A PRELIMENERY REPORT ON

# Thermal Anomaly detection and Breast Cancer Risk Prediction: Using DITI and Deep Learning Approach

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

**BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)**

**SUBMITTED BY**


1. **Mr.Samyak Gangwal** [Exam Seat No: 71816475L ]
2. **Mr.Yash Kewlani** [Exam Seat No: 71816555B ]
3. **Mr.Nikhil Pawar** [Exam Seat No: 71816658C ]
4. **Mr.Prasanna Dixit** [Exam Seat No: 71816454H ]




**DEPARTMENT OF COMPUTER ENGINEERING**

**Maratha Vidya Prasarak Samaj's**

**Karmaveer Adv. Baburao Ganpatrao Thakare**

**College of Engineering, Nashik-13**

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**2020-21**


# CERTIFICATE

This is to certify that the project report entitles

" **Thermal Anomaly detection and Breast Cancer Risk Prediction: Using DITI and Deep Learning Approach** "

Submitted by

1. **Mr.Samyak Gangwal** [Exam Seat No: 71816475L ]
2. **Mr.Yash Kewlani** [Exam Seat No: 71816555B ]
3. **Mr.Nikhil Pawar** [Exam Seat No: 71816658C ]
4. **Mr.Prasanna Dixit** [Exam Seat No: 71816454H ]

are a bonafide students of this institute and the work has been carried out by him/her under the supervision of **Dr. V. S. Pawar** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering**(Computer Engineering).

**Dr. V. S. Pawar**
Guide
Department of Computer Engineering

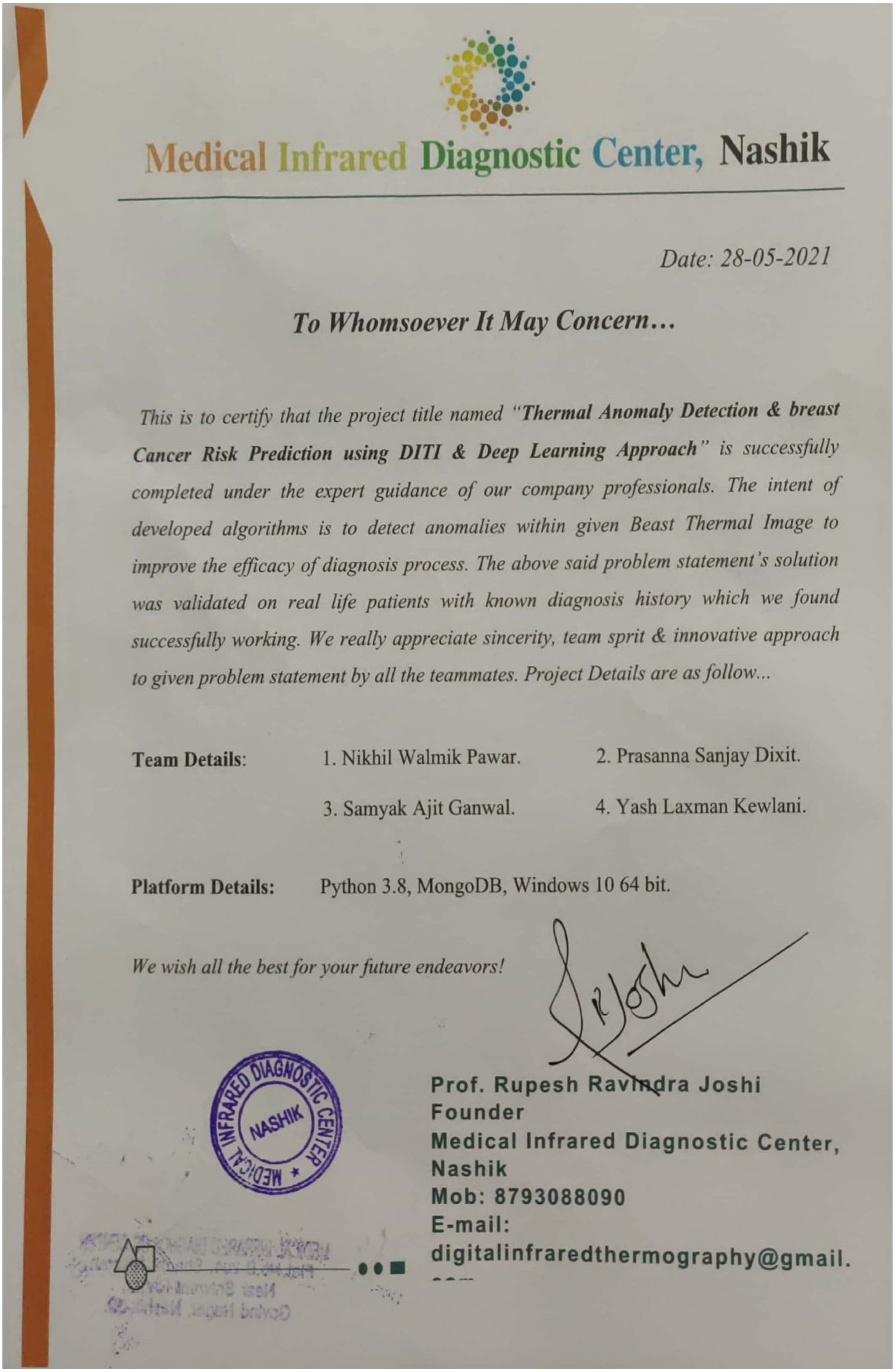
**Prof. R. Joshi**
External Guide
Medical Infrared Diagnostic Center

**Dr. V. S. Pawar**
Head
Department of Computer Engineering

**Dr. S. R. Devane**
Principal
MVPS's K.B.T. College of Engineering
Nashik – 13

Place: Nashik

Date:



# Medical Infrared Diagnostic Center, Nashik

*Date: 28-05-2021*

***To Whomsoever It May Concern…***

*This is to certify that the project title named "**Thermal Anomaly Detection & breast Cancer Risk Prediction using DITI & Deep Learning Approach**" is successfully completed under the expert guidance of our company professionals. The intent of developed algorithms is to detect anomalies within given Beast Thermal Image to improve the efficacy of diagnosis process. The above said problem statement's solution was validated on real life patients with known diagnosis history which we found successfully working. We really appreciate sincerity, team sprit & innovative approach to given problem statement by all the teammates. Project Details are as follow...*

**Team Details**: 1. Nikhil Walmik Pawar. 2. Prasanna Sanjay Dixit.
3. Samyak Ajit Ganwal. 4. Yash Laxman Kewlani.

**Platform Details:** Python 3.8, MongoDB, Windows 10 64 bit.

*We wish all the best for your future endeavors!*

MEDICAL INFRARED DIAGNOSTIC CENTER NASHIK

**Prof. Rupesh Ravindra Joshi**
**Founder**
**Medical Infrared Diagnostic Center, Nashik**
**Mob: 8793088090**
**E-mail: digitalinfraredthermography@gmail.**

# Acknowledgements

Foremost, We would like to express our gratitude to all those people who have helped me. It would have been insurmountable without the friendly guidance of professors and people at **Medical Infrared Diagnostic Center** and assistance of all the people involved in the work.
We are profoundly grateful to Dr.V.S. Pawar, Head of Department, Computer Engineering and Prof. Rupesh Joshi, Director, Medical Infrared Digital Imaging, Nashik for permitting us to undertake this work together with their kind and inspiring advice which helped us to understand the subject and its semantic significance. They enriched us with valuable additions to our topic furthermore ameliorating presentation skills. Finally, We would like to appreciate all the teachers and mentors throughout our academic career without whom we would not have achieved this milestone.
We are also tremendously grateful to all of our peers who helped and assisted us in completing this project by their active participation.

Samyak gangwal
Yash Kewlani
Nikhil Pawar
Prasanna Dixit

# Abstract

Breast cancer is one of the most fatal diseases responsible for the death of women all over the world. About one in eight of women is subject to breast cancer over the course of her lifetime. There is no effective method to prevent or know the reasons for the growth of these cancerous cells, however the number of deaths can be reduced by early detection. Early detection of breast cancer will facilitate the treatment process. For this purpose we are using DITI and a deep learning approach. Digital Infrared Thermal Imaging (or DITI) is a non-invasive clinical imaging procedure used to detect and monitor many diseases and physical injuries. The images generated by the infrared scanning device allow us to to visualize and quantify changes in skin surface temperature. The approach we are using will consist of two outputs. In the first output we are detecting thermal anomaly in the Region of Interest(ROI). For obtaining this output, we will preprocess the thermal images to obtain ROI and will apply a detection algorithm to find the thermal anomaly. The second output will consist of a prediction of the risk of breast cancer. For this we would be considering the thermal pre-processed images. These images would be used for extracting characteristic features using image processing. These collected features are used as input for the Back Propagation Neural Network(BPNN) to predict the risk of Breast Cancer.

**Index Terms: Digital Infrared Thermal Imaging, Back Propagation Neural Network, Region of Interest, Breast Cancer, Thermal Anomaly**

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| DITI | Digital Infrared Thermal Imaging |
| ROI | Region Of Interest |
| GLCM | Grey Level Co-occurence Matrix |
| BPNN | Back Propagation Neural Network |
| NN | Neural Network |
| GUI | Graphical User Interface |
| MRI | Magnetic Resonance Imaging |
| ReLU | Rectified Linear Unit |

# Chapter 1

# Introduction

## 1.1 Overview

The purpose of this project is to find thermal anomalies in the thermal scans generated using DITI and finding a novel set of useful features which will be used for predicting the probability of positive breast masses using back propagation neural network.

The solution deals in medical diagnostic field to give alert in early stages & detect future occurrence of life threatening disease and eventually reduce the mortality rate. The operational model is very simple, we need to follow some protocol before taking scan by the doctor. The principle of operation of this modality is unique when compared with current available modalities in medical field. The important operational difference between our solution and present modalities is that we monitor physiological aspect and interpretations are based on this fact. The present scenario for detection of Breast Cancer is based on appreciable anatomical changes, which causes quite delay in detection resulting increase in mortality rates and serious medical complications. As our solution monitors precancerous physiological changes on early stages which helps physician to decide line of medication well in advance to save the life of patient. Added advantages of this solution are Radiation Free, No Touch – No Pain, Non-Invasive, suites for wide category of patient, can be used repeatedly, portable.

