****

**BREAST CANCER DETECTION PROJECT.**

**SYSTEM DOCUMENTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR DIPLOMA IN SOFTWARE ENGINEERING**

**THE DEPARTMENT OF INFORMATION COMMUNICATION TECHNOLOGY IN ZETECH UNIVERSITY.**

## **DECLARATION**

We here, declare that this project is entirely our own original work and that we have done

to the best of our knowledge, it has not been presented or submitted for any Degree or

examination in any other university, and that all the sources we have used or quoted have been

indicated and acknowledged by complete references.

Signature…………………………………

Date…………………………………….

Jeffrey Rowlings DSE-01-0106/2019

**Declaration by the supervisor**

This project has been submitted for examination with my approval as the university supervisor.

Signature……………………………………….

Date……………………………………

Mr. Felix Weru

## 

## **ACKNOWLEDGEMENTS.**

For the success of this research project, it has taken many people their time and resources, either directly or indirectly and the researchers are grateful.

First, I thank Thee, God for His seeing me through the start to the end of this wonderful project and making all things possible for the project to be of great success.

I sincerely acknowledge the support of Ruth Osolo, Mum, in providence of all the resources I needed to make this project a success and the encouragement to continue flexing despite the challenges I encountered.

I would like to express our appreciation to our project supervisor, Mr. Felix Weru, whose constant advise and continuous direction on the research study was utmost important and indeed mandatory for the successful completion of this project, GOD bless you. We would also like to extend the gratitude to ourselves for having the courage to work for the common goal of learning through the turbulence of time to the end of the project. In this regard, not forgetting my classmates, Komora Bada, Kelvin Bett, for their continued support, and with whom we’ve consulted and discussed over issues pertaining to the success of this project. Finally, I also wish to thank the entire Zetech University Fraternity, my lecturers Richard Mathenge, Titus Njeru, just to mention a few, for the continued support and instilling the knowledge that helped me come this far.

**ABSTRACT**

This documentation introduces the process of classifying breast cancer using a machine learning model used by doctors and lab technicians in hospitals. The major part of this project is the machine learning created from jupyter notebook manipulating data to predict type of breast cancer i.e., malignant or benign.

The model is deployed using flask in Heroku.**Heroku** is a platform as a service (PaaS) that enables developers to build, run, and operate applications entirely in the cloud. The implementation uses a tool called Flask Framework micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions

The major part of this report will introduce how to use Flask and Heroku to deploy the machine learning model in order to predict which type of breast cancer a patient has.

.

# **CHAPTER ONE: INTRODUCTION.**

**1.1. Introduction.**

This research aims to classify breast tumors into two categories; malignant and benign, by analyzing data from ultrasound images of breast scan using MATLAB Image Processing Toolbox.

The significance of this research paper is to create a completely new method of data analysis that would produce accurate and credible results that would be on par with results produced by biopsies. With that, oncologists and physicians could be able to confidently rely on the analysis results; the results would be a support to the physician’s decision in proceeding with the treatment for breast cancer. Consequently, patients would not be required to undergo as many biopsies as before this just to confirm the presence of cancer. This would reduce the patients’ risks of the side effects of biopsies and also reduce the overall costs for diagnosis.

This document will propose all features and procedures to develop the system. This document specially containing details about objectives, scope limitation, process model, primary requirements, team development, possible project risks, project schedule, and finally monitoring and reporting mechanisms.

## **1.2 Statement of the problem.**

At the present, oncologists solely depend on biopsy results to determine whether a tumor is cancerous or benign before they could proceed with the treatment. This has proven to be a major inconvenience to patients and oncologists as breast biopsies pose certain risks and disadvantages. First of all, breast biopsies cause bruising or scarring to the patient’s breasts, which consequently may cause infections. Besides that, there is high risk probability that this procedure may cause hematoma; a phenomenon referred to pooling of blood trapped inside the biopsy area.

Apart from health issues, financial issues also cause concern amongst patients due to the high cost of the procedure. The cost of a single biopsy ranges from Ksh 28,000 to Ksh 40,000. The bad news is the state of a tumor cannot usually be determined by a single biopsy; it usually requires a series of biopsies to confirm the presence of cancerous tumor. Commonly, patients are forced to undergo at least three biopsies before oncologists could determine for certain the malignance of the tumor. This directly increases cost of the procedure as well as amplifying the chances of patients to be affected by the risks and side effects of undergoing biopsies.

