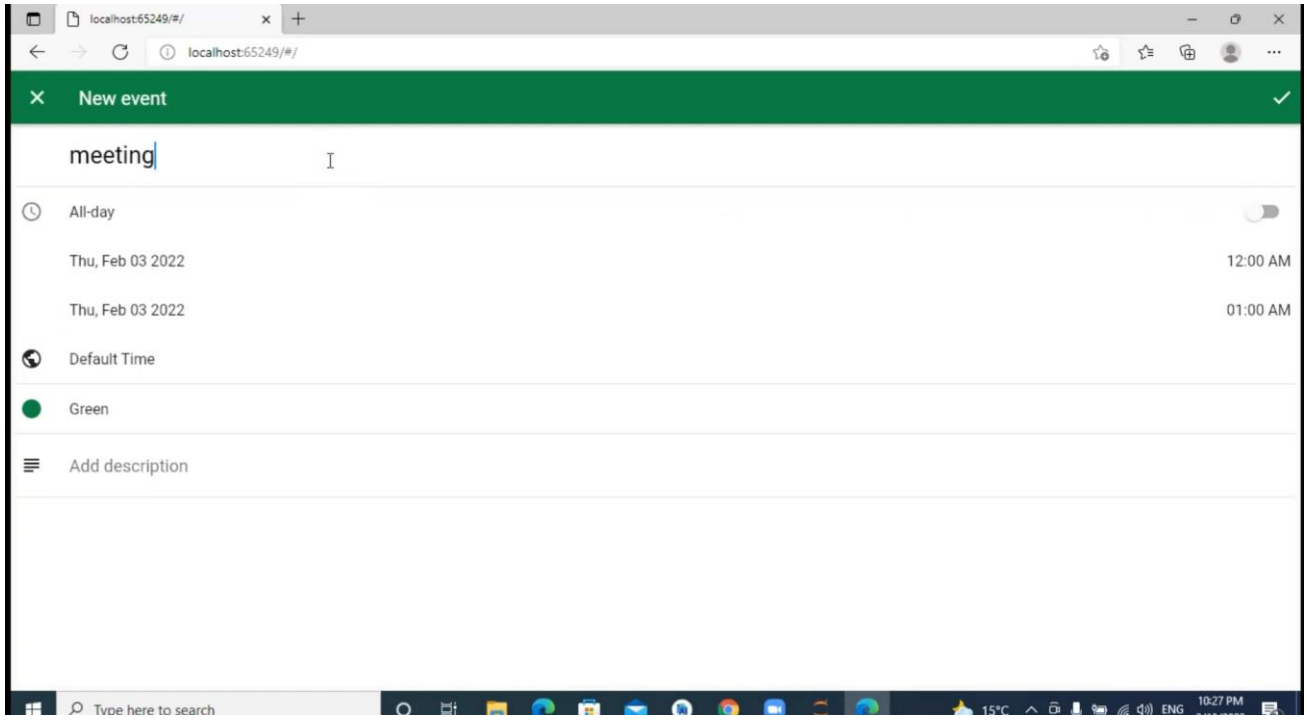


Figure 60 Timetable details

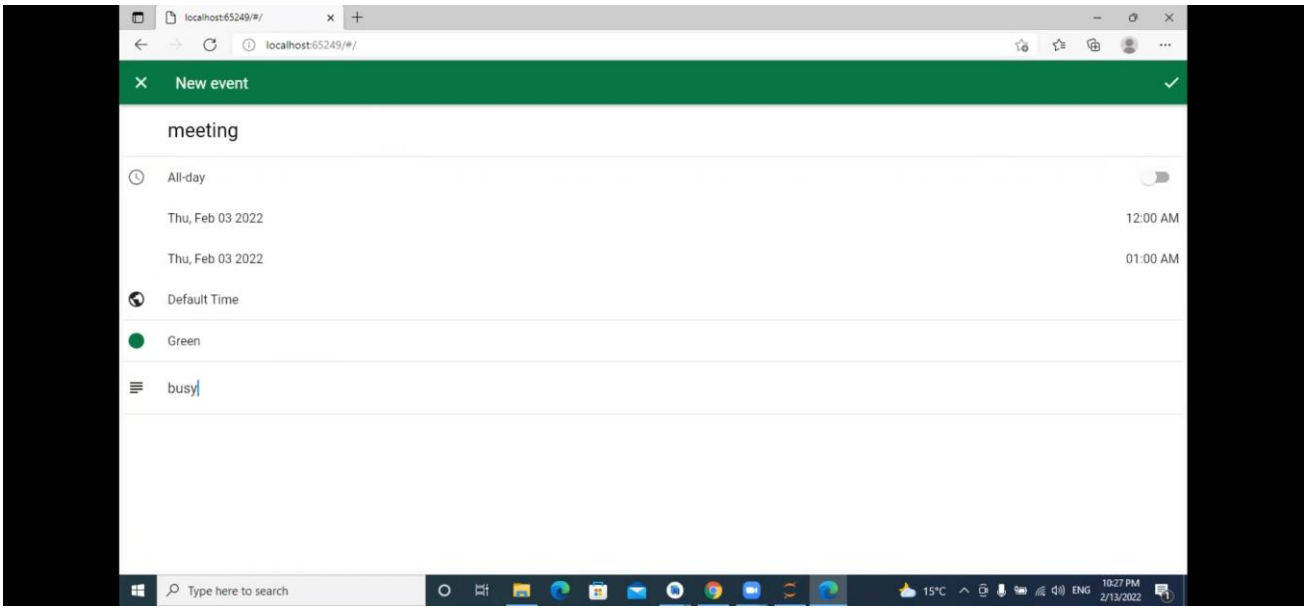


Figure 61 Timetable details

-Timetable details screen allows doctors to put all appointments which they have as shown in figures 60 and 61.

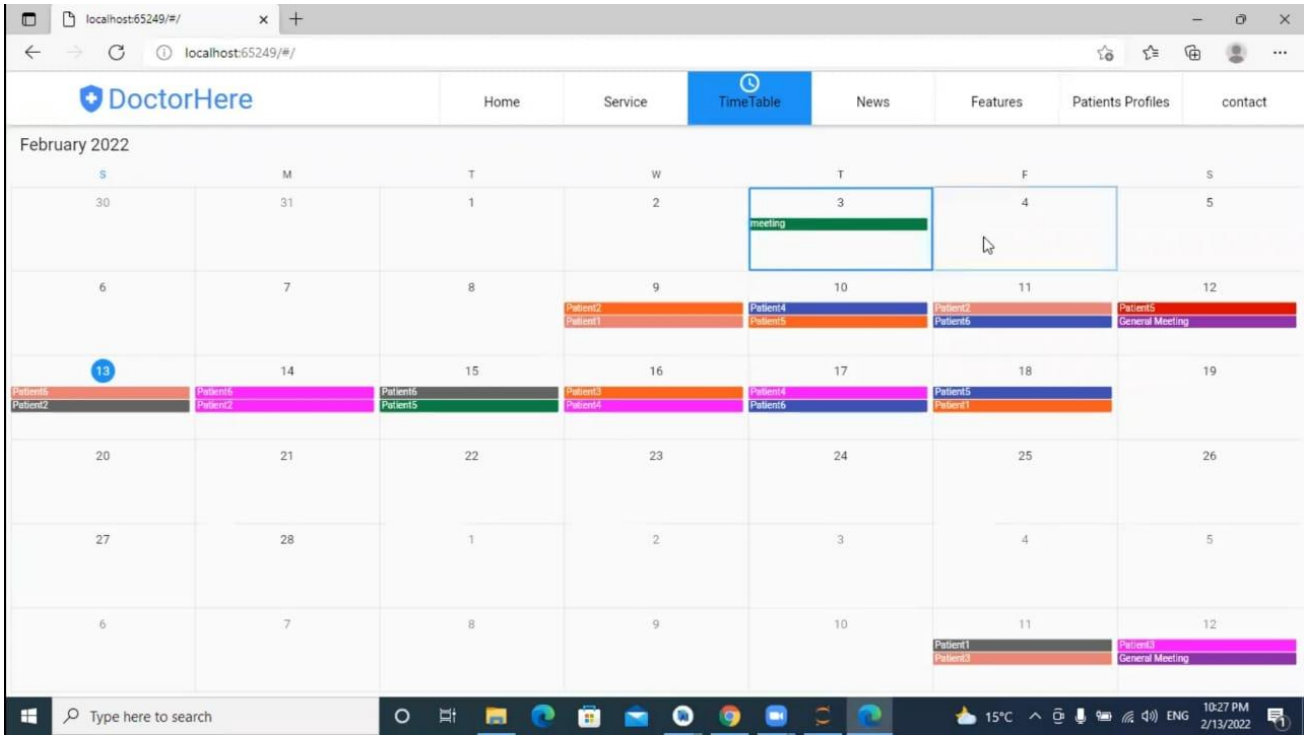


Figure 62 Timetable

-After the doctor put his appointments, it is displayed at his time table as shown in figure 62.