Medical infrared diagnostic center uses the technique of DITI for diagnosing different type of medical conditions like broken bones, thyroid related problems, early cancer detection, diabetes, etc. The organization is led by Prof. Rupesh Joshi who is also the mentor for this project.

## 1.2 Motivation

Cancer is the name given to a collection of related diseases. In all types of cancer, some of the body's cells begin to divide without stopping and spread into surrounding tissues. Cancer can start almost anywhere in the human body, which is made up of trillions of cells. Normally, human cells grow and divide to form new cells as the body needs them. When cells grow old or are damaged, they die, and new cells replace them. When cancer develops, however, this orderly process breaks down. As cells become more and more abnormal, old or damaged cells survive when they should die, and new cells form when they are not needed. These extra cells can divide without stopping and may form growths called tumors. When we consider the cancer cells occurring in women, Breast cancer is among the most frequently occurring disease. Breast cancer is cancer that forms in the cells of the breasts. After skin cancer, breast cancer is the most common cancer diagnosed in women in the United States. Breast cancer can occur in both men and women, but it's far more common in women. Early detection and treatment of cancer minimize the risk of deaths and increase the survival rate[1].

Traditional ways of detecting breast cancer:

**Mammogram** is an x-ray picture of the breast. It can be used to check for breast cancer in women who have no signs or symptoms of the disease. It can also be used if you have a lump or other sign of breast cancer. Mammogram involves radiation and can only detect if the cancer cells are big enough.

**Breast MRI** (magnetic resonance imaging) uses radio waves and strong magnets to make detailed pictures of the inside of the breast. only for high risk due to high rate of false positive.

**Digital Infrared Thermal Imaging** (or DITI) is a non-invasive clinical imaging procedure used to detect and monitor many diseases and physical injuries. The images generated by the infrared scanning device allow a certified examiner to visualize and quantify changes in skin surface temperature. The scanning device converts heat (infrared radiation) emitted from the skin surface into electrical impulses that are displayed in color on a monitor. This visual image graphically maps the body temperature of an individual (this image is called a thermo-gram). The spectrum of colors indicates an increase or decrease in the amount of heat emitted from the body surface. Since there is a high degree of thermal symmetry in the normal body, subtle abnormal temperature asymmetries are identified easily by a trained clinician.

## 1.3 Problem Statement and Objectives

### 1.3.1 Problem Statement

To automate the process of early detection of cancer using DITI by finding Breast anomalies using image processing and probability of positive breast cancer masses using BPNN.

### 1.3.2 Objectives

- Early prediction of Breast cancer risk to minimize the death rate using DITI and BPNN.
- To automate the process of examining the thermographic images to find anomalies using image processing.

## 1.4 Project Scope and Limitations

### 1.4.1 Project Scope

This project will consist of creating an automated process for finding anomalies and risks in thermal scans of patients for early detection of severe medical conditions, which aims to streamline the already existing process. The project includes finding thermal anomalies and predicting probability values of positive thermograms containing breast masses. The proposed project will help in early detection of cancer before it causes severe damage to the patient. It is also a gateway to automate future processes for detecting thermal anomalies and ultimately removing human intervention. For future scope, the project can be deployed commercially for all the infrared diagnostic centers across the country which can help to diagnose different medical conditions in early phases.

### 1.4.2 Limitations

The DITI method can show some false positives and false negatives. Therefore, it can only be used as a preliminary test. The equipment for capturing thermal images is costly. So availability across different regions is a concern. Since the software will be built using python, it may be slower than other potential candidates.

## 1.5 Methodologies of Problem Solving

### 1.5.1 Python 3.8 :

Python[2] is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming.

### 1.5.2 MongoDB :

MongoDB[3] is a source-available cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with optional schemas. MongoDB is developed by MongoDB Inc. and licensed under the Server Side Public License (SSPL).

# Chapter 2

# Literature Survey

This chapter tells us about the detailed study of the research papers studied for this project and the years of publication and the technology used.

In [4], Ola O. Soliman used the concept of symmetry for finding the probability of positive breast mass characteristics using a BPNN with input layer consisting of 32 neurons and a hidden layer consisting of 8 neurons followed by output layer consisting 1 neuron. This technique was tested on 63 IR single breast images(29 Healthy and 34 malignant) and had an accuracy of 96.51%.

In [5], Nader Abd El-Rahman Mohamed used the concept of image processing to find 32 features which is feed to BPNN with input layer consisting of 32 neurons and a hidden layer consisting of 15 neurons. The output layer consist of a 1 neuron. The technique was tested on 206 thermography images of the breast(187 normal and 19 abnormal patterns) and had an accuracy of 96.12%.

| Title | Author | Year | Survey | Remark |
|---|---|---|---|---|
| Automatic Breast Cancer Detection Using Digital Thermal Images | Ola O. Soliman, Nasser H. Sweilam, Doaa M. Shawky | 2018 | Use of thermal asymmetry property for detecting Breast Cancer. | Accuracy of 96.51%. |
| Breast Cancer risk detection Using Digital Infrared Thermal Images | Nader Abd El-Rahman Mohamed | 2015 | Use of classification technique for detecting Breast cancer. | Accuracy of 96.12%. |

Table 2.1: Literature Survey

# Chapter 3

# Software Requirements Specification

This chapters holds the Introductory part of the system which is to be implemented,its scope,characteristics,requirements etc.The detailed study of the model is been imbibed in the model.

## 3.1 Introduction

The software requirement specification document consistent of all necessary requirements required for project development. To develop the software system we should have clear understanding of Software system. To achieve this we need to continuous communication with customers to gather all requirements.

The purpose of this project includes, early prediction of Breast cancer risk to minimize the death rate using DITI and BPNN and automate the process of examining the thermographic images to find anomalies using image processing.

### 3.1.1 Project Scope

This project will consist of creating an automated process for finding anomalies and risks in thermal scans of patients for early detection of severe medical conditions, which aims to streamline the already existing process. The project includes finding thermal anomalies and predicting probability values of positive thermograms containing breast masses. The proposed project will help in early detection of cancer before it causes severe damage to the patient. It is also a gateway to automate future processes for detecting thermal anomalies and ultimately removing human intervention. For future scope, the project can be deployed commercially for all the infrared diagnostic centers

across the country which can help to diagnose different medical conditions in early phases.

### 3.1.2 User Classes and Characteristics

The user of the system would be the staff member of the organization. He/She will use this system to find anomalies in the thermogram. The user will have the following functions:

- Create a new patient
- Get information about an existing patient
- Generate Report about a patient
- Update patient's history

### 3.1.3 Assumptions and Dependencies

The following assumptions are made while developing this project,

- For Database Management the software should be running on a system that has MongoDB locally installed.
- For the software to run, the system should have Python 3 installed.
- Python libraries like opencv, pillow, numpy, etc are also required.

## 3.2 Functional Requirements

The primary requirement of the project is to automate the process of finding thermal anomaly in the thermographs and thus reducing the human intervention in the process, increasing the accuracy and speed of diagnostic process. The product will also provide a facility which will predict the possibility of positive breast masses.

### 3.2.1 System Feature 1 (Functional Requirement)

The proposed system will consist a GUI through which the organization employees can interact with the system and database, providing functionalities such as operating on an existing patient data as well as their thermal images, adding a new patient along with their thermal images to the database and retrieving patient data and reports from the database as per their needs.

### 3.2.2 System Feature 2 (Functional Requirement)

The proposed system will take the thermal images of the patients and then perform image pre-processing on it.This pre-processed image will be used to find thermal anomalies and probability of positive breast mass characteristics in the breast. After obtaining the results from the previous function, a report will be generated which will be stored in the database.There will also be a separate functionality provided for the radiologist to enter his/her remarks in the final report.

## 3.3 External Interface Requirements

### 3.3.1 User Interfaces

In this section we will discuss about how the user will interact with the system.

#### 3.3.1.1 Home Screen

The user interface will consist of a GUI through which the employee of the organization can interact with the system. The Main screen of the GUI will display 3 options, the options are as follows:

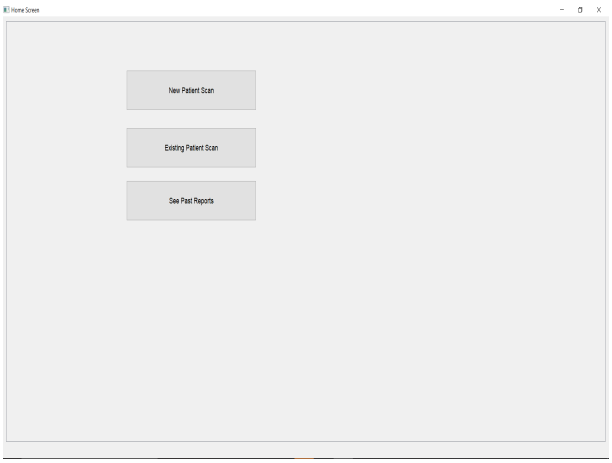



Figure 3.1: Home Screen

#### 3.3.1.2 New patient Screen

This option will display a new page which will allow the employee to enter information of new patient along with his/her thermographs and give an option to generate the related report.

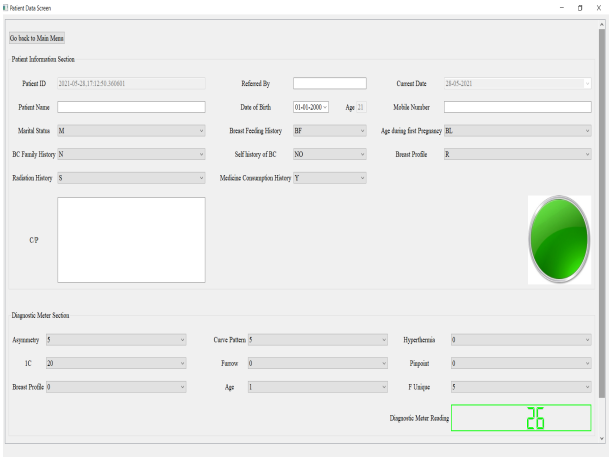



Figure 3.2: Patient Screen

#### 3.3.1.3 Existing patient Screen

This option will display a new page where the employee can search existing patient using patient id, name, mobile number. After finding the existing patient the employee can add new thermographs of the patient and generate new report.

