جامعة حلوان
HELWAN UNIVERSITY
Faculty of Engineering
Computer Engineering Department

Breast cancer detection system

Supervisor: Prof. Dr/ Zaki

Team members:

- Sama Akram NasrAllah
- Taha Ramadan
- Omar Khaled Mohamed Hemdan

Contents

Introduction

Introduction

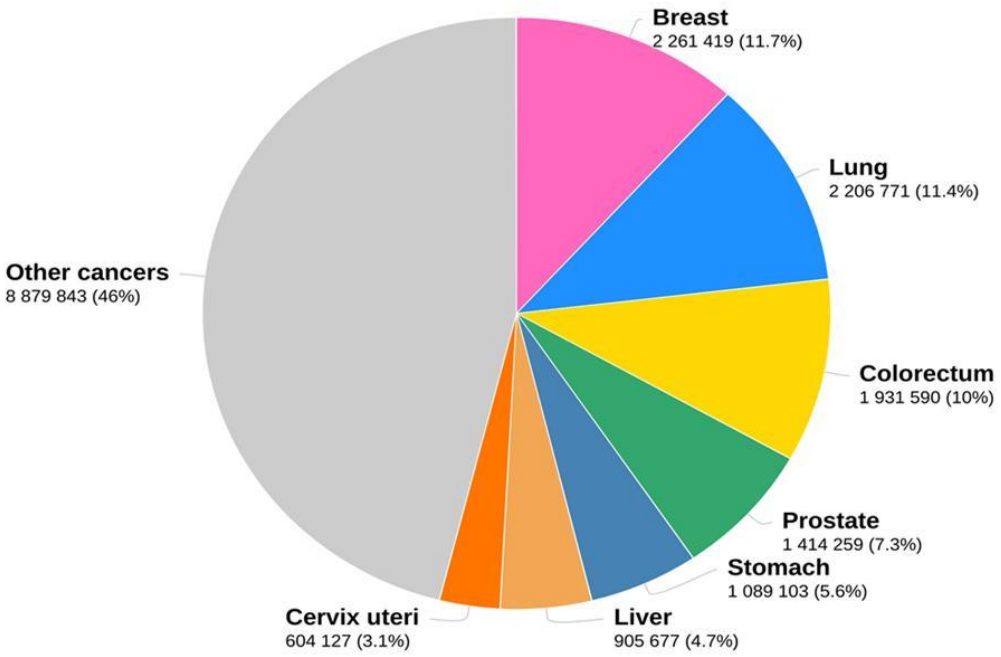
- Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer-related death among women worldwide .
- Biopsy is the main way doctors diagnose most types of cancer. Other tests can suggest that cancer is present, but only a biopsy can make a diagnosis
- There are three types of biopsies: [Core needle biopsy , Fine Needle Aspirate(FNA), Surgical Open Biopsy (SOB)]
- Our project focuses on the latter two, besides predicting the risk of having breast cancer

Problem statement

Problem Statement

- The number of breast cancer cases is significantly increasing, where breast cancer is the most common in terms of new cases of cancer in 2020 and it is the second most common cancer type for women.
- The mortality of BC is very high when compared to other types of cancer due to late and wrong detection.