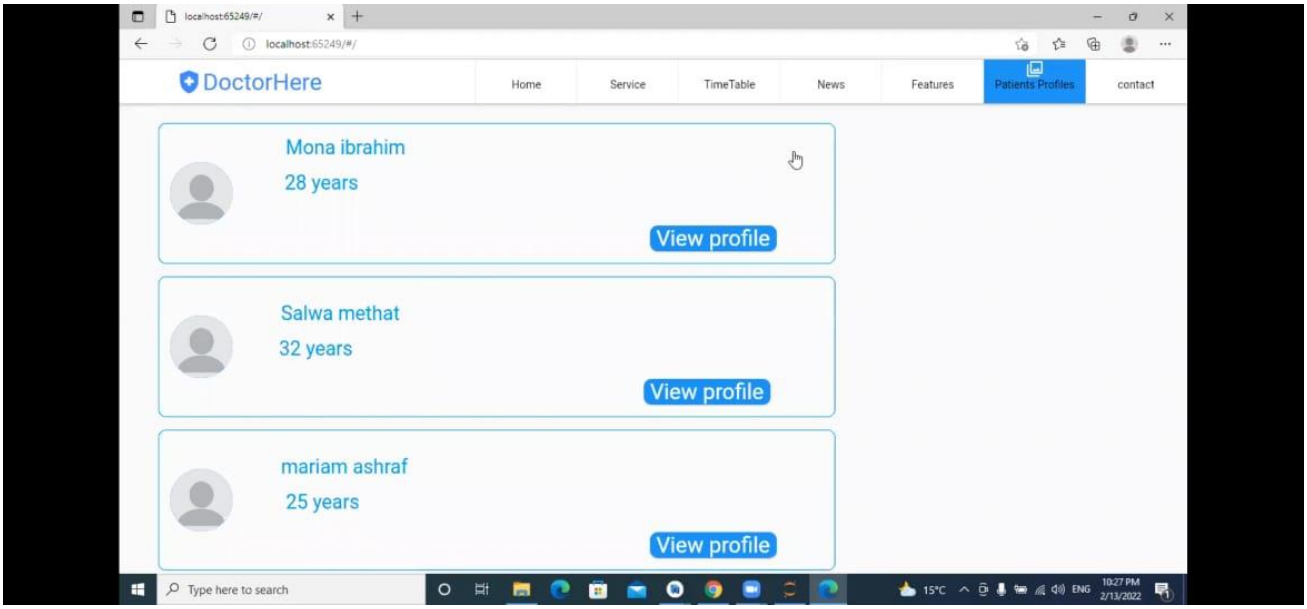


Figure 63 Patient profiles

-Patient profiles screen contains all the profiles of the patients to be examined by the doctor as shown in figure 63.