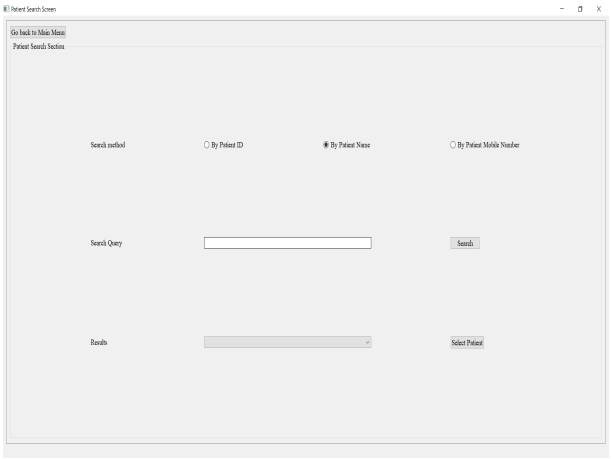



Figure 3.3: Existing patient Screen

#### 3.3.1.4 View Reports Screen

This option allows the employee to view the report of a particular patient from the database.

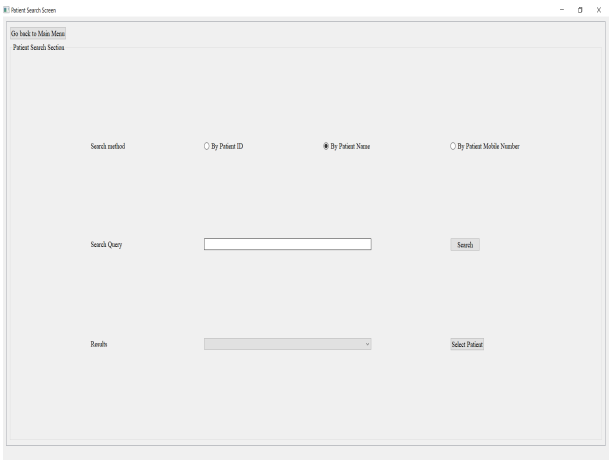



Figure 3.4: View Report Screen

### 3.3.2 Hardware Interfaces

The system consist of a dedicated computer which will contain the processing application and a thermal camera for capturing thermal images. The images captured using thermal camera are stored in the local memory of the camera. The employee can use the micro-SD card reader to transfer the images to the computer for further processing.

### 3.3.3 Software Interfaces

The system and the database will communicate using a library available to us. Here MogoDb database will communicate with the system using the PyMongo library. We need image processing libraries like opencv, PILLOW for image preporccesing. For GUI implementation, we need tkinter.

## 3.4 Nonfunctional Requirements

### 3.4.1 Performance Requirements

The product must perform all the specified requirements of the sponsoring organisation like finding thermal anomalies and also the given software should complete its execution in less than 15 minutes.

### 3.4.2 Safety Requirements

If there is extensive damage to a wide portion of the database due to catastrophic failure, such as a disk crash, the recovery method restores a past copy of the database that was backed up to archival storage (typically tape) and reconstructs a more current state by reapplying or redoing the operations of committed transactions from the backed up log, up to the time of failure.

### 3.4.3 Security Requirements

Security systems need database storage just like many other applications. However, the special requirements of the security market mean that vendors must choose their database partner carefully.

### 3.4.4 Software Quality Attributes

- Portability : Program should work on all computer systems given that execution environment and database requirements are satisfied.
- Maintainability : The Database system shall address the maintenance of data storage and administer data retrieval and removal.
- Availability : The patient shall be available at the instant of request without any hindrance.

## 3.5 System Requirements

### 3.5.1 Database Requirements

We are using MongoDB as our database. The database schema will mainly consist of two document,

- Patient Document: This document will store the information of the patient.

- Image Document: This will contain the thermographs of the patient.

### 3.5.2 Software Requirements(Platform Choice)

- OS - Windows 10
- Python 3.6 interpreter and associated libraries
- Associated Database management System.

### 3.5.3 Hardware Requirements

- CPU - i3 $6^{th}$ generation and above
- RAM - 4 GB
- Hard Disk - 500GB
- Printer for report generation

## 3.6 Analysis Models: SDLC Model to be applied

The development model used for this project is called Evolutionary Development Model.This model is a combination of Iterative and Incremental model of software development life cycle.

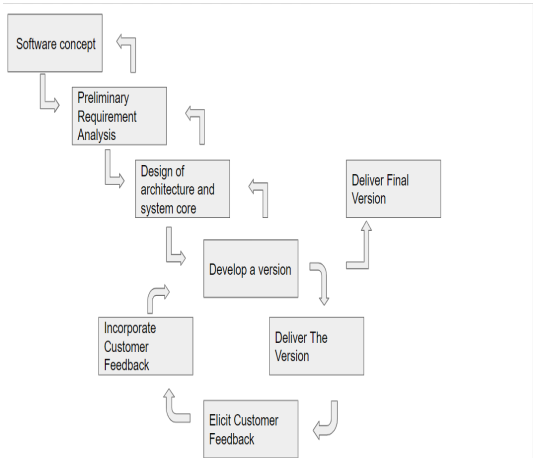



Figure 3.5: Evolutionary Development Model

The Evolutionary development model divides the development cycle into smaller, incremental waterfall models in which users are able to get access to the product at the end of each cycle. Feedback is provided by the users on the product for the planning

stage of the next cycle and the development team responds, often by changing the product, plan or process. Therefore, the software product evolves with time.

Evolutionary model suggests breaking down of work into smaller chunks, prioritizing them and then delivering those chunks to the customer one by one. The number of chunks is huge and is the number of deliveries made to the customer. The main advantage is that the customer's confidence increases as he constantly gets quantifiable goods or services from the beginning of the project to verify and validate his requirements. The model allows for changing requirements as well as all work in broken down into maintainable work chunks.

In evolutionary model, a user gets a chance to experiment partially developed system.It reduces the error because the core modules get tested thoroughly.

# Chapter 4

# System Design

In this chapter we will discuss about architecture of the system, different UML Diagrams and Entity Relationship Diagram.

## 4.1 System Architecture

The architecture of the proposed system is shown in Fig 4.1. From the architecture diagram we can clearly see how the different components of the system as well as the stakeholders interact with each other.

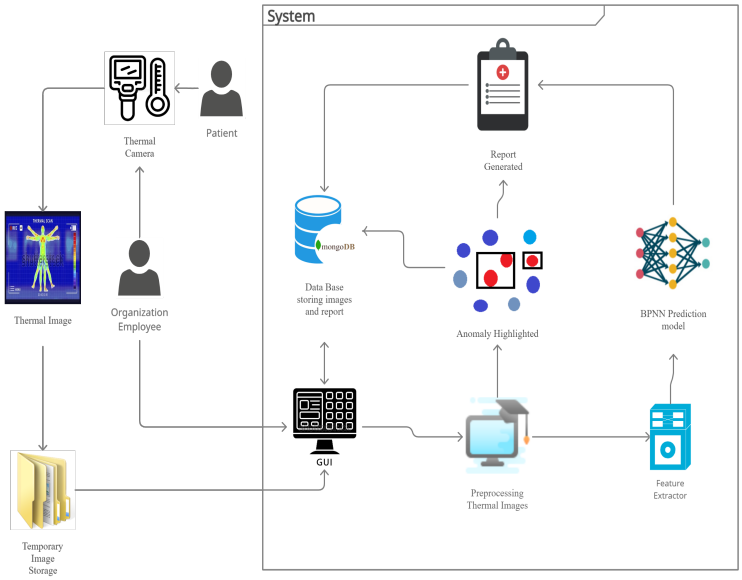



Figure 4.1: System Architecture Diagram

As per the architecture, the employee will interact with the system through GUI. With the help of this GUI, he/she can retrieve as well as store patient information,

thermographs and medical reports. The patient will interact with the thermal camera to generate thermal images which will be further diagnosed. The GUI will take thermal images from the temporary Image storage as an input and will send these images to the image preprocessing module. This module will remove the back-ground, extract the region of interest(ROI).This ROI is given as input to thermal anomaly module as well as to feature extraction module.The thermal anomaly module will highlight the temperature anomaly in the ROI. The feature extractor module will take the ROI as an input and extract the 32 Characteristic features from it. These features are used as an input for the BPNN which will generate a probability value of the positive breast mass being present in the breast. This value is used to define the risk of breast cancer which will be add to the report.The temperature anomaly obtained through thermal anomaly module and probability value generated by BPNN are included in the patient report which is stored in the database.

## 4.2 Data Flow Diagrams Level-1

DFD graphically representing the functions, or processes, which capture, manipulate, store, and distribute data between a system and its environment and between components of a system. The visual representation makes it a good communication tool between User and System designer.

