SDGP 2020 - G-Force - Report Submission

by Sothiingam Santosh

Submission date: 27-Jan-2020 06:37AM (UTC+0000)
Submission ID: 119149908
File name: SDGP_2020_-_G-Force_-_Report_Submission.pdf (2.54M)
Word count: 21363
Character count: 131376

UNIVERSITY OF WESTMINSTER

Informatics Institute of Technology

Department of Computing

Module: 5COSC009C

Software Development Group Project

Project Proposal

"PREDICTING THE POSSIBILITY OF GETTING BREAST CANCER THROUGH ANALYZING USER'S DAILY LIFESTYLE"

Team Name: G-FORCE

Team Members:

Full name	UoW Number
S Ziyam Santhosh	w1742090
Senura Nisal Weerasinghe	w1742092
Gayan Sasanka Dissanayake	w1742065
G. D. Dinali Uththama Kumarasiri	w1742096
A. S. M. Fernando	w1742066
Mohamed Boosary Fathima Farheen	w1742063

Contents

List of Figures

List of Tables

Chapter 01: Introduction

1.1 Problem Background

Breast Cancer is the most common cancer among women and one out of every eight women is found to be a victim of it. In the year 2018 alone there had been more than 620,000 deaths due to breast cancer which is 15% of all cancer deaths reported among women. There is a tendency in increasing rates in the number of diagnoses over the past years and over 2.1 million women are diagnosed every year. (World Health Organization, 2019) Breast Cancer is not limited to a particular race, age, region or not even gender, it is a global issue. The survival rate after being diagnosed as a breast cancer victim is less than 90% in the next 5 years and around 80% in the next 10 years. (American Society of Clinical Oncology, 2019) However, if you can identify Breast Cancer in very early stages there is a high possibility to completely cure Cancer. In the early stages of Breast Cancer, patients can use fewer radical treatments which cause less side effects as well. (Taylor, 2014)

1.1.1 Breast cancer

Cancer cells are usually formed from a lump or mass called a tumour and it is named after the body part which it is originated. Breast cancer is found in the breast tissue made up of glands for milk production. Breast cancer can be detected from mammogram and breast lumps. Sometimes this is in the very early stages where there is a high chance in curing. Screening is important when it comes to identifying breast cancer as the tumour could be really small. A small painless lump is commonly found as a physical signature. (American Cancer Society, 2017-2018)

1.1.2 Breast cancer and lifestyle

Breast depends on a lot of lifestyle factors. Some of them are in our control as well as some of them are not. Growing older, Genetic mutations, reproduction history, previous treatment using radiation therapy and family history are some of the factors that we cannot change that affects to increase the possibility of being a victim of Breast Cancer. Not being physically active, taking hormones, consumption of alcohol, smoking, bad food habits and reproduction history are some of the factors that we can control that has a huge impact on increasing the possibility of conceiving Breast Cancer. (Division of Cancer Prevention and Control, 2018)

1.2 Related work

1.2.1 Gail model

There have been attempts made by researchers in order to predict the cancer possibility before the person actually gets infected by breast cancer. The Gail model is the widely known tool which is implemented in order to predict the possibility of women being a victim of breast cancer. (Anon., 2019) The creator of this model Dr Mitchell Gail has used personal details of women such as the age, when was the first menstruation was present, when did that woman first deliver a baby and relatives who have

been diagnosed with breast cancer by extracting all the above mention details and some extra data from the user the Gail model will first predict the possibility of the user being diagnosed within the first five years as an average and then it will give the overall prediction of the user getting breast cancer throughout their lifespan (Wacholder, 2010). To train the model the creator has used the data of 280 000 women aged 35-70, but he has only taken white women into consideration. So the success rate of using this model in Asian countries stands at a lower rate. Since this model has been there for more than two decades it has gone through various developments, so there is an extensive validation in this model. But the problem with this model is as it has mentioned above it can be only used with western people, a study was conducted by a particular group of researchers in Qatar but the results was not successful like it is with white people because the data which was used to train the model was very much different from the answers given from the Arabian Gulf women (Erbil, 2015) so as for the researches done with the Gail model it is clear that it is highly recommended to use only with western people with similar lifestyles.