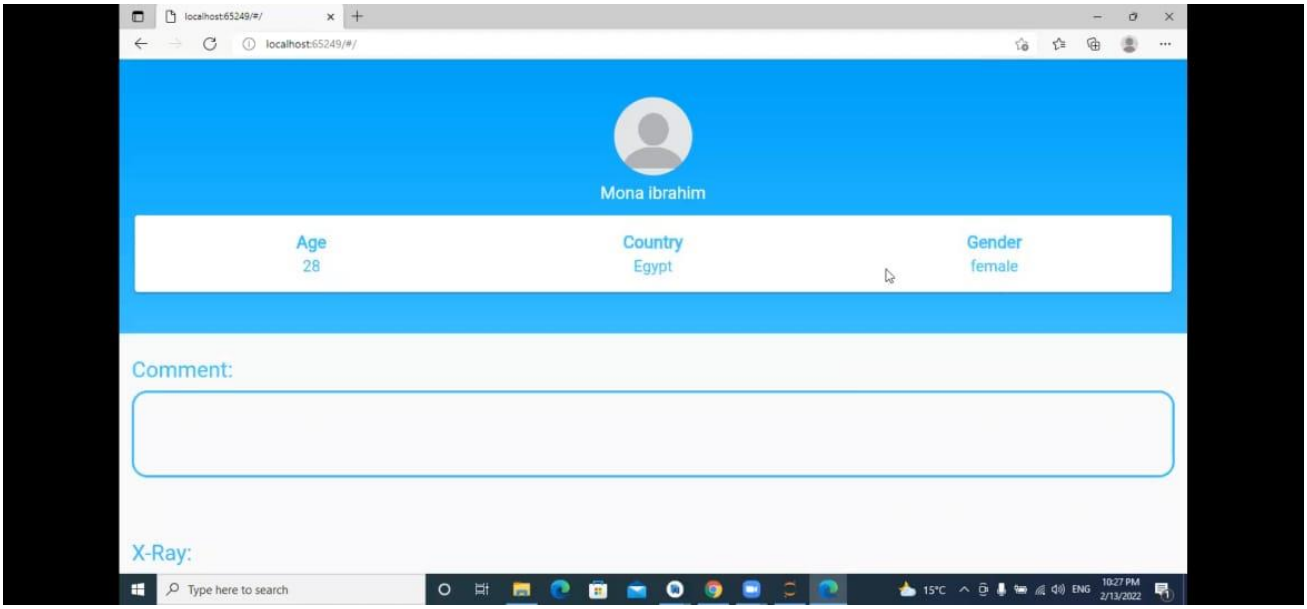


Figure 64 Patient profile details

CHAPTER 6 Experimental design and result

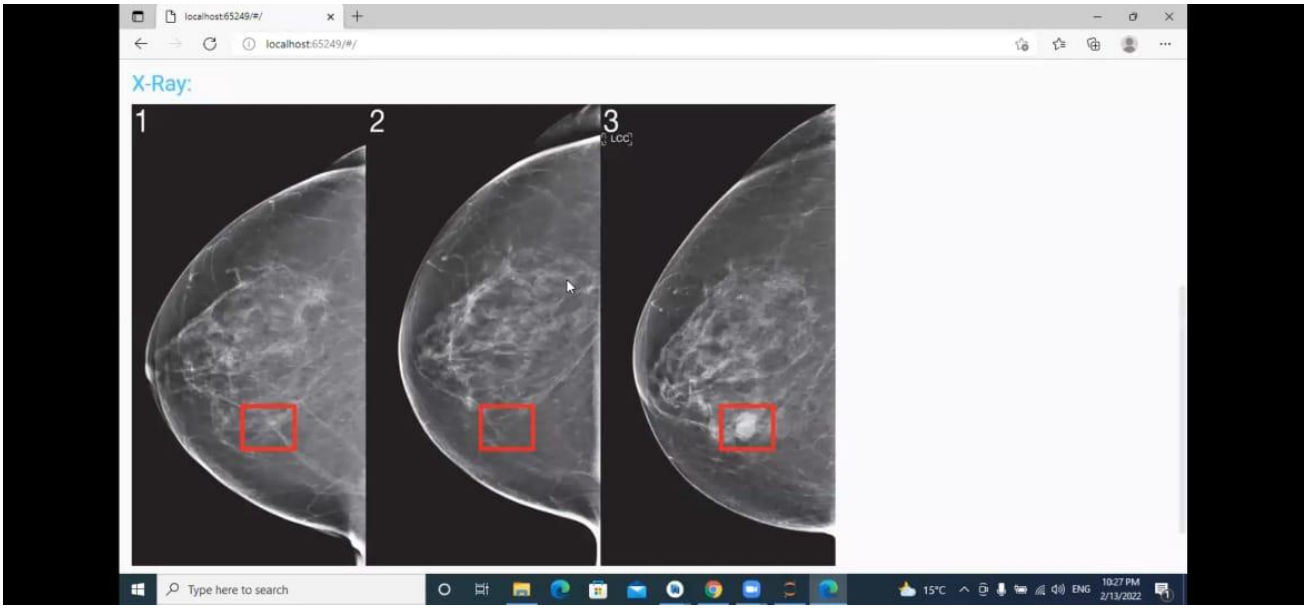


Figure 65 Patient profile details

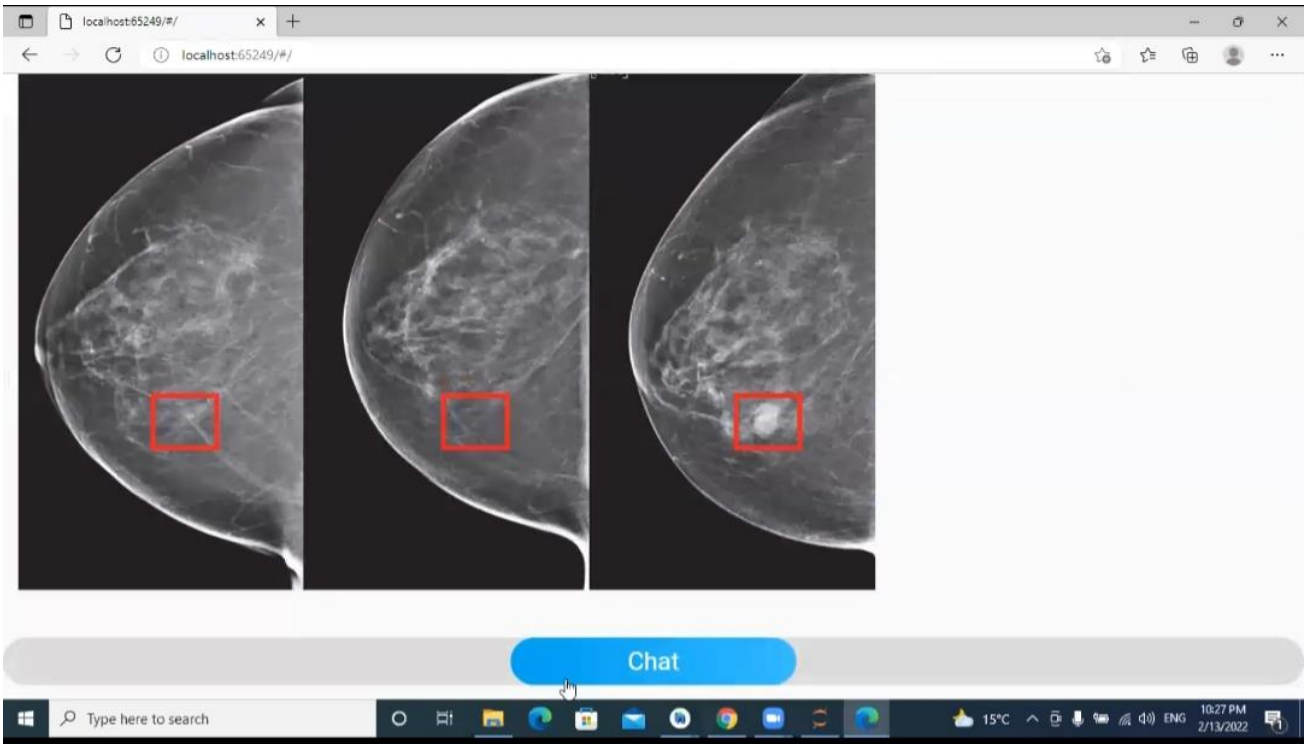


Figure 66 Patient profile details

-Patient profile details screen displays the profile of the patient which contains his X-Ray and the dr. can comment or start chatting with the patient in figure 64, 65 and 66.

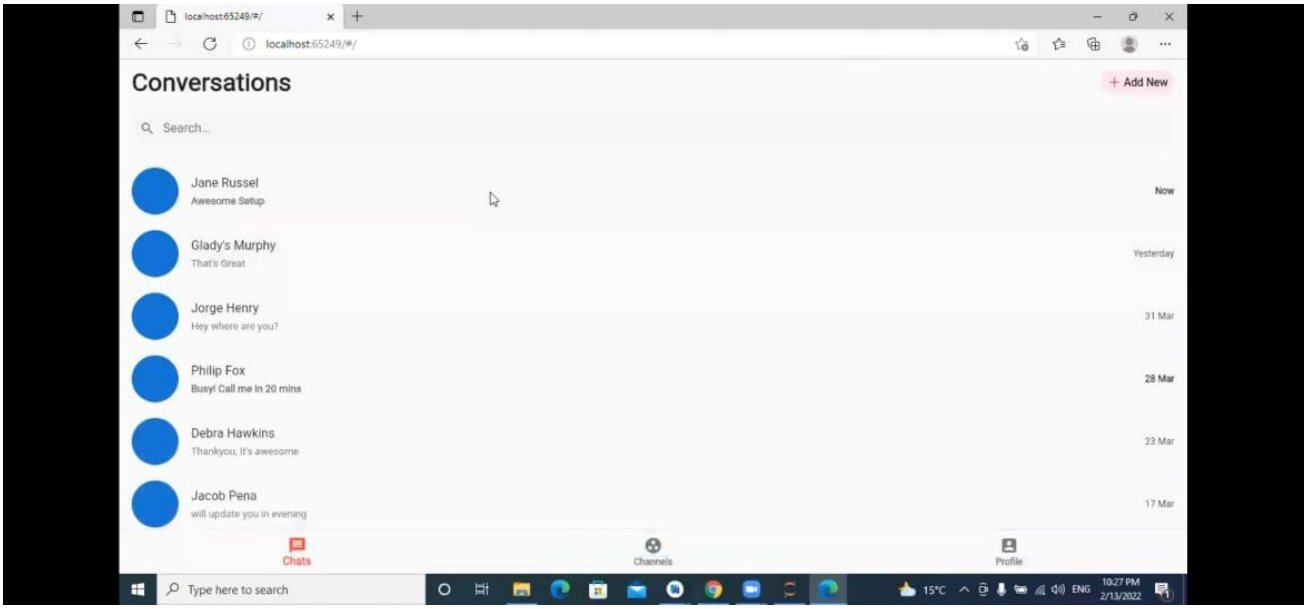


Figure 67 Chat

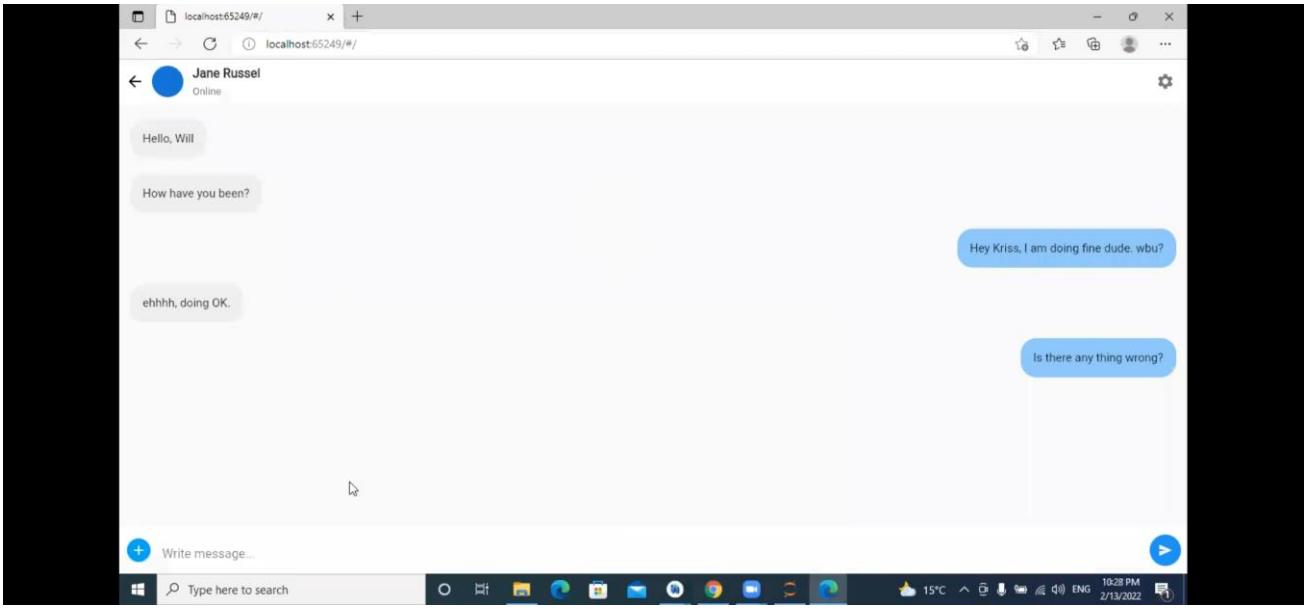


Figure 68 Chat

-Those screenshots display the chat screens as shown in figure 67 and 68.

Welcome To Yo_

Figure 69 welcome page

The Welcome Page of the patient mobile application.

Doctor Here

Welcome to Docotor Here, Let's Go!

Continue

Figure 70 welcome page 2

The second welcome page in the patient mobile application.

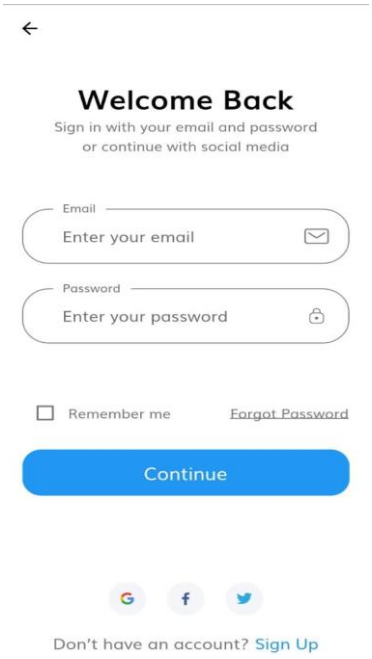


Figure 71 Sign in screen

The sign in screen includes patient name and password.

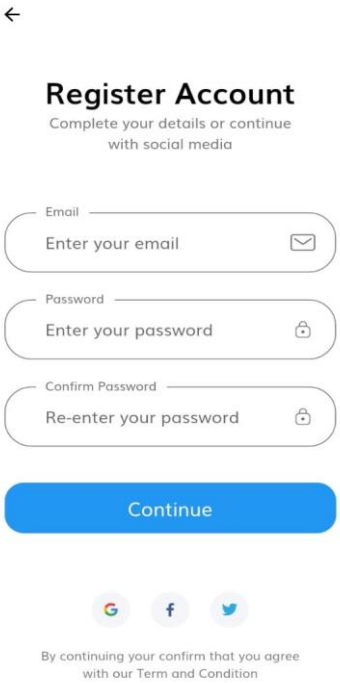


Figure 72 Sign up screen

The registration screen includes patient mail, password and confirm password.