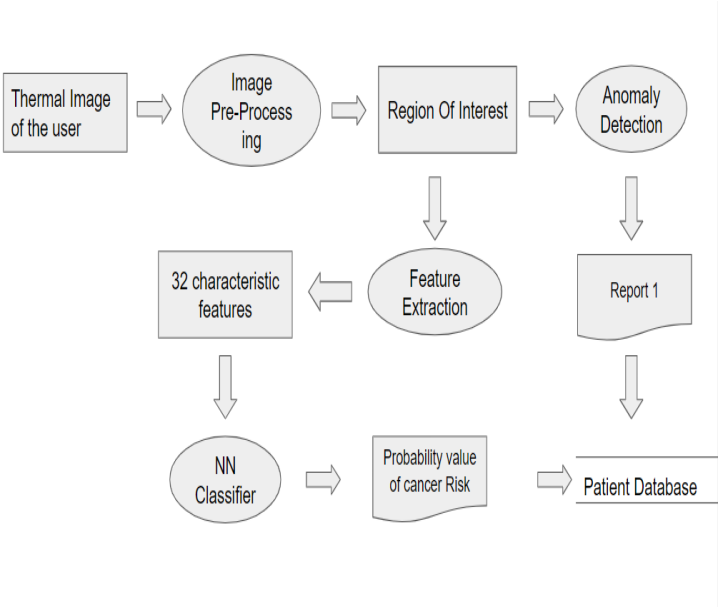



Figure 4.2: DFD Level-1

This diagram shows how the data is shared between the modules. As we can see from the given diagram Fig. 4.2, the thermal images are captured and then passed

to preprocessing class which then finds the ROI. This ROI is given as an input to two classes, Anomaly Detector and Feature Extraction. The anomaly detector finds the anomaly in the ROI and then adds this information to the report. The Feature Extraction class extracts 32 characteristic features from the ROI and sends it Neural Network classifier which then uses these 32 features for finding the probability value of cancer risk in the patient. This value along with the anomaly detected is then stored in the database which is then used to generate medical report.

## 4.3 Entity Relationship Diagrams

An entity relationship diagram (ERD) shows the relationships of entity sets stored in a database. An entity in this context is an object, a component of data. An entity set is a collection of similar entities. These entities can have attributes that define its properties. The ER-Diagram for our project as shown in Fig. 4.3.

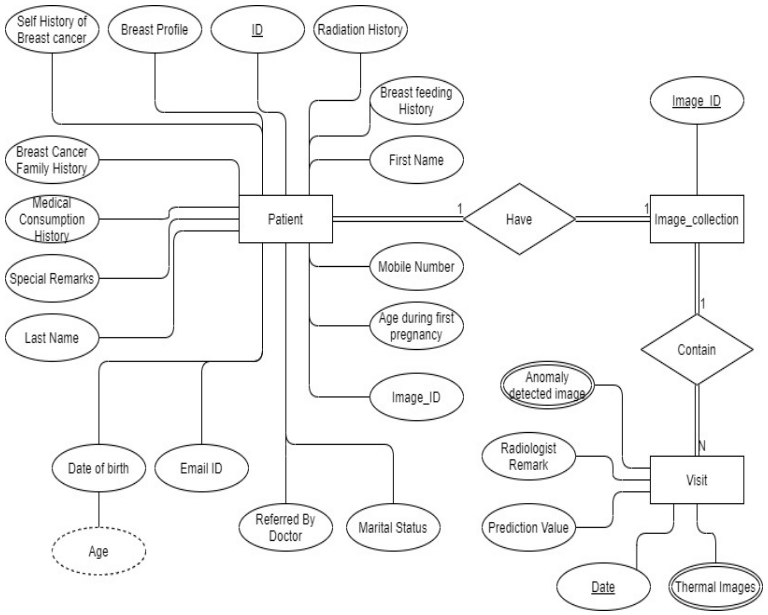



Figure 4.3: Entity Relationship Diagram

## 4.4 UML Diagrams

A UML diagram is a diagram based on the UML (Unified Modeling Language) with the purpose of visually representing a system along with its main actors, roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system.

### 4.4.1 Activity Diagram

An activity diagram is a behavioral diagram i.e. it depicts the behavior of a system. It portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed. We can depict both sequential processing and concurrent processing of activities using this diagram. In Fig 4.4, we have show the activity diagram for our project.

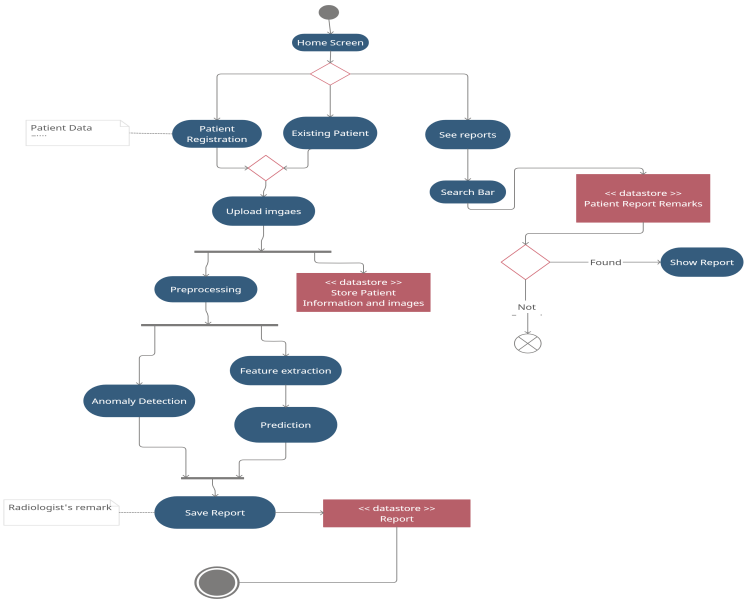



Figure 4.4: Activity Diagram

### 4.4.2 Class Diagram

A class diagram in the Unified Modeling Language is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations, and the relationships among objects.

Fig. 4.5 show the class diagram of our project. According to our class diagram there five classes which are stated as follows:

- GUI : This class will handle the user interface.
- Patient : This class will handle the images and patient information
- Preprocessing : This class will process the image and extract the ROI.
- Feature_Extraction : This class will take ROI as an input and calculate the 32 features of the ROI.

- Anomaly_Detection : This class will take the ROI and find the thermal anomaly.
- Prediction : This class will take the 32 features to predict breast cancer.

The interaction between the different classes is also shown in the Fig. 4.5.

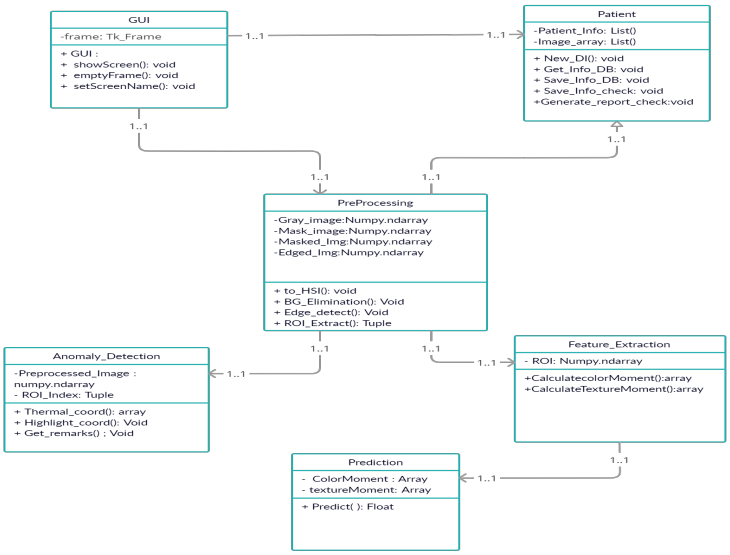



Figure 4.5: Class Diagram

### 4.4.3 Use Case Diagram

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved.

As per the Fig. 4.6, there are three actors namely, patient, employee and radiologist. The actors can do the following activities:

- The employee can register new patient, work on existing patient or view patient reports.
- The patient can see the reports which are generated by the system.
- The radiologist can enter his/her remarks on the patient medical condition and can view the patient reports generated


Figure 4.6: Use Case Diagram

### 4.4.4 Sequence Diagram

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario.

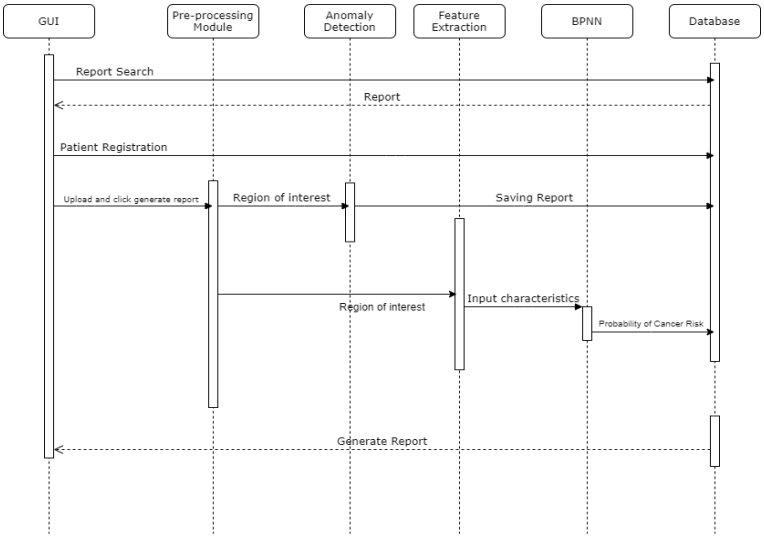



Figure 4.7: Sequence Diagram

# Chapter 5

# Project Plan

## 5.1 Project Estimate

### 5.1.1 Reconciled Estimates

**Approximate Time Estimate**

Effective working hours/week/person = 3

Total number of weeks worked on the project = 28

So, total number of hours H = 28*3 = 84 hours

Hence total number of man hours = 4*84 = 336 hours

As no hardware is used in this project, therefore the Hardware/equipment Cost = 0

Table 5.1: Project implementation plan

| SR.NO | Activity | Days Required |
|---|---|---|
| 1 | Project Initiation and study | 10 Days |
| 2 | Requirement Analysis | 15 Days |
| 3 | Installation of required software | 5 days |
| 4 | Implementation of Pre-processing | 10 days |
| 5 | Implementation of GUI | 10 days |
| 6 | Implementation of Neural Network and Anomaly Detection | 35 days |
| 7 | Generation of reports | 10 days |
| 8 | Testing of product | 5 days |
| 9 | Correction of errors | 10 days |

### 5.1.2 Project Resources

#### 5.1.2.1 Software Requirements(Platform Choice)

- OS - Windows 10
- Python 3.6 interpreter and associated libraries
- Associated Database management System.