Number of new cases in 2020

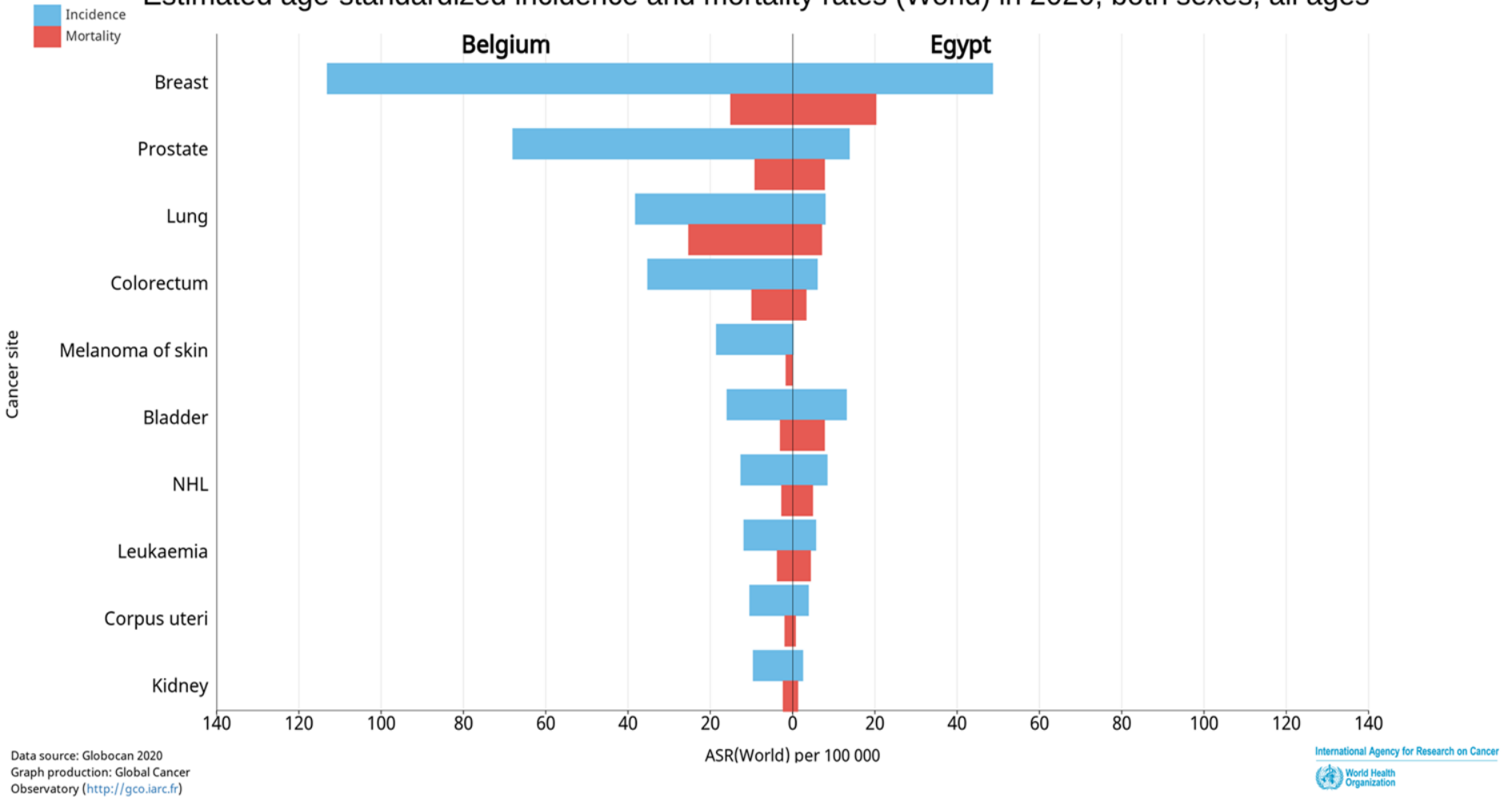


Total : 19 292 789

Data source: Globocan 2020
Graph production: Global Cancer Observatory (http://gco.iarc.fr)

International Agency for Research on Cancer
World Health Organization

Estimated age-standardized incidence and mortality rates (World) in 2020, both sexes, all ages



Data source: Globocan 2020
Graph production: Global Cancer Observatory (http://gco.iarc.fr)

International Agency for Research on Cancer
World Health Organization

Objectives

Objectives

- Our project aims to improve the process of breast cancer detection and increase the number of early detected cases of breast cancer.
- The system delivers an accurate detection of the type of tumor (benign or malignant)
- The system can help any woman to do self-examination.
- The system can help women to determine their likelihood to have breast cancer in the future

Related Work

Related work

1. **Breast cancer detection using the FNA test results and machine learning algorithms.**
   In this area we reviewed 5 research papers that used ML algorithms like (ANN, SVM, LR, Decision tree, KNN) to detect breast cancer.

2. **Breast Cancer Histopathological Image Classification using deep learning.**
   In this area we reviewed 2 research paper that used deep learning algorithms like CNN to determine the type of tumors in using the images of SOB test.

3. **Breast cancer prediction using personal health data and machine learning algorithms.**
   In this area we reviewed 2 research paper that used machine learning regressors like (Decision tree, SVM, LR, Random forest) to determine the likelihood of a woman to have BC in the future .

Related work

.1.1 Comparative Study Between Artificial Neural Networks and Conventional Classifiers for Predicting Diagnosis of Breast Cancer."
They applied 4 machine learning algorithms (SVM, LR, Random forest,ANN) on the Wisconsin breast cancer diagnostic dataset, and the ANN algorithm had the best results with 97.2% testing accuracy.

2. On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset.
They applied 6 machine learning algorithms (KNN, Linear Regression, MLP, GRU-SVM, SVM, Softmax Regression) on the same dataset, and the MLP had the best results

3. A Comparative Study of Machine Learning Algorithms applied to Predictive Breast Cancer Data.

They used 3 machine learning algorithms (KNN, ANN, NB) on the same dataset, and the ANN had the best result with 100% training accuracy.