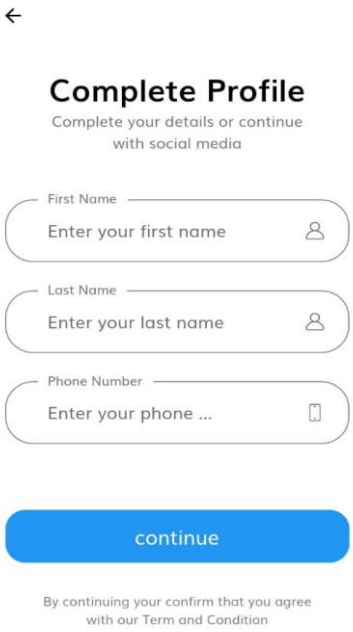


Figure 73 Sign up screen 2

The registration screen includes first name, last name and phone.

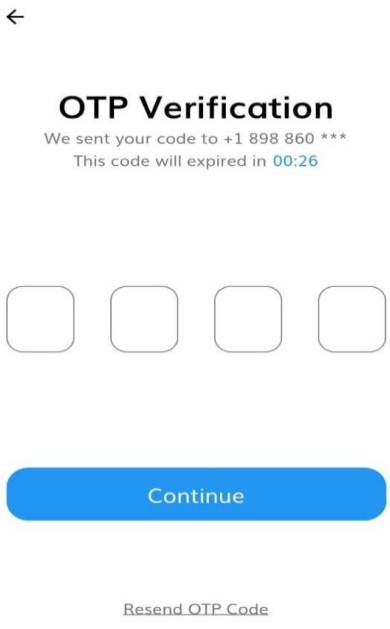


Figure 74 OTP verification

The patient verified his phone number.

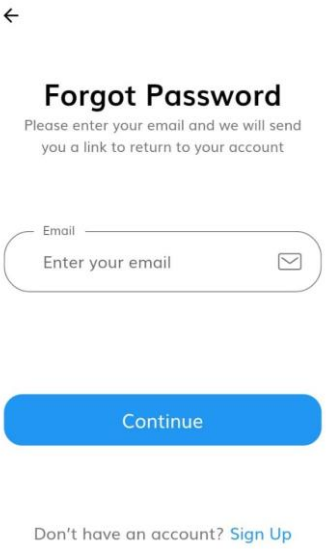


Figure 75 Forget password screen

The forget password screen which the patient can get the old password.

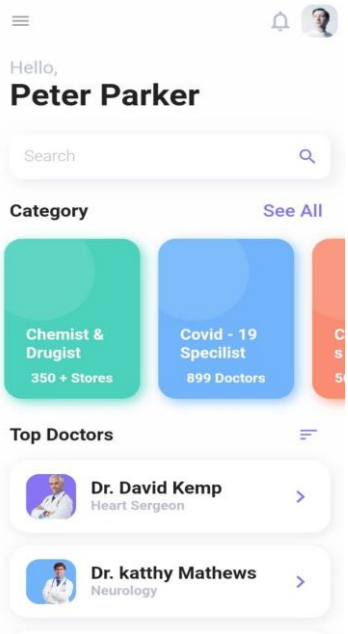


Figure 76 Home screen

The home page includes doctors and categories.

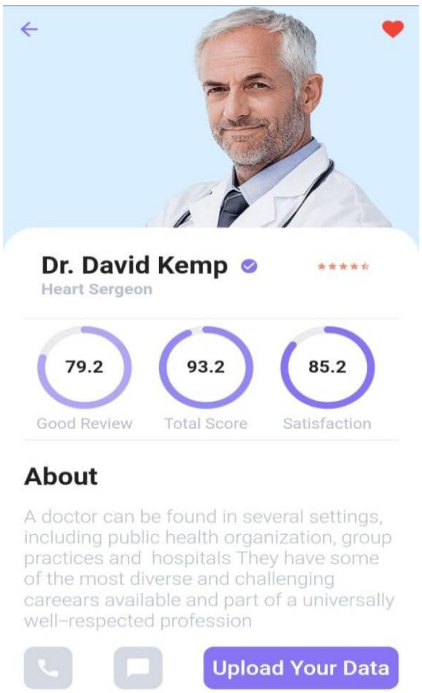


Figure 77 Doctor's profile

The doctor profile screen which displays the doctor information.

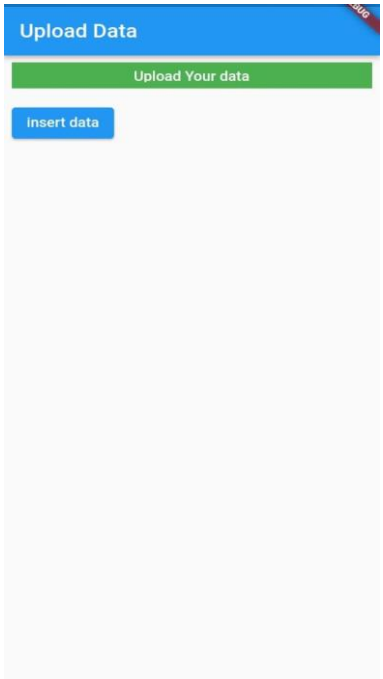


Figure 78 Uploading data
The patient upload csv file.

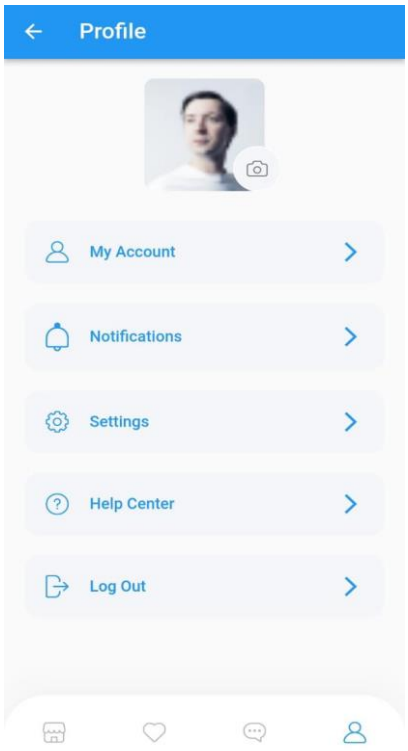


Figure 79 Patient's profile

The patient profile which he can edit his information.

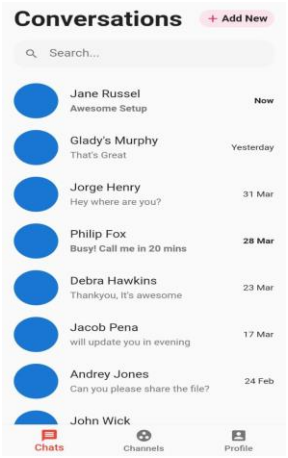


Figure 80 Chat

The chat which the patient can communicate with the doctor.