In some cases, it is possible for biopsies to reveal false positive or false negative results. False positive is a phenomenon when the biopsy results indicate a cancer when it is not actually present, whereas false negative is the phenomenon when the biopsy results miss a cancer when it is actually present. This in turn reduces the accuracy and the reliability of biopsies. Due to these disadvantages and inconveniences, physicians are finding the new means for reducing the number of biopsies, at the same time acquiring accurate results. Though it is impossible to completely eliminate biopsy procedure, it is possible to use image analysis tools to reduce the number of biopsies as well as provide reliable results as it would in a normal biopsy.

## **1.3 Proposed solution.**

This new project will integrate all other ways of predicting breast cancer and implement all the existing data to accurately predict the type of breast cancer a patient has. Hospitals will find it easy to administer fast treatment to patient with the disease. This will be more convenient and significantly improve their efficiency

## **1.4 Objectives**

1. To learn to use Machine learning in the creation of real and practical solutions and improvements to existing problems.
2. To provide a system that will improve efficiency in the hospitals by early detection of breast cancer in a good way that’s harmless to the health of patients.
3. To give hope to breast cancer patients of early detection of breast cancer and administration of treatment.

**PLATFORM**

## **Steps to Install Anaconda and Set Up a Virtual Environment**

Here, are the steps to install Anaconda in your system.

### **Step 1. Installing Python 3**

Building the system, Python 3 would need to be installed on your system. You just have to download the package from the official website, www.python.org, according to your operating system. Keep in mind you have to install a Python 3 version, not the version 2, so just download the stable Python 3 version or the version which it recommends you when you hover the cursor on the downloads link from the website.

While installing Python 3 don’t forget to tick the option “Add Python 3 to your path” when the installation prompt opens.

This should also install pip (pip installs packages) which is the**Python package**manager and we will be using the same to install other software packages.

### **Step 2. Installing Virtualenv**

Virtualenv is a Python package that lets you create different virtual environments for multiple projects requiring different versions of the software.

Its main function is to isolate your project environment with the rest of the system so whatever you install inside this virtual environment will not affect the rest of your system.

### **Step 3. Installing Anaconda**

To access Jupyter lab, Anaconda needs to be installed.

Downloaded from the official website: <https://www.anaconda.com/>

We will be installing Anaconda and some extensions globally on your system, after completing this tutorial you will be able to install anaconda and setup a virtual environment.

### **Step 4. Installing a Python IDE**

This is also an important part of any development environment and you can choose whichever IDE you are familiar with (for Python), for this tutorial I will be using the IDLE and PyCharm. The IDLE comes installed by default when you **install Python in your system**and if you want you can also install PyCharm or any other text editor.

**PROJECT DESIGN**

Creating and running the breast cancer model using jupyter lab.

After installation of everything python visual studio code and anaconda, we start building the model.

1. Run CMD (Command Prompt) in windows as administrator
2. Navigate to the project folder
3. Start anaconda by using the command conda activate”
4. After successful activation of anaconda, start jupyter lab by typing “jupyter lab” in the terminal
5. Jupyter lab then will launch in the computers default browser

**Python**

Python is the language used to build the Django framework. It is a dynamic scripting language. Python supports dynamic typing and has a garbage collector for automatic memory management. Another important feature of Python is dynamic name solution which binds the names of functions and variables during execution.

## 

## **Project management approach:**

**Software Process Model**:

To solve an actual problem in an industry, software developer or a team of developers must integrate with a development strategy that include the process, methods and tools layer and generic phases. This strategy is often referred to a process model or a software developing paradigm.

## **My project follows the waterfall model.**

The steps of waterfall model are:

• Requirement Definition

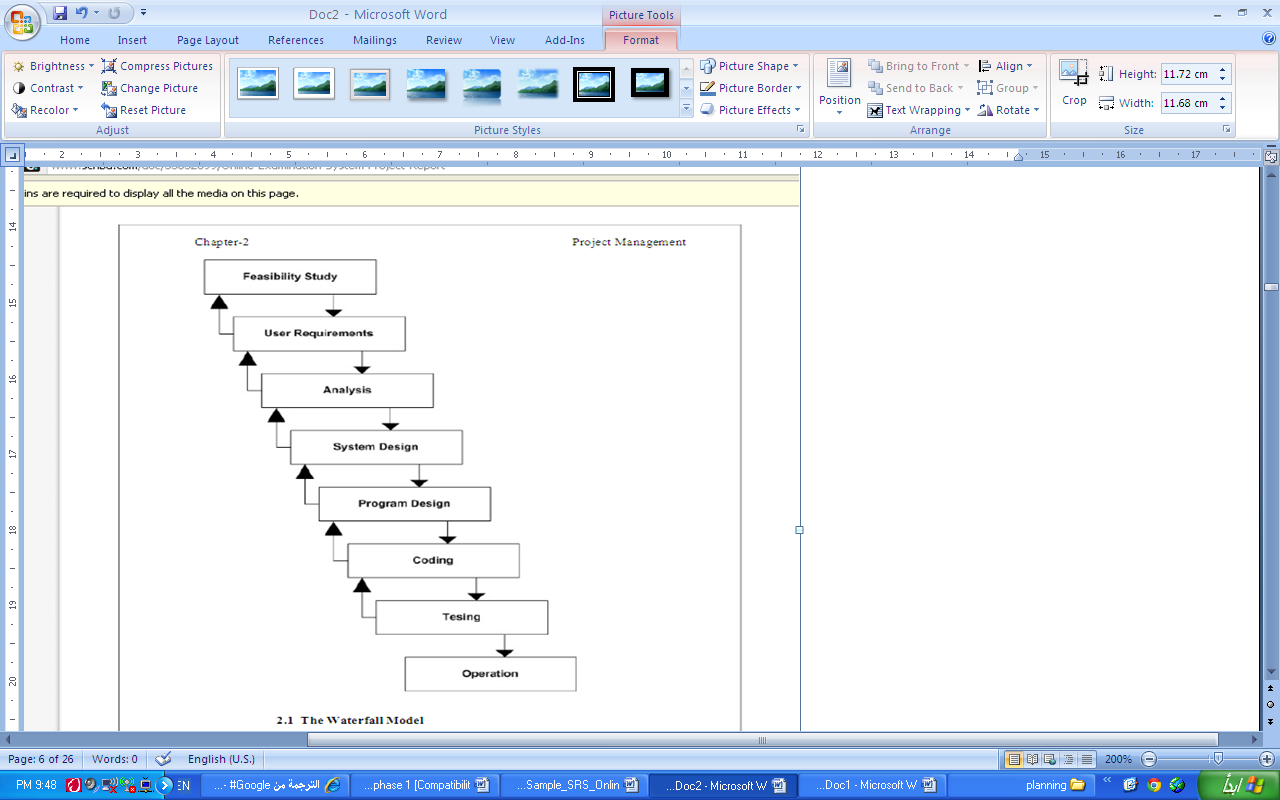
• System and Software Design

• Implementation

• Integration and System Testing

• Operation and Maintenance

**Waterfall model.**



# 

# **CHAPTER TWO: SYSTEM REQUIREMENT SPECIFICATION.**

## **2.1 Preface:**

This document is the basic intended for any individual user, developer, tester, project manager or documentation writer that needs to understand the basic system architecture and its specifications.

## **2.2 Introduction:**

The purpose of this SRS document is to write the functional and non-functional user or system requirements that represent the characteristics of Breast Cancer Classification Project.

## **2.3 The scope and limitation of this system is:**

* The system is designed for hospitals
* Predicts patients type of breast cancer

**Software Quality Attributes**

The Quality of the System is maintained in such a way so that it can be very user friendly to all the users.

The software quality attributes are assumed as under:

* Accurate and hence reliable.
* Secured.
* Fast speed.
* Compatibility.

**System Interfaces:**

This section describes how the software interfaces with other software products or users for input or output.

**User Interface**

Application will be accessed through a Browser Interface.

**System Models:**

In this system we are use waterfall model to apply these ideas. Which is help us to separate each step and when we finish a one phase the output of it is the input to the next phase. Also, we can backwards if there is a new requirement or to apply any update.

# **CHAPTER THREE: SYSTEM DESIGN:**

## **3.1 Introduction:**

Design is the abstraction of a solution; it is a general description of the solution to a problem without the details. Design is view patterns seen in the analysis phase to be a pattern in a design phase. After design phase we can reduce the time required to create the implementation.

In this chapter we are introduce context diagram, models, system architecture, principal system object, design model and object interface.

## 3.3 **Models**:

Interaction model:

Is a dynamic model that shows how the system interacts with its environment. We use a data flow diagram.

**Use case diagram:**

DOCTOR

LAB TTECHNICIAN

PATIENT







3.4 Activity Diagram**:**

DOCTOR

Results Processing

Patients details

Gtives Results

Offer Medication

Screens the patients

Gets Results

Lab Technician

Breast Cancer Prediction

Insert the screens in the system

Results

Request Patients Screens

View Report

Figure: the activity diagram for basic operation in system. (a) For the doctor, (b) for the Lab Technician .