Related work

4. Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis.
They used 5 machine learning algorithms (SVM, RF, LR, Decision tree, KNN) on the same dataset, and the SVM algorithm had the best result with 96.6% testing accuracy.

5. Comparative Study of Machine Learning Algorithms Using the Breast Cancer Dataset.
They applied 7 machine learning algorithms (KNN, GNB,RF,LR, DT, SGD, ANN) on the same dataset and the LR algorithms had the best result with 97.3% accuracy.

Related work

1. **Breast Cancer Histopathological Image Classification using Convolutional Neural Networks**

They used AlexNet CNN architecture on the BreaKHis dataset to classify the BC histopathological images.
Experimental results obtained on the BreaKHis dataset showed improved accuracy obtained by CNN when compared to traditional machine learning models trained on the same dataset.

2. **Breast Cancer Detection using Histopathological Images**

They have successfully detected malignant tissues in histopathological images using CNN.
They tried 3 different learning rates with Adam Optimizer, and the 1e-4 had the best results with accuracy .89.41%

Related work

3.1Predicting breast cancer risk using personal health data and machine learning models compared to the Gail model results

Models were trained and evaluated on the PLCO data set.

They trained two machine learning models (logistic regression, neural network) that were fed 12 inputs (5 inputs from the Gail model assessment).

logistic regression, linear discriminant analysis, and neural network models with the broader set of inputs predicted five-year breast cancer risk close to the Gail model did.

Related work

3.2Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT (Gail model)

They compared ML-based estimates and estimates from BCRAT model using US population actual observational dataset.

They used the same risk factors as BCRAT model, as input for the ML algorithms in the comparison.

There was an improvement in the accuracy of classification of women with and without breast cancer achieved with ML algorithms compared to the BCRAT (Gail model).

In the US population-based sample, YBCS had fewer affected relatives than their cancer-free relatives .

Machine learning models

Machine learning models

1. **Diagnostic test model**

   Using SVM algorithm and Wisconsin breast cancer dataset we build a model to detect the type of breast tumor (benign or malignant) according to the results of FNA test.

2. **Histopathological Image test model**

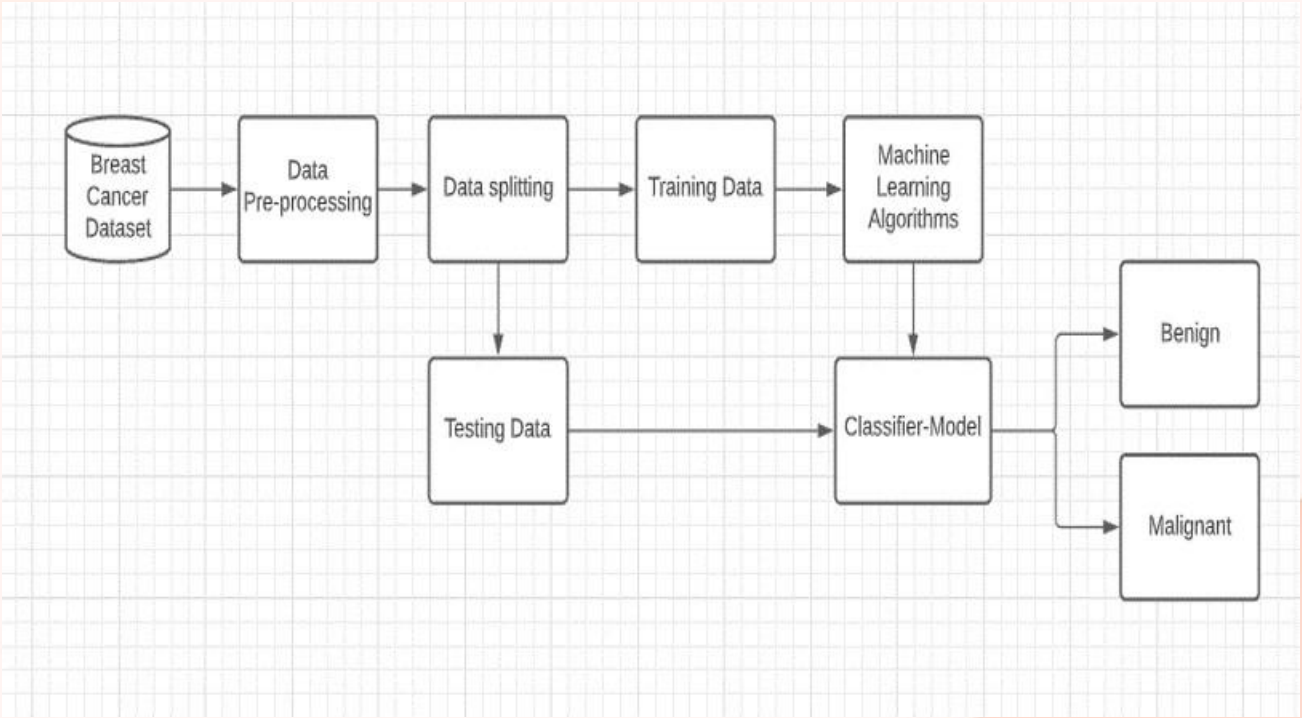
   Using CNN algorithm, transfer learning and BreKHis breast cancer dataset we build a model to detect the type of breast tumor (benign or malignant) according to the results of SOB test.