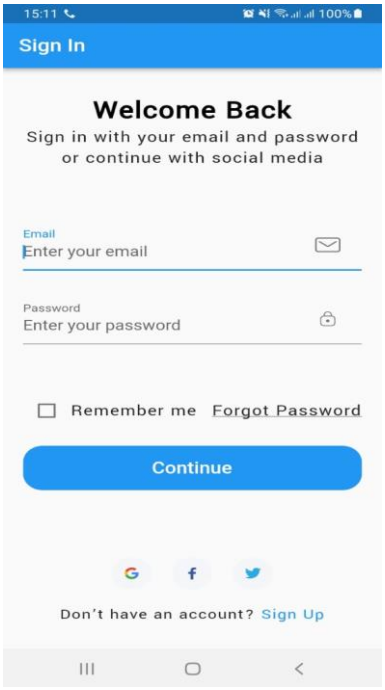


Figure 81 sign in

The sign in screen includes doctor mail and password

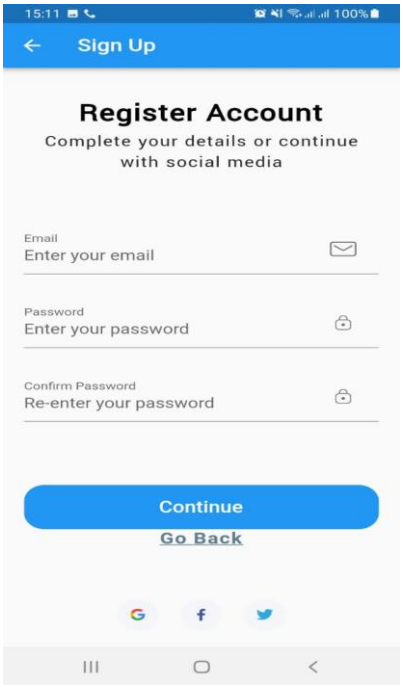


Figure 82 sign up

The registration screen includes doctor's mail , password and confirm password

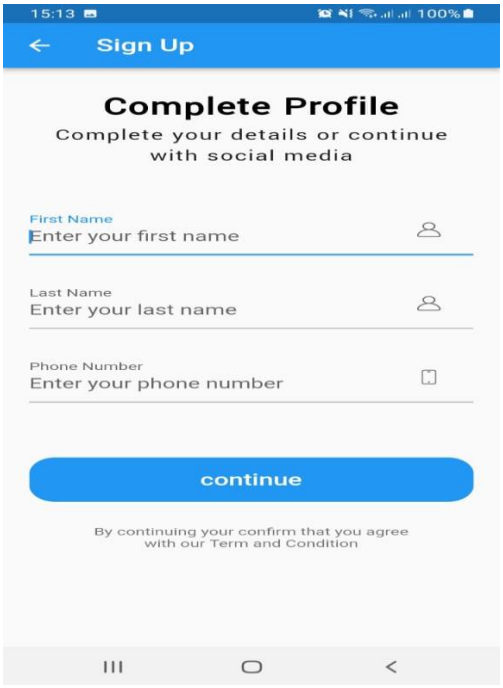


Figure 83 sign up 2

The registration screen includes doctor's name and phone number

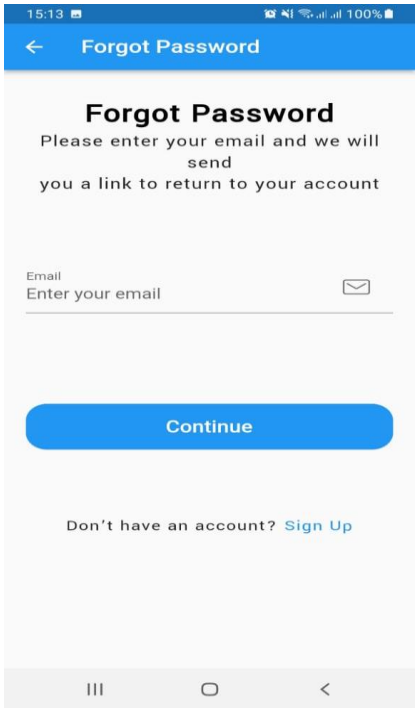


Figure 84 Forgot password

The forget password screen includes the doctor's mail.

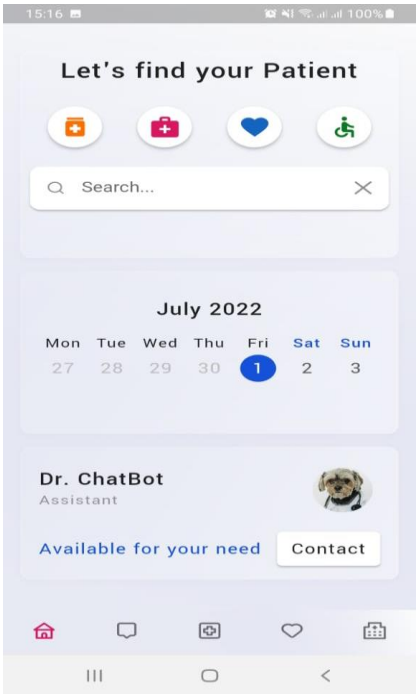


Figure 85 Home screen
The home screen includes the calendar and the search

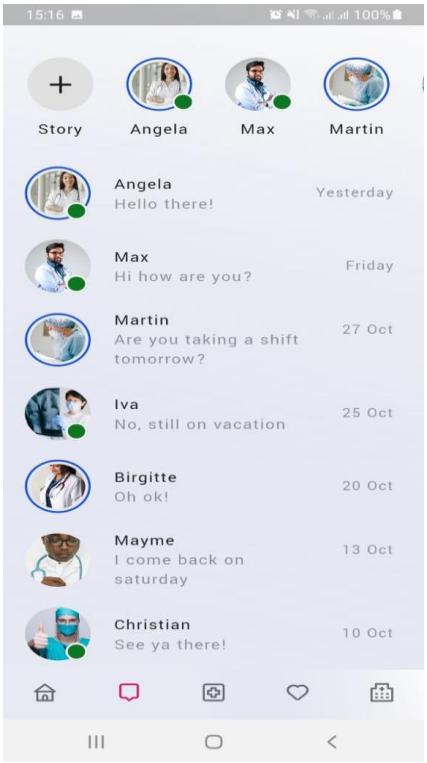


Figure 86 Chat screen
The chat screen includes the chat between doctor and patient.

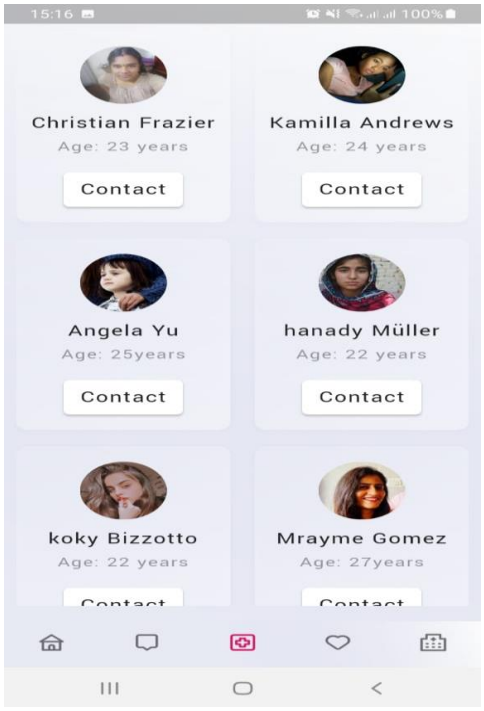


Figure 87 patient screen

The patient screen contains the patients profiles

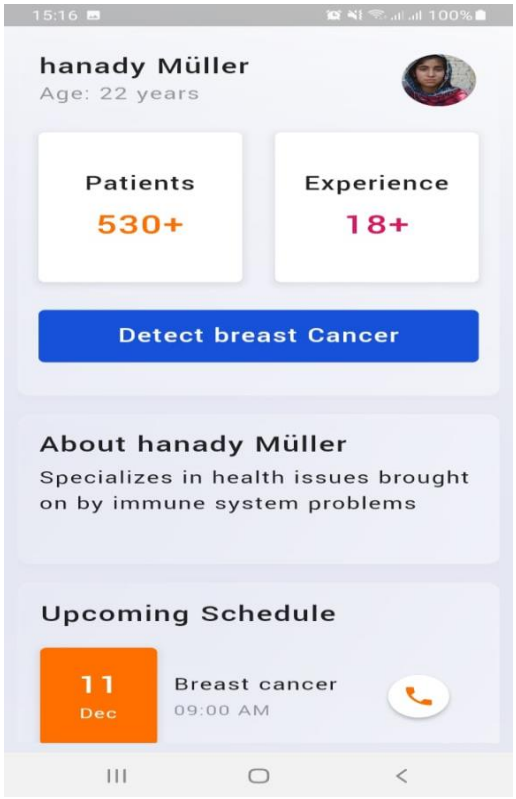


Figure 88 patient profile screen

The patient profile screen includes the patient information.

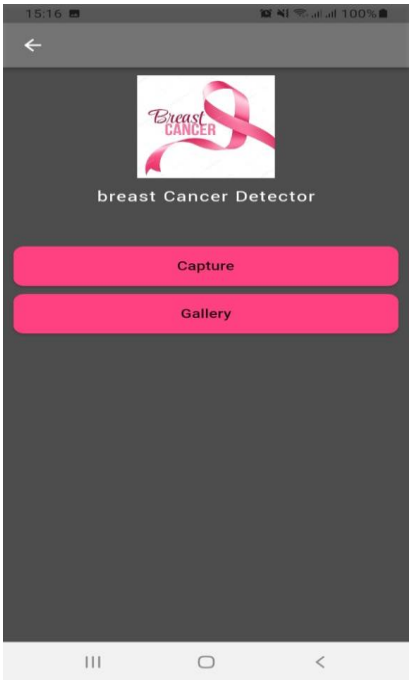


Figure 89 Check screen

Checking screen includes the images of the breast cancer

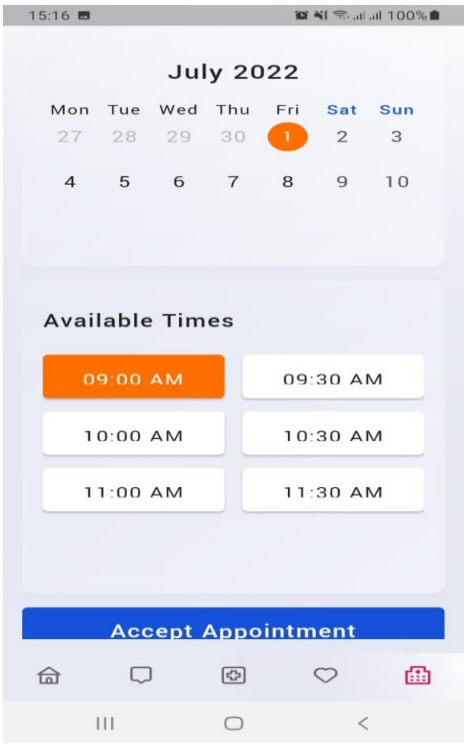


Figure 90 Appointment screen

Appointment screen includes the appointments of the doctor

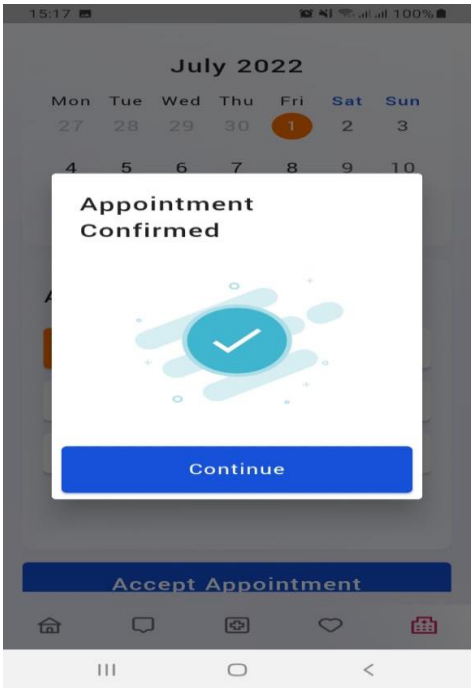


Figure 91 conforming screen

Confirming appointments screen includes the confirmation message of the appointment

6.4Model evaluation

We use eight machine learning text algorithms (KNN, Random Forest, Logistic Regression, Decision Tree, MLP, Back-Propagation) and image algorithms (CNN) and Dense net in the proposed model, dataset includes 32 columns and 469 columns.