**CHAPTER FOUR: REQUIREMENTS SPECIFICATION**

## **4.1 Introduction**

This chapter gives a detailed explanation about the requirements of the system.

## **4.2 Software Requirements**

One of the most difficult tasks is that, the selection of the software, once system requirement is known is determining whether a particular software package fits the requirements. After initial selection further security is needed to determine the desirability of particular software compared with other candidates. This section first summarizes the application requirement question and then suggests more detailed comparisons.

• Operating System-------- ------------------- Windows 10, 7 8

* Browser --------------------------- Chrome
* IDE ---------------------------- Anaconda
* Other Tools & Technologies------------ PYTHON, HTML, CSS,

## **4.3 Hardware Requirements**

The selection of hardware is very important in the existence and proper working of any software. In the selection of hardware, the size and the capacity requirements are also important.

The Web Based Manufacturing System can be efficiently run on Pentium system with at least 128 MB RAM and Hard disk drive having 20 GB. Floppy disk drive of 1.44 MB and 14-inch Samsung color monitor suits the information system operation. (A Printer is required for hard copy output).

* Pentium processor -------- 233 MHZ or above
* RAM Capacity -------- 128MB
* Hard Disk -------- 20GB
* Floppy disk -------- 1.44 MB
* CD-ROM Drive -------- 32 HZ

• Keyboard -------- 108 Standard

## 

## **4.4 Performance Requirements**

Performance is measured in terms of the output provided by the application. Requirement specification plays an important part in the analysis of a system. Only when the requirement specifications are properly given, it is possible to design a system, which will fit into required environment. It rests largely with the users of the existing system to give the requirement specifications because they are the people who finally use the system. This is because the requirements have to be known during the initial stages so that the system can be designed according to those requirements. It is very difficult to change the system once it has been designed and on the other hand designing a system, which does not cater to the requirements of the user, is of no use.

The requirement specification for any system can be broadly stated as given below:

* The system should be able to interface with the existing system
* The system should be accurate
* The system should be better than the existing system

The existing system is completely dependent on the user to perform all the duties.

## **4.5 Feasibility Study**

Preliminary investigation examines project feasibility; the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All systems are feasible if they are given unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

* Technical Feasibility
* Operation Feasibility

• Economic Feasibility

## **4.5.1 Technical Feasibility**

The technical issue usually raised during the feasibility stage of the investigation included the following:

* Does the necessary technology exist to do what is suggested?
* Do the proposed equipment have the technical capacity to hold the data required to use the new system?
* Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?
* Can the system be upgraded if developed?
* Are there technical guarantees of accuracy, reliability, ease of access and data security?

## **4.5.2 Operational Feasibility**

**User-friendly**

Customer will use the forms for their various transactions i.e. for adding new routes, viewing the routes details. Also, the Customer wanted the reports to view the various transactions based on the constraints. These forms and reports are generated as user-friendly to the Client.

**Reliability**

The package wills pick-up current transactions on line. Regarding the old transactions, User will enter them in to the system.

**Security**

The web server and database server should be protected from hacking, virus etc.

**Portability**

The application was developed using standard open source software (Except Oracle) like Java, tomcat web server, Internet Explorer Browser etc. this software will work both on Windows and Linux operating system. Hence portability problems will not arise.

**Availability**

This software will be available always.

**Maintainability**

The system uses the 2-tier architecture. The first tier is the GUI, which is said to be front-end and the 2nd tier is the database, which uses MySQL, which is the back-end.

The front-end can be run on different systems (clients). The database will be running at the server. Users access these forms by using the user-ids and the passwords.

**4.5.3 Economic Feasibility**

The computerized system takes care of the present existing system’s data flow and procedures completely and should generate all the reports of the manual system besides a host of other management reports.

It was built as a web-based application with separate web server and database server. This was required as the activities are spread throughout the organization customer wants a centralized database. Further some of the linked transactions take place in different locations.

Open source software like TOMCAT, JAVA, MySQL and Linux is used to minimize the cost for the Customer.

**CHAPTER FIVE: CODING AND TESTING OF RESULTS**

## **Introduction**

Coding is the process of creating the main system. The system was developed in accordance to all the requirements provided. Testing of the system is also discussed in this chapter.