#### 5.1.2.2 Hardware Requirements

- CPU - i3 $6^{th}$ generation and above
- RAM - 4 GB
- Hard Disk - 500GB
- Printer for report generation

### 5.1.3 Risk Management

### 5.1.4 Risk Identification

There are quite different types of risk analysis that can be used. Risk analysis is used to identify the high-risk elements of a project in software engineering. Also, it provides ways of detailing the impact of risk mitigation strategies. Risk analysis has also been found to be most important in the software design phase to evaluate the criticality of the system, where risks are analyzed and necessary countermeasures are introduced. The main purpose of risk analysis is to understand risks in better ways and to verify and correct attributes. Successful risk analysis includes important elements like problem definition, problem formulation, data collection. Risk analysis is useful in many situations:

- When you're planning projects, to help you anticipate and neutralize possible problems.
- When you're deciding whether or not to move forward with a project.
- When you're improving safety and managing potential risks in the workplace.
- When you're preparing for events such as equipment or technology failure, theft,staff.
- When you're planning for changes in your environment, such as new competitors.
- coming into the market, or changes to government policy.

### 5.1.5 Risk Analysis

It means identifying the particular hazards that their personnel may encounter in the course of their duties. It then describes the appropriate way to remove a particular risk and protect the system.

**Risk assessment:**

- Technical Risk: There is a possibility that a part of the database is lost due to hardware failure. To avoid we can use different data recovery techniques which are available with database management systems.
- Operational Risk: Even though the proposed algorithm have higher accuracy but it is still immune to failure like false positives and false negatives.

### 5.1.6 Overview of Risk Mitigation, Monitoring, Management

#### 5.1.6.1 Risk Mitigation

- Database failure : MongoDB provides a facility for data backup which can be used here.
- Human Error : The software provides backtracking which can be used here.

#### 5.1.6.2 Risk Monitoring

After risks are identified, analysed, and prioritized, and actions are established, it is essential that the team regularly monitor the progress of the system and the resolution of the risk items, taking corrective action when necessary. This monitoring can be done as part of the team project management activities or via explicit risk management activities. Often teams regularly monitor their "Top 10 risks". Risks need to be revisited at regular intervals for the team to re-evaluate each risk to determine when new circumstances caused its probability and/or impact to change. At each interval, some risks may be added to the list and others taken away. Risks need to be re-prioritized to see which are moved "above the line" and need to have action plans and which move"below the line" and no longer need action plans. A key to successful risk management is that proactive actions are owned by individuals and are monitored.

#### 5.1.6.3 Risk Management

The risk management process can be broken down into two interrelated phases, risk assessment and risk control. These phases are further broken down. Risk assessment

involves risk identification, risk analysis, and risk prioritization. Risk control involves risk planning, risk mitigation, and risk monitoring. It is essential that risk management can be done iteratively, throughout the project, as a part of the team's project management routine.

### 5.1.7 Project Schedule

#### 5.1.7.1 Project Task Set

A task is a single unit of work - an action to accomplish in a project, a single step in a multi-step project. A task is accomplished by a set deadline, and must contribute toward work-related objectives. Just as project management is the coordination of individual tasks, a task can be broken down further into subtasks, which should also have clear start and end dates for completion.
No single task set is appropriate for all projects and process models.
1. It varies depending on the project type and the degree of rigor (based on influential factors) with which the team plans to work The task set should provide enough discipline to achieve high software quality.
2. But it must not burden the project team with unnecessary work.

#### 5.1.7.2 Task Network

Figure 5.1 describes all the dependencies between different tasks.

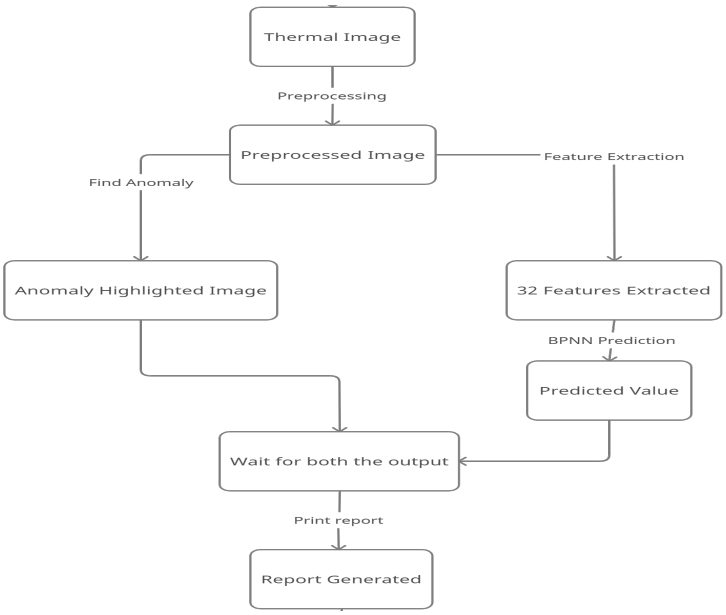



Figure 5.1: Task Network

#### 5.1.7.3 Timeline Chart

Table 5.2: Project implementation plan

| NO. | Activity | Month Started | Month Ended |
|---|---|---|---|
| 1 | Project Initialisation | October | October |
| 2 | Project Planning | October | October |
| 3 | Project Idea proposal and acceptance | November | November |
| 4 | Project Designing | November | November |
| 5 | Incorporated new changes to project design | December | December |
| 6 | Preparation of SRS | December | December |
| 7 | Modeling and Design of Project | January | January |
| 8 | Working on ROI Extraction module | January | February |
| 9 | Working on GUI Module | February | April |
| 10 | Working on Anomaly Detection module | February | May |
| 11 | Working on Neural Network Module | March | April |
| 12 | Testing | May | May |

### 5.1.8 Team Organization

#### 5.1.8.1 Team Structure

Proper project team organization is one of the key constraints to project success. If the project has no productive and well-organized team, there's an increased probability that this project will be failed at the very beginning because initially the team is unable to do the project in the right manner. Without right organization of teamwork, people who form the team will fail with performing a number of specific roles and carrying out a variety of group/individual responsibilities. Hence, when you plan for a new project, first you must take care of the best project team organization through team building activities. A Project Team is an organized group of people who are involved in performing shared/individual tasks of the project as well as achieving shared/individual

goals and objectives for the purpose of accomplishing the project and producing its results.

#### 5.1.8.2 Team Structure

All the work is divided into four members as shown below,

- Samyak Gangwal (Leader) : Implementation, Research Paper, Design Analysis, UML Diagram
- Nikhil Pawar (Member) : Implementation, Research Paper, Design Analysis, UML Diagram
- Yash Kewlani (Member) : Implementation, Research Paper, Design Analysis, UML Diagram
- Prasanna Dixit (Member) : Implementation, Research Paper, Design Analysis, UML Diagram

# Chapter 6

# Project Implementation

## 6.1 Overview of Project Modules

The proposed system consist of four modules containing :

1. Pre-Processing - This module is used to extract the the region of interest(ROI) from the thermal images.
2. Anomaly Detection - This module takes the ROI extracted from the previous module to find the thermal anomalies and highlights them.
3. Neural Network Classifier - This module is used to classify the given thermograms as positive or negative breast mass.
4. GUI - This module interacts with user of the organisation.

## 6.2 Tools and Technologies Used

- **Python**
  Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects[2]. For this project we are using python version 3.8.
- **OpenCV**
  OpenCV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real-time computer vision. Originally developed by Intel, it was later supported by Willow Garage then Itseez. The library is cross-platform and free for use under the open-source Apache 2 License[6].

OpenCV-Python is a library of Python bindings designed to solve computer vision problems.
It is one of the most important library for our project. Almost all the modules use opencv for manipulating the images.

- **TensorFlow**
  TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications[7].
  We are using TesnsorFlow and Keras for implementing neural network classifier in our project. This classifier will predict the possibility of positive or negative breast mass.

- **PyQt5**
  Qt is set of cross-platform C++ libraries that implement high-level APIs for accessing many aspects of modern desktop and mobile systems. These include location and positioning services, multimedia, NFC and Bluetooth connectivity, a Chromium based web browser, as well as traditional UI development. PyQt5 is a comprehensive set of Python bindings for Qt v5[8].
  This module is used for the implementation of the GUI for the proposed project.

- **Reportlab**
  An Open Source Python library for generating PDFs and graphics[9].
  This library is used for the creation of the report which is the final output of the project.

- **Scikit-image**
  scikit-image is an open-source image processing library for the Python programming language. It includes algorithms for segmentation, geometric transformations, color space manipulation, analysis, filtering, morphology, feature detection, and more. It is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy[10].
  This library is used for the implementation of edge detection and feature extraction in our project.

- **numpy**
  NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays[11].

The images which are imported using opencv are all handled as numpy array. This library was also used in all the modules.

- **MongoDB**
  MongoDB is a source-available cross-platform document-oriented database program. It is classified as a NoSQL database program. It uses JSON-like documents with optional schemas. MongoDB is developed by MongoDB Inc. and licensed under the Server Side Public License[2]. This is the database for the our proposed system. It will store patient information, thermal images and past reports.

- **PyMongo**
  The PyMongo distribution contains tools for interacting with MongoDB database from Python. The bson package is an implementation of the BSON format for Python. The pymongo package is a native Python driver for MongoDB. The gridfs package is a gridfs implementation on top of pymongo[12].
  This library is used for integrating MongoDB database within the application.

## 6.3 Algorithm Details

The algorithms used in this project are as follows:

### 6.3.1 Pre-Processing

- Step 1 : Import the image using OpenCV
- Step 2 : Display the input image to the user
- Step 3 : Ask the user to select ROI plotting points around the breast.
- Step 4 : Apply convex hull on the selected points to generate a ROI mask.
- Step 5 : Use the ROI mask to extract the region of interest.
- Step 6 : Select a pixel having normal temperature.