KNN made 98.24% accuracy, KNN made 0.9649% accuracy, Logistic Regression made 0.9912% accuracy, Decision Tree made 0.9035% accuracy, MLP made 0.9824% accuracy, Back-Propagation made 0.9824% accuracy, CNN before augmentation made 83% accuracy, CNN after augmentation made 75% accuracy and Dense Net made 0.9471% accuracy.

-Back propagation and Logistic Regression made highest accuracy 0.9912 and Decision Tree made lowest accuracy 0.9035.

Algorithms	KNN	Random Forest	Logistic Regression	Decision Tree	MLP	Back-Propagation	CNN before augmentation	CNN Traditional augmentation	Dense net
Proposed Model	98.24	0.9649	0.9912	0.9035	0.9824	0.9912	83	75	0.9471
Related work	96.4706	93.8596	95.2940	70	99.12	__	58	62	__

Table 1: Comparative Analysis

Algorithms	CNN Before augmentation	CNN After Augmentation	DenseNet
Accuracy	83%	75%	94%

Table 2: Proposed model for image dataset accuracy

CHAPTER 6 Experimental design and result

Machine Learning:

Artificial Neural Network

Multilayer Perceptron (MLP): A multi-layer artificial neuron network is an integral part of deep learning. It is a combination of multiple perceptron models MLP Classifier implements a multi-layer perceptron (MLP) algorithm that trains using Back-propagation.

Back-propagation: is an algorithm for supervised learning of artificial neural networks that uses the gradient descent method to minimize the cost function. It searches for optimal weights that optimize the mean-squared distance between the predicted and actual labels.

Convolutional Neural Network (CNN) is a Deep Learning algorithm specially designed for working with Images. It takes images as inputs, extracts and learns the features of the image, and classifies them based on the learned features.

DenseNet is a network that portrays the importance of having dense connections in a network using dense blocks. This helps in feature-reuse, better gradient flow, reduced parameter count and better transmission of features across the network.

Machine learning algorithms:

In logistic regression, we pass the weighted sum of inputs through an activation function that can map values in between 0 and 1. Such activation function is known as sigmoid function.

The equation of logistic regression:

$$\text{Logit}(p) = \log(p/1-p)$$

Decision Tree: C 5 algorithm is a tree-structured plan of a set of attributes to test in order to predict the output. To decide which attribute should be tested first, simply find the one with the highest information gain.

Information gain is so important in Decision Tree.

Information Gain= 1-Entropy

KNN algorithm: is used for Select the number K of the neighbours and calculate the distance of K number of neighbours.

Random Forest: A forest is comprised of trees. Random forests create decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

6.5Result:

Algorithms	KNN	Random Forest	Logistic Regression	Decision Tree	MLP	Back-Propagation	CNN before augmentation	CNN Traditional augmentation	Dense net
Proposed Model	98.24	0.9649	0.9912	0.9035	0.9824	0.9912	83	75	0.9471
Related work	96.4706	93.8596	95.2940	70	99.12	__	58	62	__

Table 3 Comparative analysis

Algorithms	CNN Before augmentation	CNN After Augmentation	DenseNet
Accuracy	83%	75%	94%

Table 4 Proposed model for image dataset accuracy

The idea of ensemble classification is to learn not just one classifier but a set of classifiers, and them to combine their predictions for the classification of unseen instances using some form of voting.

There are several methods to combine the outputs of the various classifiers.

Voting or averaging of predictions of multiple trained models.

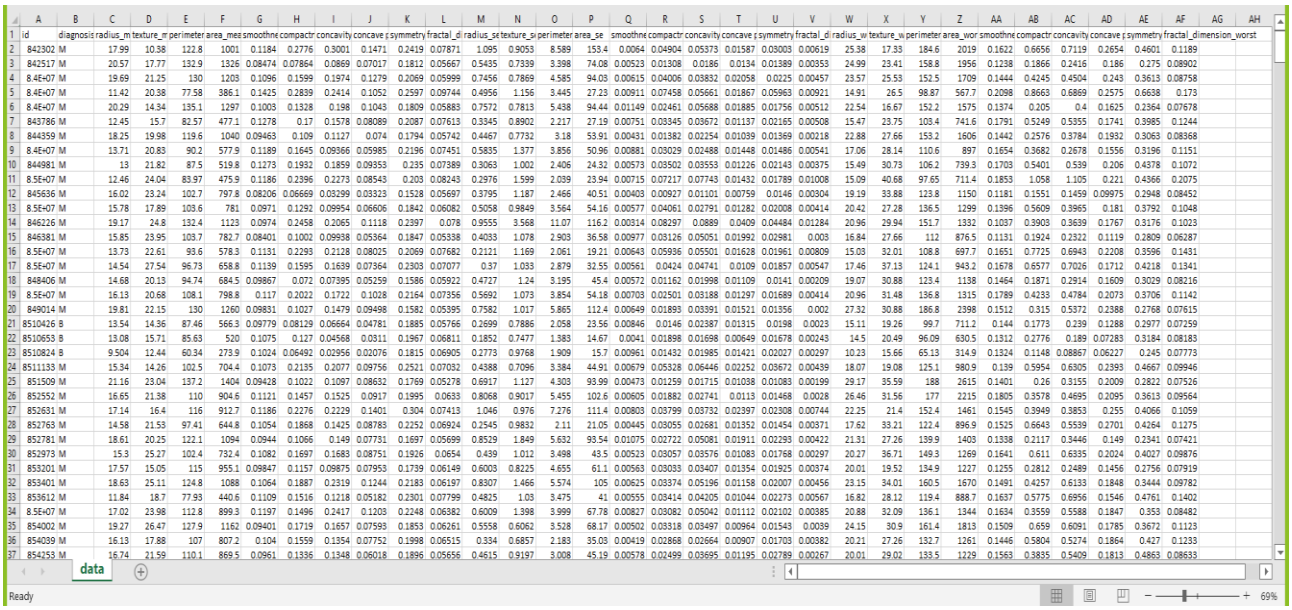
It is used to improve model performance, ideally achieving better performance than any single model used in the ensemble.

Using voting back-propagation is the best result without overfitting with 0.9912 accuracy.