**Requirements for deployment**

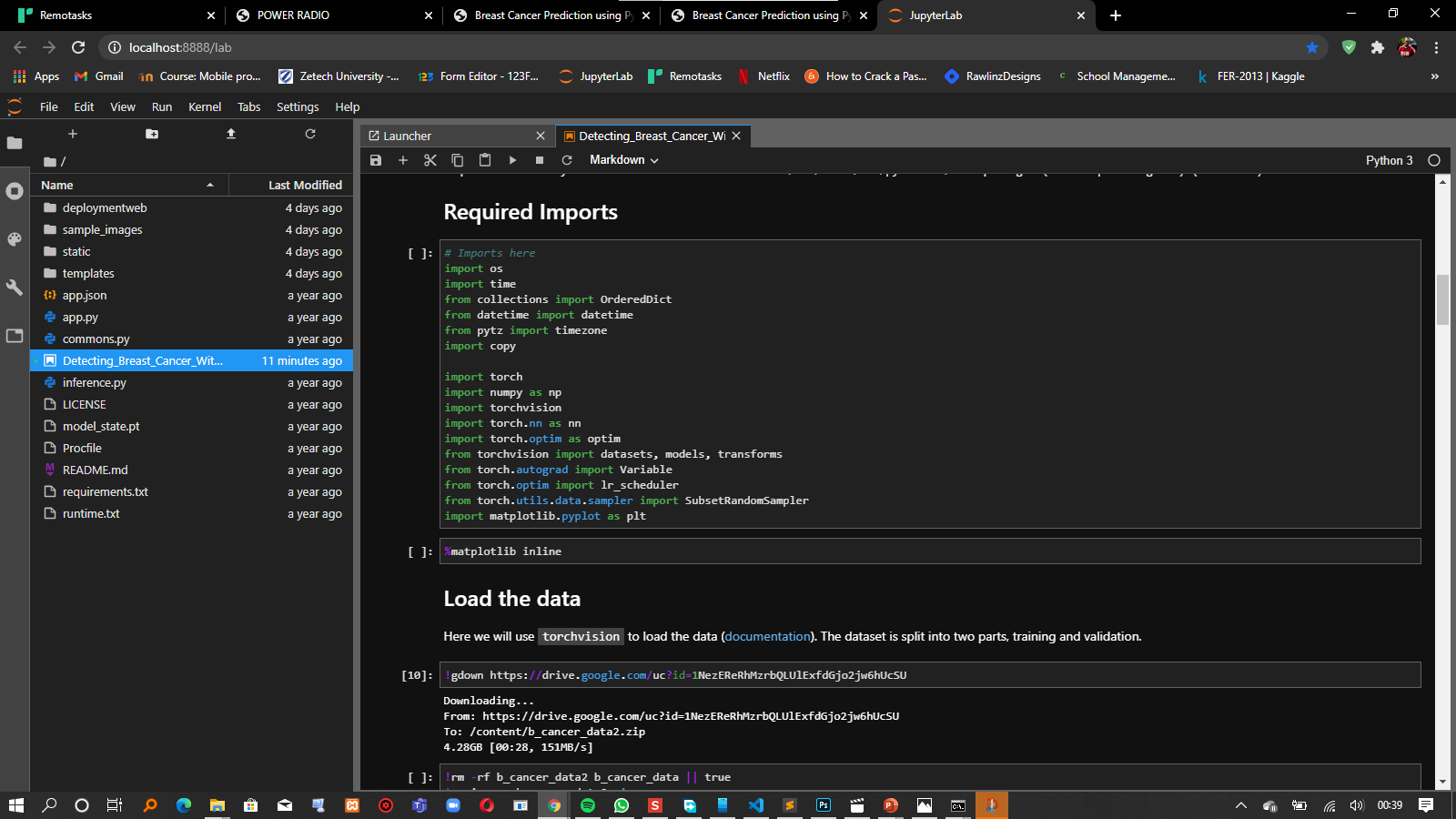
* Flask
* NumPy
* Torch

**MODEL IN JUPYTER LAB**

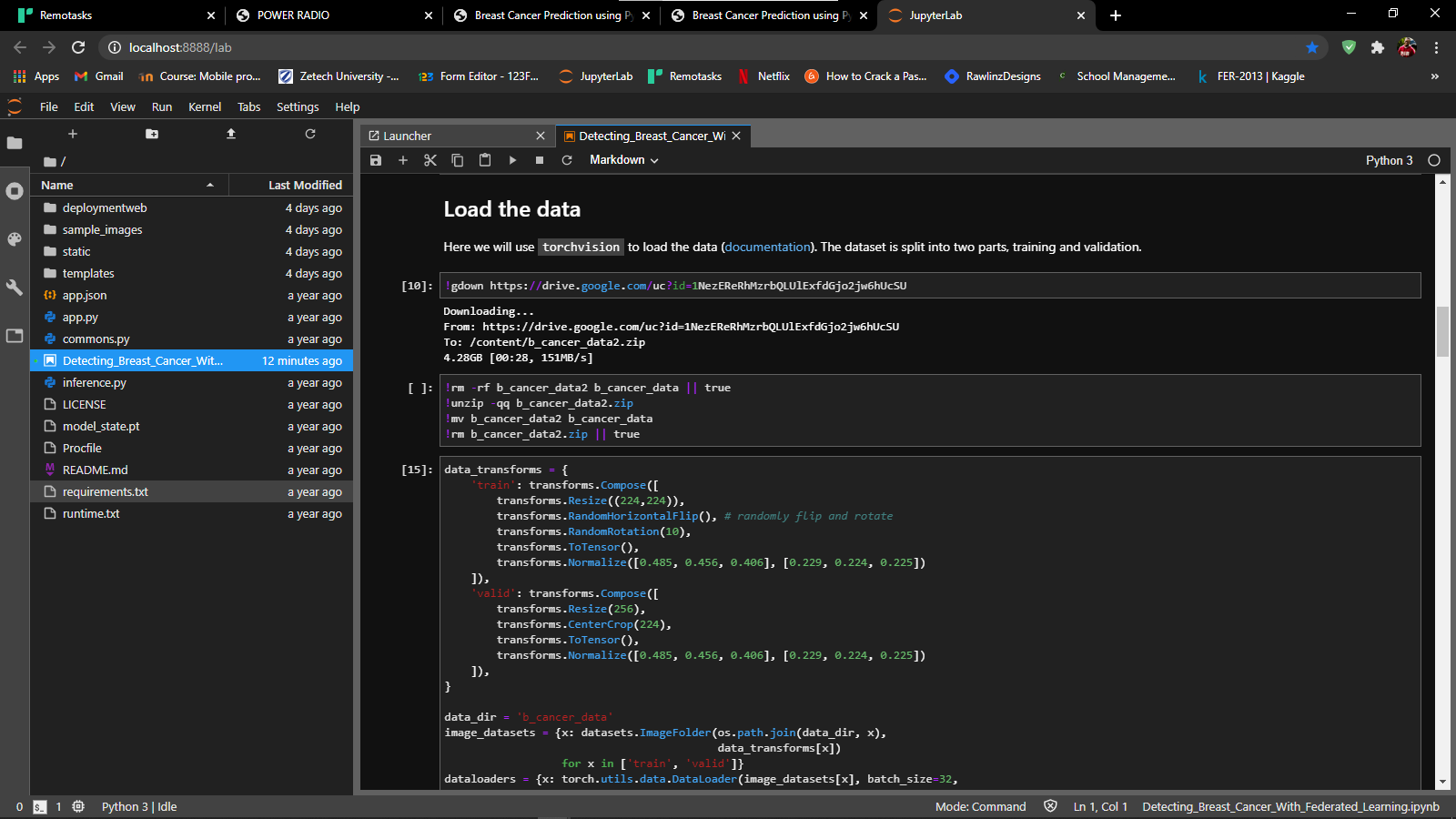
**Importing Libraries**

**The Libraries:**

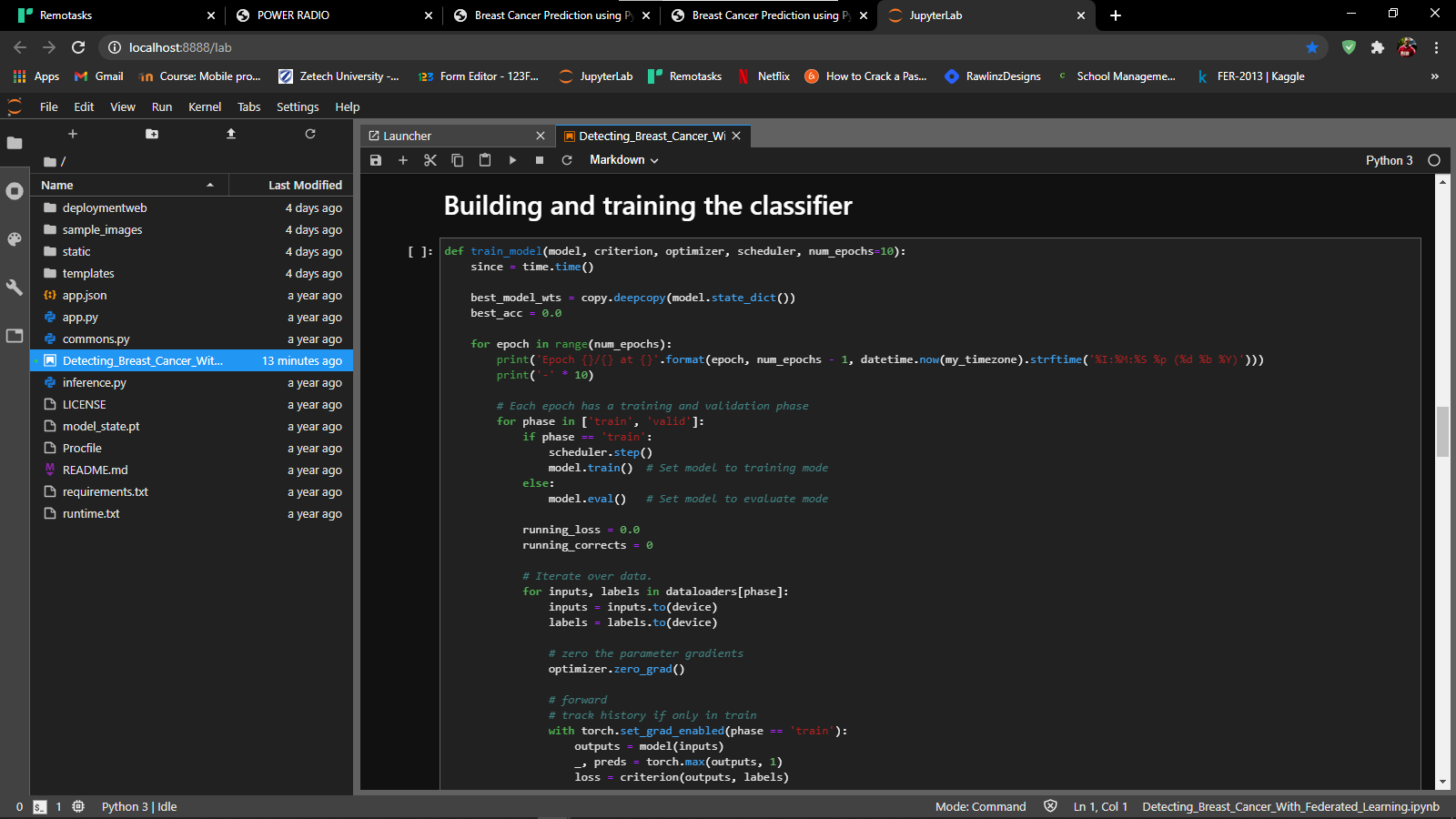
* Torch
* NumPy
* Matplotlib
* Torchvision