3. **Risk assessment model**

   Using decision tree regressor and reverse engineering we build a model to determine the likelihood of a woman having breast cancer in the future according to personal health information .

1- Diagnostic test model

1. Diagnostic test model architecture



1. Diagnostic test model

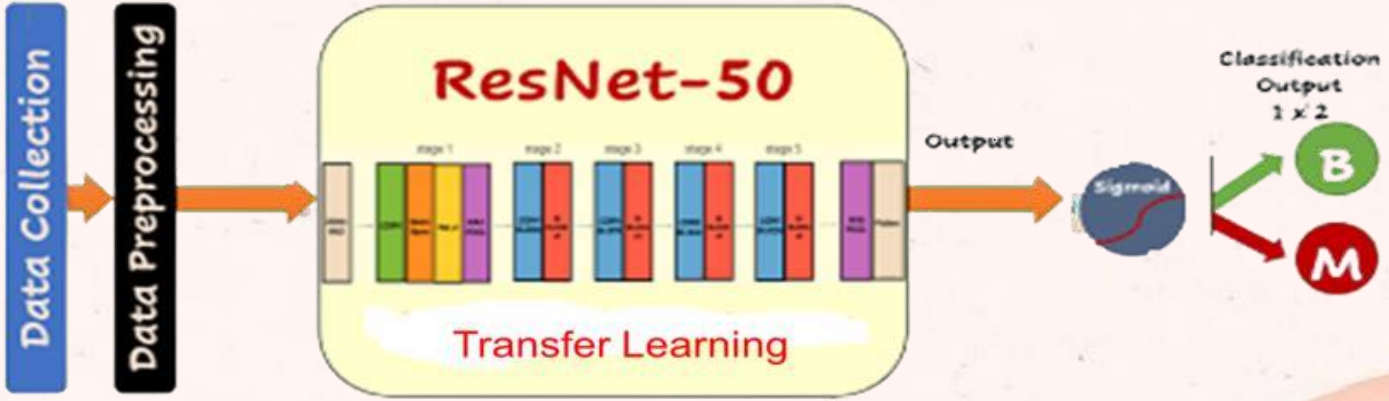
- Based on Wisconsin Dataset - computed from a digitized image of a fine needle aspirate (FNA) of a breast mass
- 569 patient's entries (357 Benign - 212 Malignant) - resulting in 30 features
- Support Vector machine (SVM) obtained the best results with testing accuracy 97.2%
- By using correlation, we obtained the most important 8 features that affect the diagnosis

```
In [102]: df.columns

Out[102]: Index(['diagnosis', 'radius_mean', 'texture_mean', 'smoothness_mean',
                 'compactness_mean', 'symmetry_mean', 'radius_se', 'compactness_se',
                 'fractal_dimension_se'],
                dtype='object')
```

Algorithms	Accuracy Testing Set (%)
Logistic Regression	%95.1
Decision Tree	%86.7
Random Forest	%95.8
KNN	%95.1
SVM	%97.2