CHAPTER 6 Experimental design and result

6.6 Dataset

Number of attributes: 32
Number of Rows: 56
We take the dataset from Kaggle



id	diagnosis	radius_m	texture_m	perimeter	area_mea	smoothne	compactr	concavity	concave p	symmetry	fractal_di	radius_se	texture_s	perimeter	area_se	smoothne	compactr	concavity	concave p	symmetry	fractal_di	radius_w	texture_w	perimeter	area_wor	smoothne	compactr	concavity	concave p	symmetry	fractal_dimension_worst
842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871	1.095	0.9053	8.589	153.4	0.0064	0.04904	0.05373	0.01587	0.03003	0.00619	25.38	17.33	184.6	2019	0.1622	0.6656	0.7119	0.2654	0.4601	0.1189
842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.00523	0.01308	0.0186	0.0134	0.01389	0.00353	24.99	23.41	158.8	1956	0.1238	0.1866	0.2416	0.186	0.275	0.08902
8.4E+07	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.7869	4.585	94.03	0.00615	0.04006	0.03832	0.02058	0.0225	0.00457	23.57	25.53	152.5	1709	0.1444	0.4245	0.4504	0.243	0.3613	0.08758
8.4E+07	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.156	3.445	27.23	0.00911	0.07458	0.05661	0.01867	0.05963	0.00921	14.91	26.5	98.87	567.7	0.2098	0.8663	0.6869	0.2575	0.6638	0.173
8.4E+07	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.7813	5.438	94.44	0.01149	0.02461	0.05688	0.01885	0.01756	0.00512	22.54	16.67	152.2	1575	0.1374	0.205	0.4	0.1625	0.2364	0.07678
843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.8902	2.217	27.19	0.00751	0.03345	0.03672	0.01137	0.02165	0.00508	15.47	23.75	103.4	741.6	0.1791	0.5249	0.5355	0.1741	0.3985	0.1244
844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.7732	3.18	53.91	0.00431	0.01382	0.02254	0.01039	0.01369	0.00218	22.88	27.66	153.2	1606	0.1442	0.2576	0.3784	0.1932	0.3063	0.08368
8.4E+07	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451	0.5835	1.377	3.856	50.96	0.00881	0.03029	0.02488	0.01448	0.01486	0.00541	17.06	28.14	110.6	897	0.1654	0.3682	0.2678	0.1556	0.3196	0.1151
844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.002	2.406	24.32	0.00573	0.03502	0.03553	0.01226	0.02143	0.00375	15.49	30.73	106.2	739.3	0.1703	0.5401	0.539	0.206	0.4378	0.1072
8.5E+07	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243	0.2976	1.599	2.039	23.94	0.00715	0.07217	0.07743	0.01432	0.01789	0.01008	15.09	40.68	97.65	711.4	0.1853	1.058	1.105	0.221	0.4366	0.2075
845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697	0.3795	1.187	2.466	40.51	0.00403	0.00927	0.01101	0.00759	0.0146	0.00304	19.19	33.88	123.8	1150	0.1181	0.1551	0.1459	0.09975	0.2948	0.08452
8.5E+07	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842	0.06082	0.5058	0.9849	3.564	54.16	0.00577	0.04061	0.02791	0.01282	0.02008	0.00414	20.42	27.28	136.5	1299	0.1396	0.5609	0.3965	0.181	0.3792	0.1048
846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078	0.9555	3.568	11.07	116.2	0.00314	0.08297	0.0889	0.0409	0.04484	0.01284	20.96	29.94	151.7	1332	0.1037	0.3903	0.3639	0.1767	0.3176	0.1023
846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338	0.4033	1.078	2.903	36.58	0.00977	0.03126	0.05051	0.01992	0.02981	0.003	16.84	27.66	112	876.5	0.1131	0.1924	0.2322	0.1119	0.2809	0.06287
8.5E+07	M	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069	0.07682	0.2121	1.169	2.061	19.21	0.00643	0.05936	0.05501	0.01628	0.01961	0.00809	15.03	32.01	108.8	697.7	0.1651	0.7725	0.6943	0.2208	0.3596	0.1431
8.5E+07	M	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077	0.37	1.033	2.879	32.55	0.00561	0.0424	0.04741	0.0109	0.01857	0.00547	17.46	37.13	124.1	943.2	0.1678	0.6577	0.7026	0.1712	0.4218	0.1341
848406	M	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586	0.05922	0.4727	1.24	3.195	45.4	0.00572	0.01162	0.01998	0.01109	0.0141	0.00209	19.07	30.88	123.4	1138	0.1464	0.1871	0.2914	0.1609	0.3029	0.08216
8.5E+07	M	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	0.5692	1.073	3.854	54.18	0.00703	0.02501	0.03188	0.01297	0.01689	0.00414	20.96	31.48	136.8	1315	0.1789	0.4233	0.4784	0.2073	0.3706	0.1142
849014	M	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498	0.1582	0.05395	0.7582	1.017	5.865	112.4	0.00649	0.01893	0.03391	0.01521	0.01356	0.002	27.32	30.88	186.8	2398	0.1512	0.315	0.5372	0.2388	0.2768	0.07615
8510426	B	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05766	0.2699	0.7886	2.058	23.56	0.00846	0.0146	0.02387	0.01315	0.0198	0.0023	15.11	19.26	99.7	711.2	0.144	0.1773	0.239	0.1288	0.2977	0.07259
8510653	B	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967	0.06811	0.1852	0.7477	1.383	14.67	0.0041	0.01898	0.01698	0.00649	0.01678	0.00243	14.5	20.49	96.09	630.5	0.1312	0.2776	0.189	0.07283	0.3184	0.08183
8510824	B	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076	0.1815	0.06905	0.2773	0.9768	1.909	15.7	0.00961	0.01432	0.01985	0.01421	0.02027	0.00297	10.23	15.66	65.13	314.9	0.1324	0.1148	0.08867	0.06227	0.245	0.07773
8511133	M	15.34	14.26	102.5	704.4	0.1073	0.2135	0.2077	0.09756	0.2521	0.07032	0.4388	0.7096	3.384	44.91	0.00679	0.05328	0.06446	0.02252	0.03672	0.00439	18.07	19.08	125.1	980.9	0.139	0.5954	0.6305	0.2393	0.4667	0.09946
851509	M	21.16	23.04	137.2	1404	0.09428	0.1022	0.1097	0.08632	0.1769	0.05278	0.6917	1.127	4.303	93.99	0.00473	0.01259	0.01715	0.01038	0.01083	0.00199	29.17	35.59	188	2615	0.1401	0.26	0.3155	0.2009	0.2822	0.07526
852552	M	16.65	21.38	110	904.6	0.1121	0.1457	0.1525	0.0917	0.1995	0.0633	0.8068	0.9017	5.455	102.6	0.00605	0.01882	0.02741	0.0113	0.01468	0.0028	26.46	31.56	177	2215	0.1805	0.3578	0.4695	0.2095	0.3613	0.09564
852631	M	17.14	16.4	116	912.7	0.1186	0.2276	0.2229	0.1401	0.304	0.07413	1.046	0.976	7.276	111.4	0.00803	0.03799	0.03732	0.02397	0.02308	0.00744	22.25	21.4	152.4	1461	0.1545	0.3949	0.3853	0.255	0.4066	0.1059
852763	M	14.58	21.53	97.41	644.8	0.1054	0.1868	0.1425	0.08783	0.2252	0.06924	0.2545	0.9832	2.11	21.05	0.00445	0.03055	0.02681	0.01352	0.01454	0.00371	17.62	33.21	122.4	896.9	0.1525	0.6643	0.5539	0.2701	0.4264	0.1275
852781	M	18.61	20.25	122.1	1094	0.0944	0.1066	0.149	0.07731	0.1697	0.05699	0.8529	1.849	5.632	93.54	0.01075	0.02722	0.05081	0.01911	0.02293	0.00422	21.31	27.26	139.9	1403	0.1338	0.2117	0.3446	0.149	0.2341	0.07421
852973	M	15.3	25.27	102.4	732.4	0.1082	0.1697	0.1683	0.08751	0.1926	0.0654	0.439	1.012	3.498	43.5	0.00523	0.03057	0.03576	0.01083	0.01768	0.00297	20.27	36.71	149.3	1269	0.1641	0.611	0.6335	0.2024	0.4027	0.09876
853201	M	17.57	15.05	115	955.1	0.09847	0.1157	0.09875	0.07953	0.1739	0.06149	0.6003	0.8225	4.655	61.1	0.00563	0.03033	0.03407	0.01354	0.01925	0.00374	20.01	19.52	134.9	1227	0.1255	0.2812	0.2489	0.1456	0.2756	0.07919
853401	M	18.63	25.11	124.8	1088	0.1064	0.1887	0.2319	0.1244	0.2183	0.06197	0.8307	1.466	5.574	105	0.00625	0.03374	0.05196	0.01158	0.02007	0.00456	23.15	34.01	160.5	1670	0.1491	0.4257	0.6133	0.1848	0.3444	0.09782
853612	M	11.84	18.7	77.93	440.6	0.1109	0.1516	0.1218	0.05182	0.2301	0.07799	0.4825	1.03	3.475	41	0.00555	0.03414	0.04205	0.01044	0.02273	0.00567	16.82	28.12	119.4	888.7	0.1637	0.5775	0.6956	0.1546	0.4761	0.1402
8.5E+07	M	17.02	23.98	112.8	899.3	0.1197	0.1496	0.2417	0.1203	0.2248	0.06382	0.6009	1.398	3.999	67.78	0.00827	0.03082	0.05042	0.01112	0.02102	0.00385	20.88	32.09	136.1	1344	0.1634	0.3559	0.5588	0.1847	0.353	0.08482
854002	M	19.27	26.47	127.9	1162	0.09401	0.1719	0.1657	0.07593	0.1853	0.06261	0.5558	0.6062	3.528	68.17	0.00502	0.03318	0.03497	0.00964	0.01543	0.0039	24.15	30.9	161.4	1813	0.1509	0.659	0.6091	0.1785	0.3672	0.1123
854039	M	16.13	17.88	107	807.2	0.104	0.1559	0.1354	0.07752	0.1998	0.06515	0.334	0.6857	2.183	35.03	0.00419	0.02868	0.02664	0.00907	0.01703	0.00382	20.21	27.26	132.7	1261	0.1446	0.5804	0.5274	0.1864	0.427	0.1233
854253	M	16.74	21.59	110.1	869.5	0.0961	0.1336	0.1348	0.06018	0.1896	0.05656	0.4615	0.9197	3.008	45.19	0.00578	0.02499	0.03695	0.01195	0.02789	0.00267	20.01	29.02	133.5	1229	0.1563	0.3835	0.5409	0.1813	0.4863	0.08633

Figure 92 Text dataset

Flutter:

Mongo DB: MongoDB stores data in flexible, JSON-like documents, meaning fields can vary from document to document and data structure can be changed over time The document model maps to the objects in your application code, making data easy to work with ad hoc queries, indexing, and real time aggregation provide powerful ways to access and analyze your data. MongoDB is a distributed database at its core, so high availability, horizontal scaling, and geographic distribution are built in and easy to use mongo DB is free to use. Versions released prior to October 16, 2018 are published under the AGPL. All versions released after October 16, 2018, including patch fixes for prior versions, are published under the Server-Side Public License (SSPL) v1.