* pandas



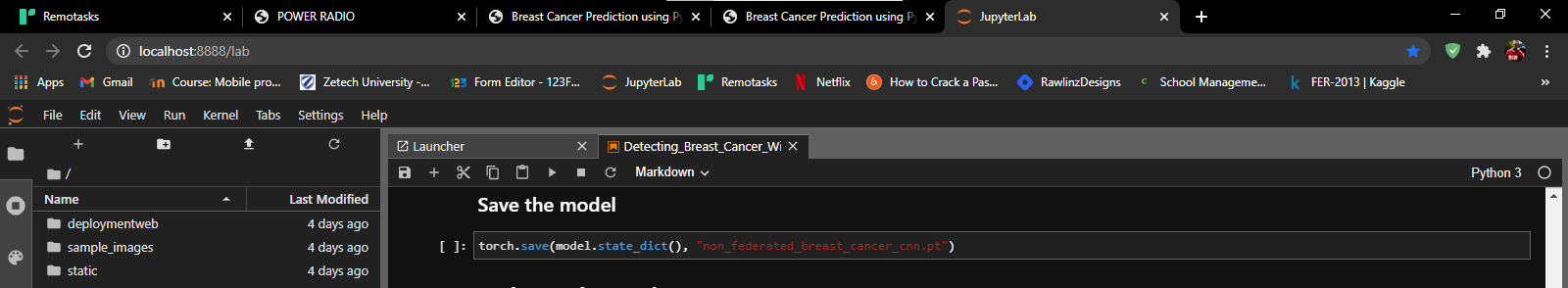
**LOADING THE DATA**



**Building and Training The model**



**SAVING THE MODEL FOR DEPLOYMENT**



**HOME PAGE**

**OUTPUT SCRENS**

## 

**SYSTEM TESTING**

## **Introduction to Testing:**

Testing is a process, which reveals errors in the program. It is the major quality measure employed during software development. During software development. During testing, the program was executed with a set of test cases and the output of the program for the test cases was evaluated to determine if the program was performing as it is expected to perform.