2-Histopathological Image test

2. Histopathological Image test architecture



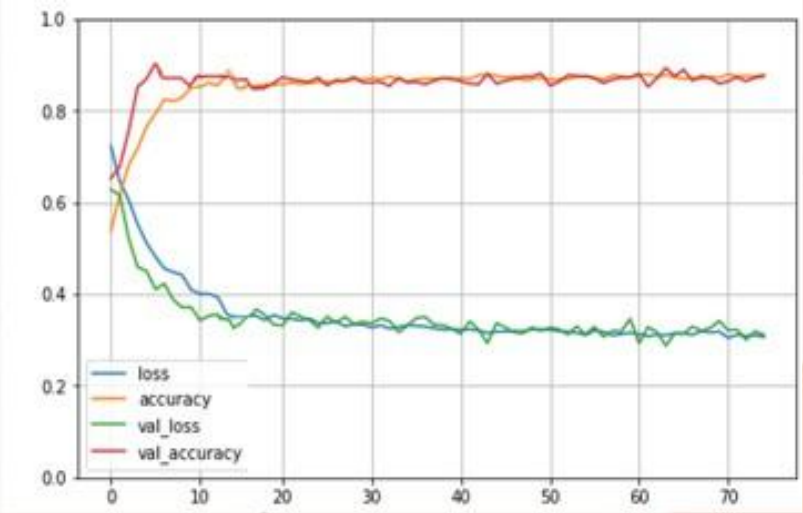
Histopathological Image test

- Histopathology is the study of the signs of the disease using the microscopic images
- Based on BreakHis Dataset - 7909 microscopic images of breast tumor (2,480 benign and 5,429 malignant) tissue collected from 82 patients collected by surgical open biopsy (SOB)
- Transfer learning technique using ImageNet database and RESNET-50 convolution neural network

```
Downloading data from https://storage.googleapis.com/tensorflow/keras-
94674944/94668760 [==============================] - 1s 0us/step
94683136/94668760 [==============================] - 1s 0us/step
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_2 (InputLayer)        [(None, 224, 224, 3)]     0

 resnet50v2 (Functional)     (None, 7, 7, 2048)        23564800

 global_average_pooling2d (G  (None, 2048)             0
 lobalAveragePooling2D)

 dropout (Dropout)           (None, 2048)              0

 dense (Dense)               (None, 1)                 2049

=================================================================
Total params: 23,566,849
Trainable params: 2,049
Non-trainable params: 23,564,800
_________________________________________________________________
```

Histopathological image test Results

```
Epoch 71/75
212/212 [==============================] - 195s 922ms/step - loss: 0.3047 - accuracy: 0.8793 - val_loss: 0.3211 - val_accuracy: 0.8644
Epoch 72/75
212/212 [==============================] - 199s 941ms/step - loss: 0.3115 - accuracy: 0.8759 - val_loss: 0.3223 - val_accuracy: 0.8737
Epoch 73/75
212/212 [==============================] - 200s 944ms/step - loss: 0.3092 - accuracy: 0.8761 - val_loss: 0.3005 - val_accuracy: 0.8644
Epoch 74/75
212/212 [==============================] - 197s 928ms/step - loss: 0.3103 - accuracy: 0.8756 - val_loss: 0.3190 - val_accuracy: 0.8723
Epoch 75/75
212/212 [==============================] - 199s 940ms/step - loss: 0.3066 - accuracy: 0.8792 - val_loss: 0.3114 - val_accuracy: 0.8750
```



3- Risk assessment

3. Risk assessment model

- **What is risk assessment tool?**

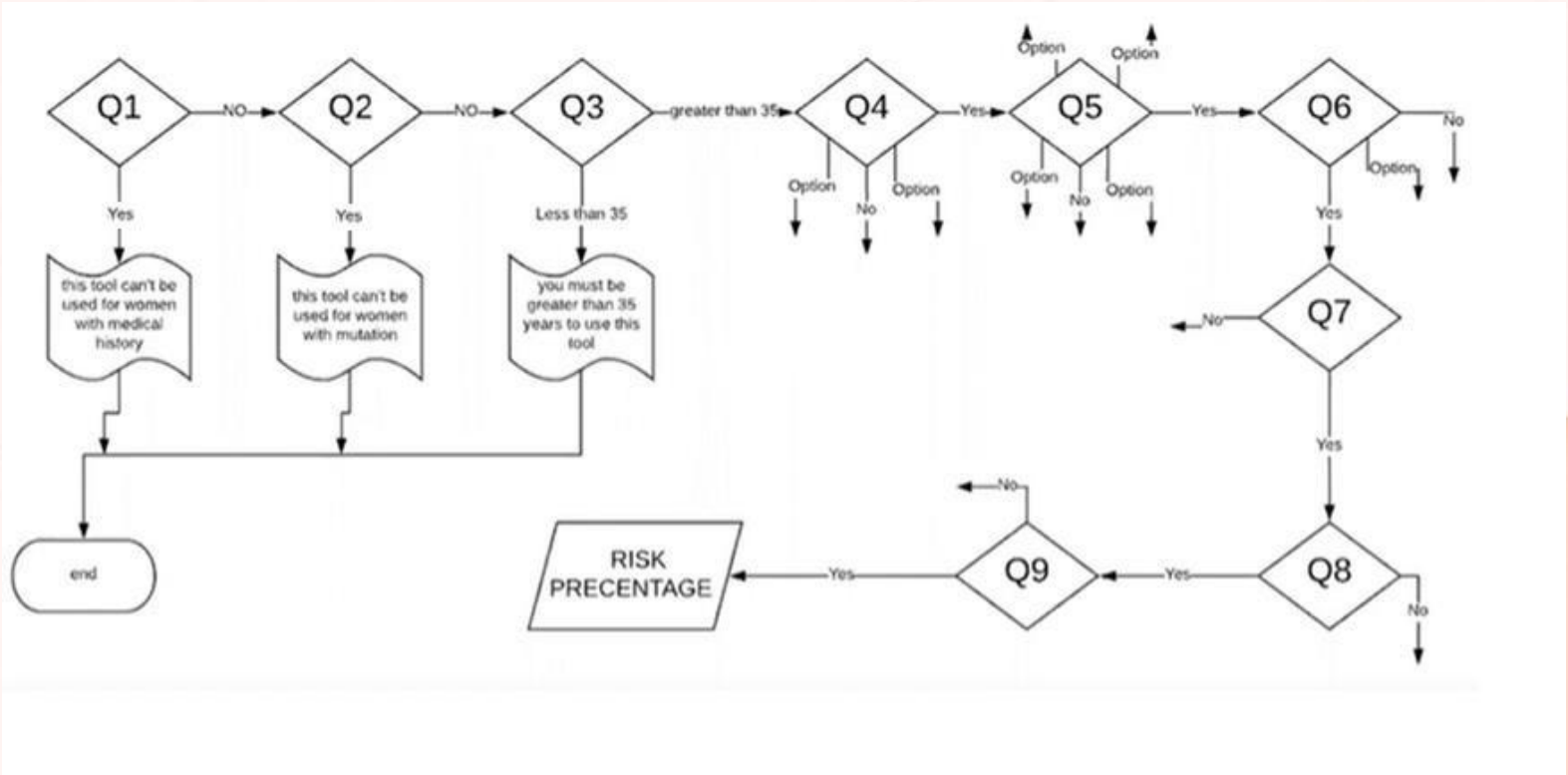
◄ Based on a statistical model known as the Gail Model
◄ The tool estimates patients' risk based on the following personal and family information:

1. Age
2. Age at the start of menstruation
3. Age at first live birth of a child
4. Number of first-degree relative (mother, sisters, daughters) with breast cancer
5. Number of previous breast biopsies (whether positive or negative)
6. Presence of atypical hyperplasia in a biopsy

◄ Combination of question, every question has a weight according to its answer
◄ If the risk ratio were more than 12.5% that means the woman should consult a specialist.

Risk assessment

- Flowchart



Risk assessment

- **problems**

◄ All the research papers worked on private datasets.

- **Creating a dataset (reverse engineering)**

◄ Using Brexa app (online risk assessment tool).
◄ By building a truth table with some of the possible combination of answers.
◄ Try all the answers in the truth table on the app manually to find the risk values.

Risk assessment

- After creating the dataset using reverse engineering as mentioned we tried 4 different machine learning regressors.
- Regression algorithms comparison

Algorithm	variance	Mean sqrd error
Linear regression	0.39	0.02
Decision tree	0.57	0.01
SVM	0.69	0.01
Random forest	0.88	0.00

Risk assessment

- we choose the decision tree regressor for our model to avoid overfitting, and the final coefficients were.

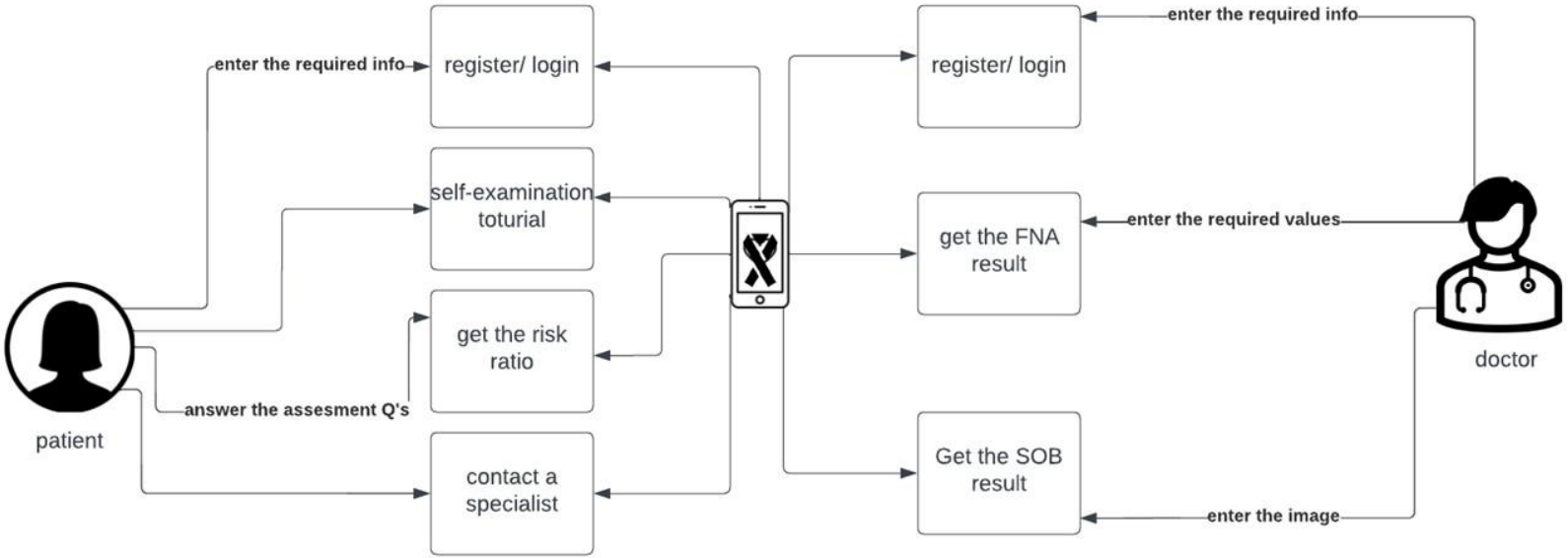
C1	C2	C3	C4	C5	C6
0.04900	0.08818	0.46331	0.0266	0.28366	0.11584