We have used Mongo Atlas DB because it is a fully managed cloud database developed by the Mongo DB Team. This handles the complexity of deploying, managing, and heal your deployments on the cloud service. Developers, deploy and scale the MongoDB cluster in just a few clicks. This MongoDB Atlas is a global cloud database service built and run by the Mongo DB team.

CHAPTER 6 Experimental design and result

We have 30 columns in our dataset, the data is normalized then we convert the data in the csv file to JSON file.

UIs built with Flutter always depend on the back end technology stack for core functionalities like authentication and access control. One of these back end technologies is the Flask framework. Flask is a Python web framework for building web applications.

We have to insert the predicated data in Mongo DB and view it from Mongo DB so we use flask. After that we take the local host IP address and run it on browser, the predicated data is displayed on it.

we have used Firebase, it is a Back end-as-a-Service (Baa S) app development platform that provides hosted back end services such as a real time database, cloud storage, authentication, crash reporting, machine learning, remote configuration, and hosting for your static files.

we have used the fire-base for authenticating the registered data. First, we sign up by username, phone number, mail and password but we authenticate by mail and password. Then the sign in screen we fill mail and password only and authenticate by them. Finally, the Forget password screen, we fill the data and received a support mail including the new password.

CHAPTER 7
Conclusion and future work

CHAPTER 7 conclusion and future work

Conclusion

Even if technology has developed, still lots of people are facing many issues with modern age diseases. Breast cancer has become one of the most common deadliest diseases rising over days among all countries in world.

Ratio of this disease increases due to lack of awareness and late identification. Our result reveals ANN (machine learning) plays measure factor for detection of cancer diagnosis to save the human life from the dangerous disease.

The method MLP, Back-propagation, Dense Net, K-Nearest Neighbor, Random Forest, Decision Tree and Logistic regression models are employed as the classification to predict the breast cancer tumor type with other attributes (Benign, Malignant).

The results of prediction will help to decrease rate of misdiagnoses. Through the use of voting in ensemble model back-propagation is the best result without overfitting with 0.9912 accuracy.

Using Flutter framework as a frontend, Mongo DB Atlas enables us to insert from patient and view data to doctor with predication through using flask that takes the trained model to predict the result whether it is benign or malignant.

Future Work:

Improving accuracy of a model by adding more datasets, training and running various algorithms on a high specification GPU.

Develop the flutter application to enable the application to alarm patient if his case is getting worse by sending a message to the application on his mobile phone.

Using sensors on the patient and use IOT to connect between the system and the sensors.

References

References

[1] Lingraj Dora, Sanjay Agrawal, Rutuparna Panda b, Ajith Abraham (2017), "Optimal breast cancer classification using Gauss–Newton representation-based algorithm"

[2] Shajib Ghosh, Jubaer Hossain, Dr. Shaikh Anowarul Fattah, Dr. Celia Shahnaz, Asir Intisar Khan(2017)"Efficient approaches for accuracy improvement of breast cancer classification using Wisconsin database"

[3] Abdulkader Helwana, John Bush Idokob, Rahib H. Abiyevb (2017), "Machine learning techniques for classification of breast tissue"

[4] Dada Emmanuel Gbenga, Ngene Christopher, Daramola Comfort Yetunde (2017), "Performance Comparison of Machine Learning Techniques for Breast Cancer Detection"

[5] Sulaiman Vesal, Nishant Ravikumar , AmirAbbas Davari , Stephan Ellmann , Andreas Maier (2018), "Classification of breast cancer histology images using transfer learning"

[6] Omar Ibrahim Obaid , Mazin Abed Mohammed , Salama Mostafa, Mohd Khanapi Abd Ghani (2018), "Evaluating the Performance of Machine Learning Techniques in the Classification of Wisconsin Breast Cancer"

[7] Muhammet Fatih Aslan , Yunus Celik , Kadir Sabanci , Akif Durdu(2018), "Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data"

[8] DALAL BARDOU, KUN ZHANG, AND SAYED MOHAMMAD AHMAD(2018), "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks"

[9]Yixuan Li, Zixuan Chen(2018), "Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction"

[10] Walid CHERIF(2018), "Optimization of K-NN algorithm byclustering and reliability coefficiento breast-cancer diagnosists"

[11] Wenbin Yue,Zidong Wang ,Hongwei Chen ,Annette Payne,Xiaohui Liu(2018), "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis"

[12] Rati Shukla, Vikash Yadav, Parashu Ram Pal, Pankaj Pathak(2019), "Machine Learning Techniques for Detecting and Predicting Breast Cancer"

[13] Ali Al Bataineh(2019), "A Comparative Analysis of Nonlinear Machine Learning Algorithms for Breast Cancer Detection"

[14] Bibhuprasad Sahu, Sachi Nandan Mohanty,Saroj Kumar Rout (2019), "A Hybrid Approach for Breast Cancer Classification and Diagnosis"

[15] David A. Omondiagbe ,Shanmugam Veeramani ,Amandeep S. Sidhu(2019), "Machine Learning Classification Techniques for Breast Cancer Diagnosis"

[16] Ebru Aydındag Bayrak, Pınar Kırcı, Tolga Ensari(2019), "Comparison of Machine Learning Methods for Breast Cancer Diagnosis"

[17] Ratula Ray, Azian Azamimi Abdullah, Debasish Kumar Mallick, Satya Ranjan Dash(2019), "Classification of Benign and Malignant Breast Cancer using Supervised Machine Learning Algorithms Based on Image and Numeric Datasets"

[18] Abderrahmane Eddaoudy, Khalil Maalmi(2020), "Breast cancer classification with reduced feature set using association rules and support vector machine"

[19] Nikita Rane,Rucha Kanade,Jean Sunny,Prof. Sulochana Devi(2020), "IJERT-Breast Cancer Classification and Prediction using Machine Learning"

[20] Shallu Sharma,Rajesh Mehra(2020), "Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images—a Comparative Insight"

References

[21] Sara Laghmati, Bouchaib Cherradi, Amal Tmiri, Othmane Daanouni and Soufiane Hamida(2020), “Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques”

[22] Ghada Hamed(B), Mohammed Abd El-Rahman Marey, Safaa El- Mohamed Fahmy(2020), “Deep Learning in Breast Cancer Detection and Classification”

[23] Puja Guptaa , Shruti Garga(2020), “Breast Cancer Prediction using varying Parameters of Machine Learning Models”

[24] Md. Milon Islam,Md. Rezwanul Haque, Hasib Iqbal, Md. Munirul Hasan, Mahmudul Hasan, Muhammad Nomani(2020), “Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques”

[25] Jiande Wu and Chindo Hicks(2020), “Breast Cancer Type Classification Using Machine Learning”

[26] Yadavendra , Satish Chand(2020), “A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method”

[27] Srwa Hasan Abdulla, Ali Makki Sagheer(2021), “Breast Cancer Classification Using Machine Learning Techniques”

[28] Said Boumaraf , Xiabi Liu , Yuchai Wan 2, Zhongshu Zheng , Chokri Ferkous , Xiaohong Ma ,Zhuo Li ,Dalal (2021), “Conventional Machine Learning versus Deep Learning forMagnification Dependent Histopathological Breast CancerImage Classification: A Comparative Study with Visual Explanation”

[29] Dina A. Ragab , Omneya Attallah , Maha Sharkas , Jinchang Ren , Stephen Marshall(2021), “Framework for Breast Cancer Classification using Multi-DCNNs”

[30] https://www.javatpoint.com/machine-learning-naive-bayes-classifier

[31] https://dataaspirant.com/how-decision-tree-algorithm-works/

[32] https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptron

[33]https://aipoint.tech/blog/post/backpropagation-algorithm-in-machine-learning-with-python-code/

[34] https://www.freecodecamp.org/news/what-is-flutter-and-why-you-should-learn-it-in-2020/

[35] https://www.researchgate.net/figure/Pseudocode-of-decision-tree-learning_fig1_233410133

[36] https://www.smartdraw.com/use-case-diagram/

[37] https://www.lucidchart.com/pages/uml-sequence-diagram#:~:text=A%20sequence%20diagram%20is%20a,to%20document%20an%20existing%20process.

[38] https://www.geeksforgeeks.org/introduction-machine-learning/#:~:text=The%20term%20Machine%20Learning%20was,learn%20without%20being%20explicitly%20programmed%E2%80%9D.

[39] https://www.lucidchart.com/pages/uml-activity-diagram/#section_0

[40] https://www.lucidchart.com/pages/uml-activity-diagram/#section_0

[41] https://www.cs.upc.edu/~marias/teaching/ml/1regression.pdf