## 

## **Testing Strategies.**

In order to make sure that the system does not have errors, the different levels of testing strategies that are applied at differing phases of software development are:

**Unit Testing:**

Unit Testing was done on individual modules as they are completed and become executable. It is confined only to the designer's requirements.

Each module can be tested using the following two Strategies:

**Black Box Testing:**

In this strategy some test cases were generated as input conditions that fully execute all functional requirements for the program. This testing has been used to find errors in the following categories:

i. Incorrect or missing functions

ii. Interface errors

iii. Errors in data structure or external database access

iv. Performance errors

v. Initialization and termination errors.

In this testing only the output is checked for correctness. The logical flow of the data is not checked.

**White Box testing:**

In this the test cases were generated on the logic of each module by drawing flow graphs of that module and logical decisions were tested on all the cases. It has been uses to generate the test cases in the following cases:

i. Guarantee that all independent paths have been executed.

ii. Execute all logical decisions on their true and false Sides.

iii. Execute all loops at their boundaries and within their operational bounds

iv. Execute internal data structures to ensure their validity.

**Integrating Testing:**

Integration testing ensured that software and subsystems work together a whole. It tested the interface of all the modules to make sure that the modules behave properly when integrated together.

**System Testing:**

Involved in-house testing of the entire system before delivery to the user. Its aim was to satisfy the user the system meets all requirements of the client's specifications.

**Acceptance Testing:**

It is a pre-delivery testing in which the entire system was tested at client's site on real world data to find errors

# **CHAPTER SIX: CONCLUSIONS & RECOMMENDATIONS**

## **6.1 Introduction**

This chapter gives an overview of whether the system met all the objectives, conclusions and recommendations about the system.

## **6.2 Conclusion**

Globally, breast cancer is the most common cancer amongst women. In 2012, almost 1.7 million women were diagnosed with breast cancer, a figure which accounts for 25% of cancer cases amongst women. The contention driving this ERC-funded research is that scientists still do not know enough to explain why some patients respond well to therapeutic treatments, whilst others continue to decline.

Breast cancer is the second leading cause of cancer death in women, after lung cancer according to the ACS.The society estimates that about 40,450 women and 440 men will likely die of breast cancer in 2016, but significantly more patients will defeat the disease. In fact, breast cancer deaths have been declining annually since about 1989, especially among women under age 50. Regular screenings, early detection and better treatments are credited for improved prognosis.

## **6.3 Recommendations**

* Adoption of using AI as a means to detect early breast cancer and start administering treatment so as to save lives
* Drug visualization to improve cancer treatments

**REFERENCES**

**1**. N. Salim, *Medical Diagnosis Using Neural Network*, Faculty of Information Technology University, 2013, <http://www.generation5.org/content/2004/MedicalDiagnosis.asp>.

2. A.Mert,N.Kilic¸, andA.Akan, “Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats,” *Neural Computing and Applications*, vol. 24, no. 2, pp. 317–326, 2014.

3. R. W. Brause, “Medical analysis and diagnosis by neural networks,” in *Proceedings of the 2nd International Symposium on* *Medical Data Analysis (ISMDA ’01)*, pp. 1–13, Madrid, Spain, October 2001.

4. R. Huang, L. Law, and Y. Cheung, “An experimental study: on reducing RBF input dimension by ICAand PCA,” in *Proceedings* *of the 2002 International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 1941–1945, November 2002.

5. S. Kara, A. G¨uven, and S. Ic¸er, “Classification of macular and optic nerve disease by principal component analysis,” *Computers* *in Biology and Medicine*, vol. 37, no. 6, pp. 836–841, 2007.

6. W. H.Wolberg,W. N. Street, and O. L. Mangasarian, “Machine learning techniques to diagnose breast cancer from imageprocessed nuclear features of fine needle aspirates,” *Cancer* *Letters*, vol. 77, no. 2-3, pp. 163–171, 1994

7. K. H. Liu, B. Li, Q. Q. Wu, J. Zhang, J. X. Du, and G. Y. Liu, “Microarray data classification based on ensemble independent component selection,” *Computers in Biology and Medicine*, vol.

39, no. 11, pp. 953–960, 2009.

8. 2013, <http://research.ics.tkk.fi/ica/fastica/>