- The equation of decision tree regressor:

$$[(c1*Q4)+(c2*Q5)+(c3*Q6)+(c4*Q7)+(c5*Q9)+(c6*Q9)]$$

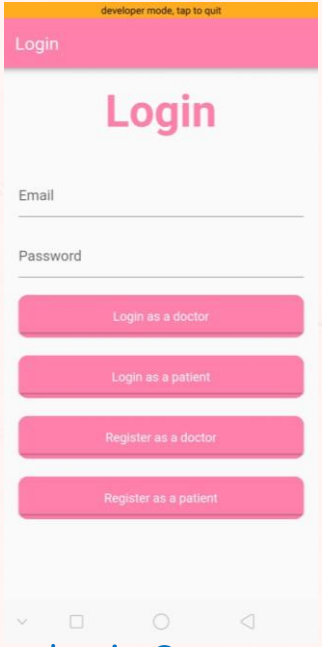
System Architecture

System Architecture

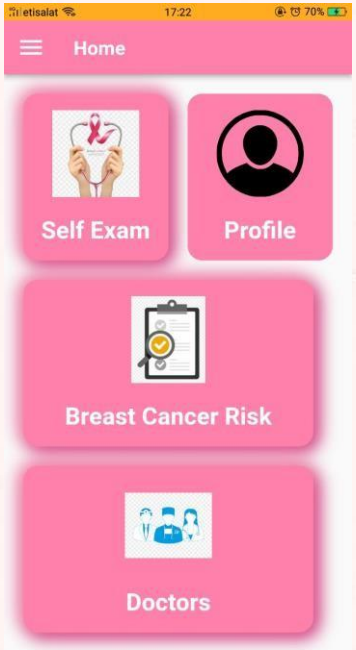


Mobile application user interface

Mobile application user interface



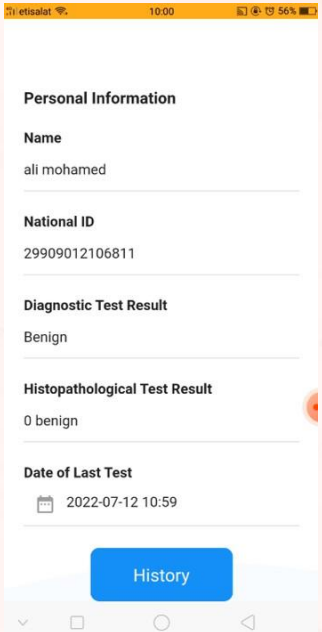
Login Screen



Patient Screen

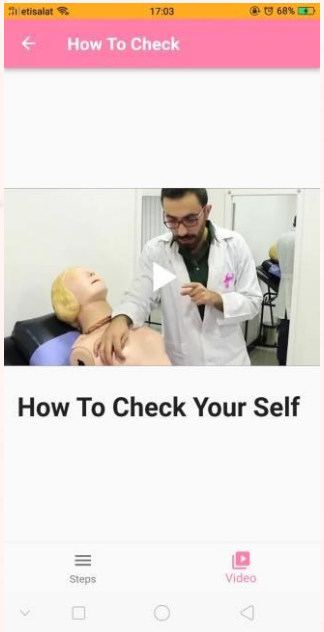
Doctor Screen

Patient user interface

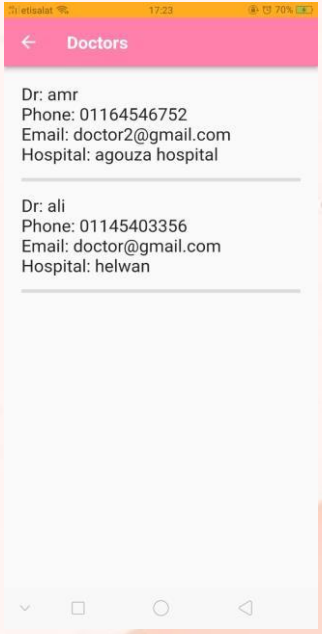


Profile Screen

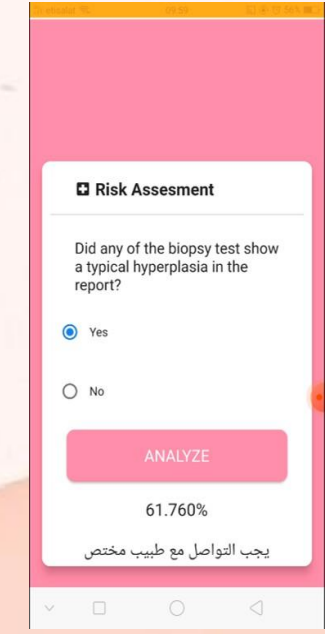
How to check Screen