### 6.3.2 Anomaly Detection

- Step 1 : Import the ROI Image and the normal image
- Step 2 : Perform detail enhancement on the ROI
- Step 3 : Apply k-means algorithm with value of k=8

- step 4 : Perform Gray-scale conversion of the image based on the flag. If the value of flag is 0 use the weights as [0.1,0.6,0.1] else if the flag is set to 1 use the weights as [-0.2,0.6,-0.2] and if the flag is equal to 1 jump to step 9.
- step 5 : Take the normal point selected in Pre-processing module and create a kernel of size 3x3 and calculate the average value of gray pixels.
- step 6 : Compare all the pixels in the ROI image with the average value calculated in previous step with some predefined threshold applied to it.
- step 7 : If the difference in the value of pixel is greater than threshold add the pixel coordinates to the anomaly list.
- step 8 : If the length of the list is less than 1500 pixels, change the flag to 1 and go back to step 4.
- step 9 : Now traverse the list and monitor the red and blue channel of the pixel of the original image. If the value of red channel is greater than the value of blue channel consider this as hot anomaly and similarly for cold anomaly the value for blue channel will be greater than red channel. Highlight the anomalies on the original image.
- step 10 : Save this image in the database.

### 6.3.3 Neural Network classifier

- Step 1 : Import the ROI image
- Step 2 : Split the 3 channels of the image as blue, green, red
- Step 3 : Ignore the non ROI pixels
- Step 4 : For each channel find mean, standard deviation, skewness and kurtosis. Store these values in a list.
- Step 5 : Now convert the BGR image to gray scale image.
- Step 6 : Construct a Gray level co-occurrence matrix using skimage library.
- Step 7 : Use GLCM to find the values of energy, entropy, contrast, homogeneity, correlation in four directions i.e Horizontal, Vertical and along the two Diagonals. Add these values in the list as well.
- Step 8 : Once the list is created, feed this list to the trained neural network classifier.

- Step 9 : The output of the classifier is added to the database.

The NN classifier is created using TensorFlow and Keras library. The classifier consist of a single hidden layer with input and output layer. The input layer consist of 32 neurons and the output layer consist of a single neuron. There are two hidden layers in this NN model containing 16 and 64 neurons each. The Error Functions used are Rectified Linear Unit(ReLU), tanh and Sigmoid.

# Chapter 7

# Software Testing

## 7.1 Type of Testing

### 7.1.1 Unit Testing

Unit testing is a level of software testing where individual units/ components of the software are tested. The purpose is to validate that each unit of software performs is designed. A unit is the smallest testable part of any software. It usually has one or a few inputs and usually a single output. In procedural programming, a unit may be an individual program, function, procedure, etc. In object-oriented programming, the smallest unit is a method, which may belong to a base/ super class, abstract class or derived/ child class.

### 7.1.2 Integration testing

Integration testing is the phase in software testing in which individual software modules are combined and tested as a group. It occurs after unit testing and before validation testing. Integration testing takes as its input modules have been unit tested, groups them in larger aggregates, applies test defined in an integration test plans to those aggregates and delivers as its output the integrated system ready for system testing. Integration testing can be done in two ways:

- **Bottom-up testing :** This is an approach to integration testing where the lowest levels components are tested first, then used to facilitate the testing of higher-level components. The process is repeated until the component at the top of the hierarchy is tested.
- **Top-down testing :** This is an approach to integrated testing where the top integrated modules are tested and the branch of the module is tested step by step

until the end of the related module.

### 7.1.3 System Testing

System testing is a level of software testing where a complete and integrated software is tested. The purpose of this test is to evaluate the system's compliance with the specified requirements. System testing is defined as the testing conducted on the complete integrated system to evaluate the working of the system as a whole. System testing is carried out in the context of functional requirements and system requirements. Type of system testing:

1. Usability Testing – Usability Testing mainly focuses on the user's ease to use the application, flexibility in handling controls, and ability of the system to meet its objectives.

2. Load Testing – Load Testing is necessary to know that a software solution will perform under real-life loads.

3. Regression Testing – Regression Testing involves testing done to make sure none of the changes made over the course of the development process have caused new bugs. It also makes sure no old bugs appear from the addition of new software modules over time.

4. Recovery Testing – Recovery testing is done to demonstrate a software solution is reliable, trustworthy, and can successfully recoup from possible crashes.

5. Functional Testing – Also known as functional completeness testing, Functional Testing involves trying to think of any possible missing functions.

## 7.2 Test cases and Test Results

### 7.2.1 Testing on Anomaly Detection

For testing anomaly detection we are using thermal images provided to us by the organisation. These images have both the hypo and hyper thermal images. The assumption made while testing is that the user will select right pixel as normal temperature.

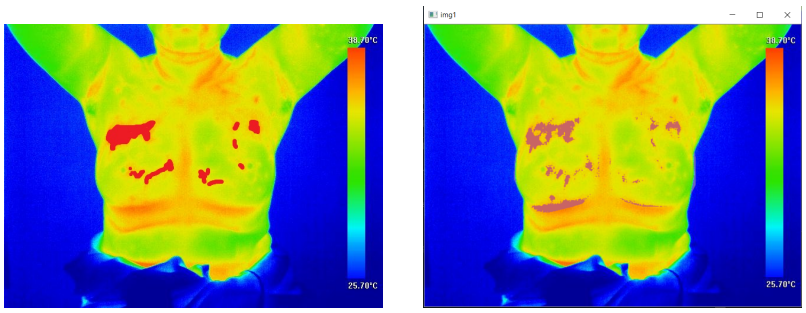
**Test Case ID** : 001

**Test Scenario** : Select a normal temperature point from the image and check if the anomaly is highlighted properly(For hot regions).

**Action** : Select normal temperature pixel.

**Input** : Thermal Image.

**Status** : Pass



(a) Expected (b) Actual

Figure 7.1: Anomaly Detection Testing

In the Figure 7.1, left figure shows the expected output of the anomaly detection module. The figure on the right i.e. Figure 7.1.b shows the actual output of the module. Here we can see that the module is clearly detecting hot patches.

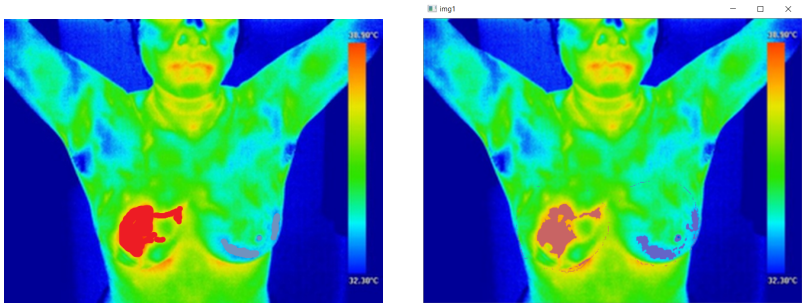
**Test Case ID** : 002

**Test Scenario** : Select a normal temperature point from the image and check if the anomaly is highlighted properly. Check for both thermal hot and cold spots in the image.

**Action** : Select normal temperature pixel.

**Input** : Thermal Image.

**Status** : Pass



(a) Expected (b) Actual

Figure 7.2: Anomaly Detection Testing

In the Figure 7.2, left figure shows the expected output of the anomaly detection module. The figure on the right i.e. Figure 7.1.b shows the actual output of the module. Here we can see that both type of anomalies are highlighted.

### 7.2.2 Testing on Neural Network Classifier

The data-set used for this project contains a total of 107 front view images[13] containing both healthy and sick patients. Out of these 107 images, 88 images were used for training the neural network classifier and the rest of 22 images were used for testing. The NN classifier showed an accuracy of 95.45%. Following are the examples of the tests carried out.

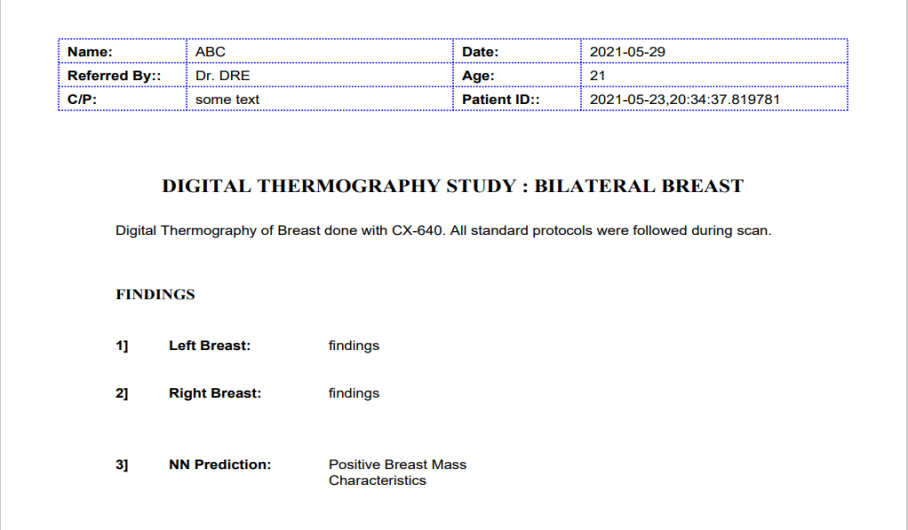
**Test Case ID** : 001

**Test Scenario** : Use the Neural Network classifier to predict the output for a sick image.

**Expected Output** : Positive

**Input** : Thermal Image.

**Status** : Pass



| | | | |
|---|---|---|---|
| **Name:** | ABC | **Date:** | 2021-05-29 |
| **Referred By::** | Dr. DRE | **Age:** | 21 |
| **C/P:** | some text | **Patient ID::** | 2021-05-23,20:34:37.819781 |

**DIGITAL THERMOGRAPHY STUDY : BILATERAL BREAST**

Digital Thermography of Breast done with CX-640. All standard protocols were followed during scan.