How to check Screen

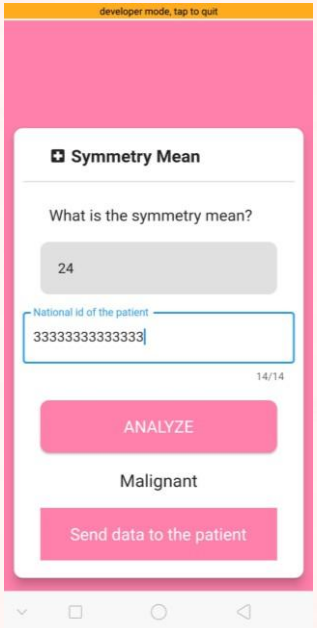


Doctors list Screen

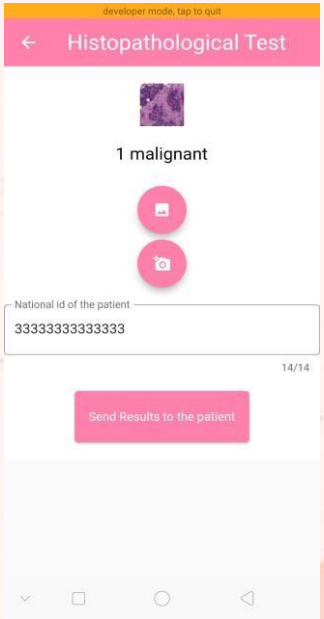


Risk assessment Screen

Doctor user interface



Diagnostic test screen



Histopathological Image test screen

Development tools

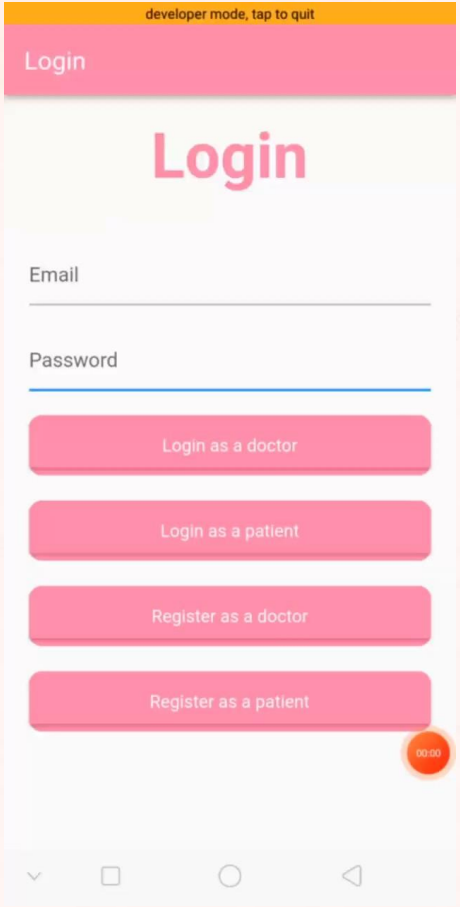
- **Development tools**

python

Future work

Future work

- Trying different algorithms
- Training on local datasets


developer mode, tap to quit
Login
Login
Email
Password
Login as a doctor
Login as a patient
Register as a doctor
Register as a patient
00:00

Thank you