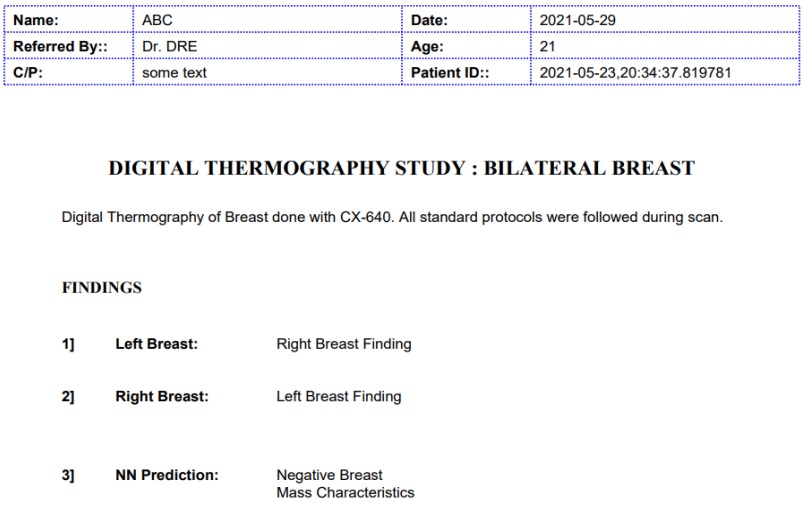
**FINDINGS**

| | | |
|---|---|---|
| **1]** | **Left Breast:** | findings |
| **2]** | **Right Breast:** | findings |
| **3]** | **NN Prediction:** | Positive Breast Mass Characteristics |

Figure 7.3: Neural Network Classifier Output in Report

In Figure 7.3 we can see that the output of the neural network classifier is positive which means that the given image contains positive breast masses which is the expected result.

**Test Case ID** : 002
**Test Scenario** : Use the Neural Network classifier to predict the output for a healthy image.
**Expected Output** : Negative
**Input** : Thermal Image.
**Status** : Pass



| | | | |
|---|---|---|---|
| **Name:** | ABC | **Date:** | 2021-05-29 |
| **Referred By::** | Dr. DRE | **Age:** | 21 |
| **C/P:** | some text | **Patient ID::** | 2021-05-23,20:34:37.819781 |

**DIGITAL THERMOGRAPHY STUDY : BILATERAL BREAST**

Digital Thermography of Breast done with CX-640. All standard protocols were followed during scan.

**FINDINGS**

| | | |
|---|---|---|
| **1]** | **Left Breast:** | Right Breast Finding |
| **2]** | **Right Breast:** | Left Breast Finding |
| **3]** | **NN Prediction:** | Negative Breast<br>Mass Characteristics |

Figure 7.4: Neural Network Classifier Output in Report

In Figure 7.3 we can see that the output of the neural network classifier is negative which means that the given image does not contain positive breast masses which is the expected result.

# Chapter 8

# Results

## 8.1 Outcomes

The output of the proposed system consist of a report. The report format is designed as per the requirement of the organisation. The report will contain the basic information of the patient followed by the findings for left breast and right breast. After findings report will display the output of neural network classifier. This report will also contain some suggestions given by the radiologist. The following report will contain anomaly highlighted images for all the views [Front view, right lateral, left lateral, right oblique, left oblique]. Followed by edge detected images. In the end the report will contain the history meter and diagnostic meter readings.

| | | | |
|---|---|---|---|
| **Name:** | Firstname middlename lastname | **Date:** | 2021-05-23 |
| **Referred By::** | Dr. Firstname Middlename Lastname | **Age:** | 21 |
| **C/P:** | This text contains information about the special remarks | **Patient ID::** | 2021-05-23,18:53:43.125389 |

**DIGITAL THERMOGRAPHY STUDY : BILATERAL BREAST**

Digital Thermography of Breast done with CX-640. All standard protocols were followed during scan.

**FINDINGS**

**1]** **Left Breast:** Finding 1
Finding 2
Finding 3

**2]** **Right Breast:** Finding 1
Finding 2
Finding 3

**3]** **NN Prediction:** Negative Breast
Mass Characteristics

**Suggest:** Finding 1
Finding 2
Finding 3

Thanks

*Many thanks for reference. Imaging findings has its own limitations and needs to be correlated clinically.*

Figure 8.1: Report Page 1

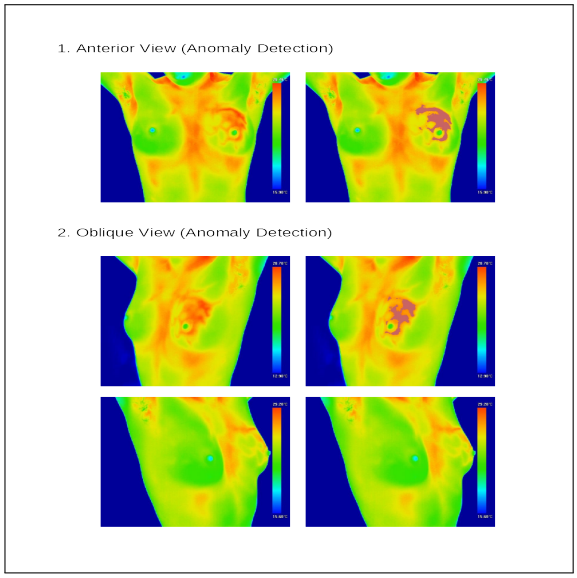



Figure 8.2: Report Page 2

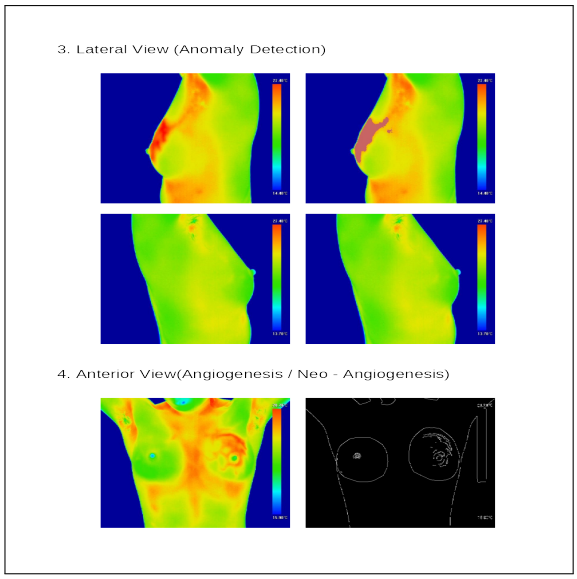



Figure 8.3: Report Page 3

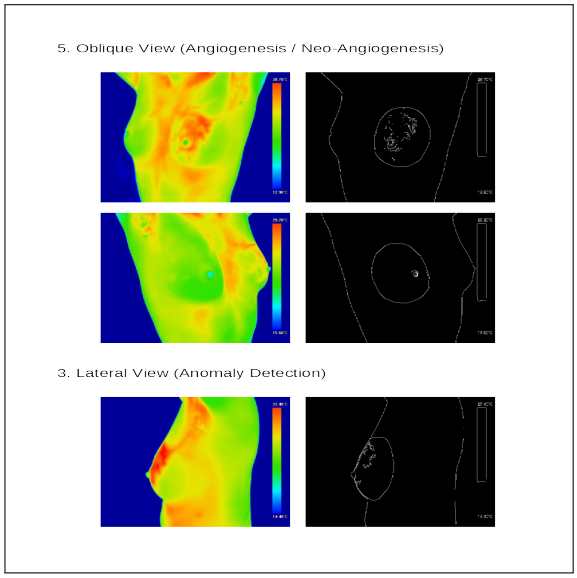



Figure 8.4: Report Page 4

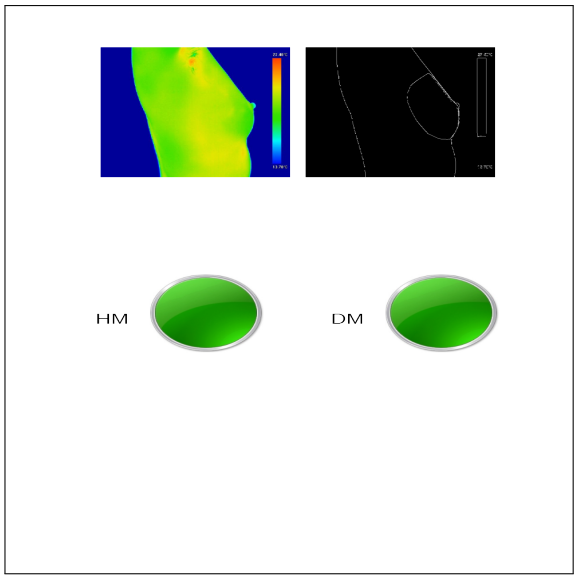



Figure 8.5: Report Page 5

## 8.2 Screen Shots

In this section we will see some working images of the software.

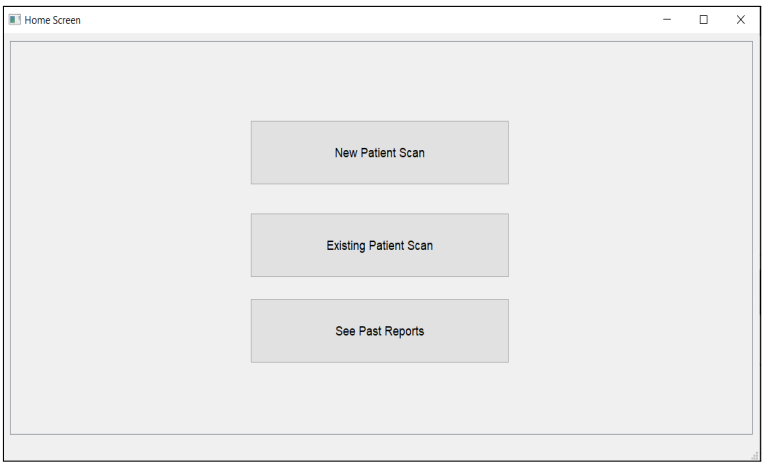



Figure 8.6: Home Screen

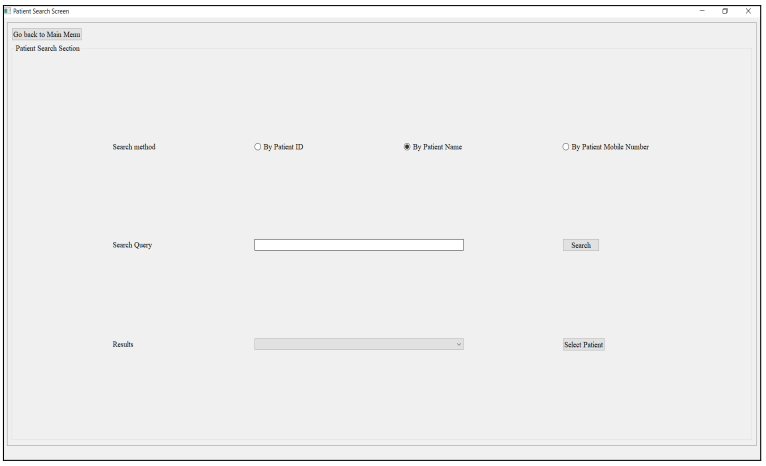



Figure 8.7: Existing Patient Screen

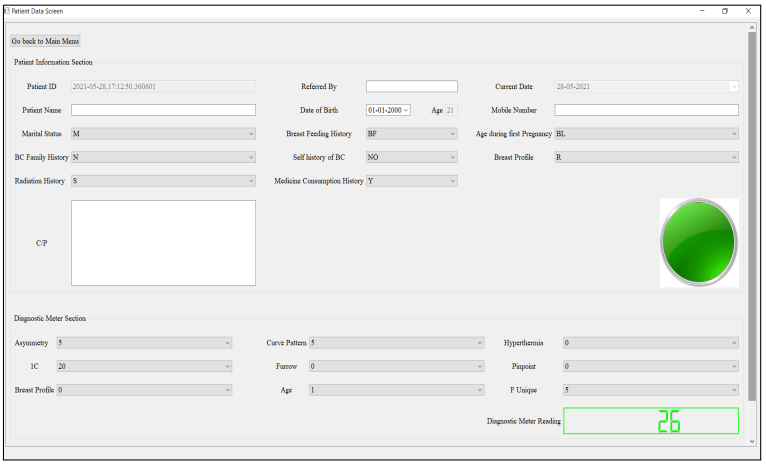



Figure 8.8: New Patient Screen

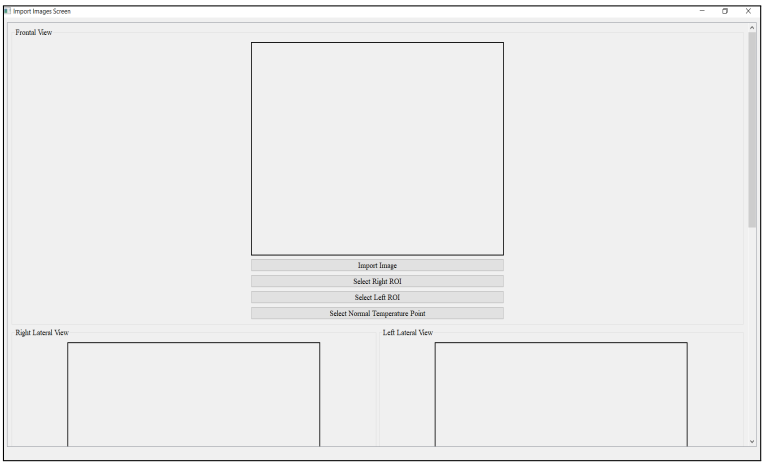



Figure 8.9: Import Image Screen

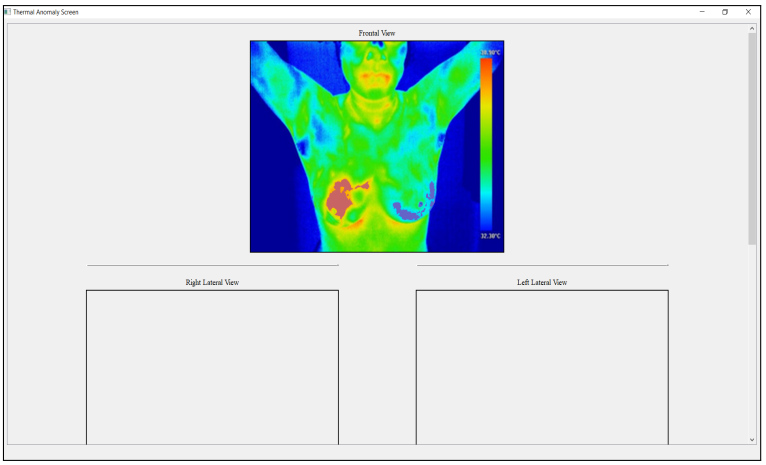



Figure 8.10: Thermal Anomaly Screen

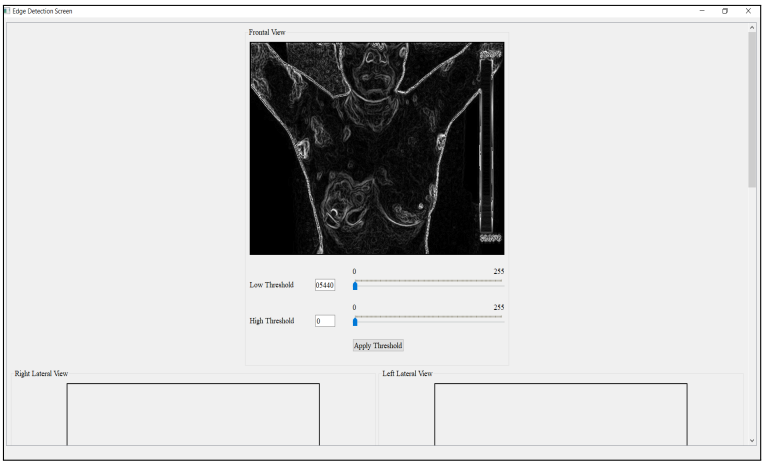



Figure 8.11: Edge Detection Screen

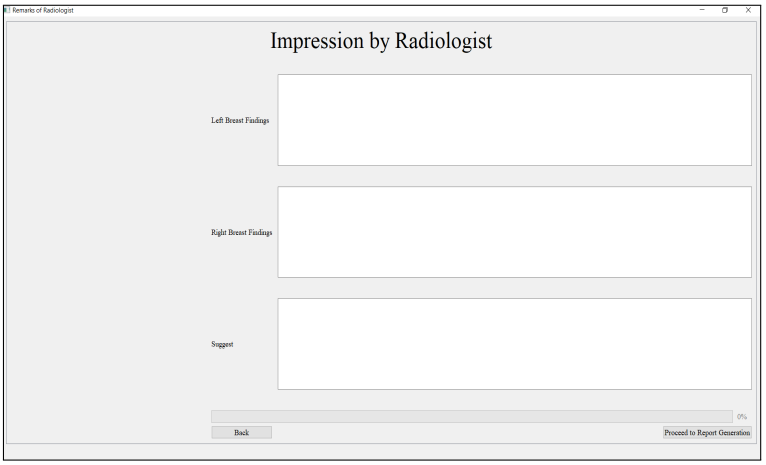



Figure 8.12: Remarks Entry Screen

# Chapter 9

# Conclusions

## 9.1 Conclusion

Hence, an approach for the early detection of breast cancer using thermal images which will detect thermal anomalies and extract a novel set of useful features to classify breast masses using back propagation neural network is proposed.

The proposed system displayed an accuracy of 95.45% on the testing data[13]. therefore this allows us to conclude that we have obtained a series of promising results with some room for improvement thus having a potential to obtain a breakthrough in cancer study.

## 9.2 Future Work

- The anomaly detection in this project is performing well. However there is still some room for improvement.
- The size of the data-set which was used for training neural network classifier was inadequate to give expected accuracy. So with a better and larger data-set the performance of neural network classifier can be improved.

## 9.3 Application

1. DITI is a non invasive method for detecting Breast cancer.
2. The proposed method is time efficient.
3. The proposed project idea automates the process of diagnosing medical conditions thus reducing human intervention.

4. The algorithm proposed for this project has an accuracy greater than 90

5. Use of Back-Propagation as training model makes it execute fast and easy to program.

# Appendix A

**Mathematical model**

- Let S be the system.
- S= {I,O,F,Success Failure }
  where,
- I={Input}
- I={Thermal Images}
- F={Function}
- F={Pre-processing,Anomaly Detection, NN Prediction}
- O={Output}
- O={Report, Anomaly detected images, Risk Prediction}
- Success- All functions executed successfully.
- Failure- Anomaly undetected, Wrong NN prediction, Database failure
- Problem is NP Hard

.

# Appendix B

We have submitted our paper titled as **"Thermal Anomaly detection and Breast Cancer Risk Prediction: Using DITI and Deep Learning Approach"** to **International Journal of Engineering Research and Applications**.

Journal Details :
International Journal of Engineering Research and Applications (IJERA) is an open access online peer reviewed international journal that publishes research and review articles in the fields of Computer Science, Neural Networks, Electrical Engineering, Software Engineering, Information Technology, Mechanical Engineering, Chemical Engineering, Plastic Engineering, Food Technology, Textile Engineering, Nano Technology science, Power Electronics, Electronics  Communication Engineering, Computational mathematics, Image processing, Civil Engineering, Structural Engineering, Environmental Engineering, VLSI Testing  Low Power VLSI Design etc.

# Appendix C

## .1 Plagiarism Report

| Title | Similarity Score |
|---|---|
| Chapter 3 | 17% |
| Chapter 1 | 50% |
| Abstract | 38% |

| Title | Similarity Score |
|---|---|
| Chapter 6 | 0% |
| Chapter 5 | 4% |
| Chapter 4 | 35% |

# Bibliography


[1] www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics, 9th Jan, 2018.

[2] https://www.python.org/

[3] https://www.mongodb.com/

[4] Ola O. Soliman, "Automatic Breast Cancer Detection Using Digital Thermal Images," 9th Cairo International Biomedical Engineering Conference (CIBEC), 2018.

[5] N. A. Mohamed, "Breast cancer risk detection using digital infrared thermal images," international journal of bioinformatics and biomedical engineering, 2015.

[6] https://docs.opencv.org/

[7] https://www.tensorflow.org/

[8] https://pypi.org/project/PyQt5/

[9] https://pypi.org/project/reportlab/

[10] https://pypi.org/project/scikit-image/

[11] https://pypi.org/project/numpy/

[12] https://pypi.org/project/pymongo/

[13] http://visual.ic.uff.br/dmi/