1.2.2 Breast Cancer Surveillance Consortium (BCSC)

Breast Cancer Surveillance Consortium (BCSC) Risk Calculator is another related work in this field. It is developed and validated among 1.1 million women undergoing mammography in the United States. It is an interactive tool designed by scientists in order to calculate a women's five years' and ten years' risk of being diagnosed as a Breast Cancer patient. This model is based on factors such as Age, Race/ethnicity, Family history (if a first-degree relative is a past Breast Cancer patient), History of breast biopsy (removing tissues), BI-RADS breast density (radiologic assessment of the density of breast tissue). (Breast Cancer Surveillance Consortium, 2016)

Constraints

- It is only applicable for women who are aged 35 to 75.
- Cannot calculate the risk of the previous diagnosis of breast cancer.
- User should have not undergone mastectomy surgery. (removing all breast tissues in order to prevent breast cancer)

1.3 Problem Definition

Breast cancer is not a disease that is only limited to a particular country or a location so it has started to gain the attention of the people all around the world because of the increased death rates reported in the recent years. (Tao, 2015) Most of the people think that breast cancer will only be diagnosed in females but they are unaware that even the males have a possibility of being a victim of this disease. All most every technology that is there in the market which is related to cancer doesn't actually predict the possibility, It only diagnoses whether the patient is a cancer victim. The most common way of doing this procedure is by getting cells from breast and through examining the cells the doctor will come to a conclusion on whether he or she is positive of having breast cancer, other technologies also provide

similar outcomes (Anon., 2019). Cancer is considered as a human tragedy and one of the most prevalent diseases in the wide, and its mortality resulting from cancer is being increased. It seems necessary to identify new strategies to prevent and treat such a deadly disease. (Elham Safarzadeh, 2014) Since cancer is considered as a deadly disease unless otherwise, you identify the sickness at a very early stage the victim is most likely to lose his or her life. Even after identifying cancer at an early stage the patient will have to go through a series of medication and treatments. These treatments and especially the medication will eventually lead the patient into various negative side effects like Hair loss, Anemia, Edema, infection and neutropenia (Anon., 2019), (Ewertz, 2011). So rather than identifying whether a specific person is suffering from breast cancer, it is better to prevent that person from being sick in the first place.

1.4 Research Motivation

Breast cancer is the most diagnosed cancer among women (24.2%, i.e. about one in 4 of all new cancer cases diagnosed in women worldwide), (cancer, 2018 September 12). According to that, the breast is the second leading cause of cancer death in women worldwide and diagnosed nearly in one out of eight women. (Muhammad Sufyian Bin Mohd Azmi, Zaihisma Che Cob, 2018 December 13). Even though the diagnosis rate is high, the survival rate of breast cancer is also increased up to 40% worldwide. (Dib, 29th April 2019). Although there are many implementations done regarding predicting breast cancer, they were all done on top of analyzing clinical data, and cancer cell images of patients. Other than these, a patient's lifestyle also plays a major role in getting breast cancer (Max Dieterich, Johannes Stubert, Toralf Reimer, Nicole Erickson, and Anika Berlingb, 2014 Nov 25). Since identifying breast cancer is a major task in the medical industry, we decided to analyze the risk factors of a user's daily lifestyle to provide a possibility prediction regarding breast cancer.

1.5 Research Question

RQ1: How daily lifestyle factors influence the breast cancer incidence possibility of a woman?

RQ2: How to analyze a user's lifestyle data using machine learning algorithms?

RQ3: What kind of data we need from the user's lifestyle to predict the possibility?

1.6 Research Aim

This research project is focused on designing and developing a predicting system to evaluate and predict the possibilities of breast cancer in women in a certain age range, considering their environmental and lifestyle factors.

For further clarification of the aim, the project will produce a predicting and recommendation system that predicts the breast cancer possibility in a woman at an early stage supporting the early diagnosis to facilitate the treatment process. This will be achieved by storing breast cancer victim lifestyle information in a database along with information of other women who have not experienced breast

cancer in their lifetime and comparing the similarities of their lifestyles to produce an accurate percentage rate of breast cancer incidence. In addition, the system will provide suggestions depending on the person's risk factor and lifestyle factors to facilitate the reduction of breast cancer risk in the individual.

The knowledge required will be researched, a machine learning model will be developed, and the accuracy of the system will be evaluated in order to validate the final result produced by the system. The system will be deployed as a cross-platform (android and IOS) mobile application available for public use. The system database will be hosted in a cloud.

1.7 Project Scope

Considering the objectives and research related to the project, the scope of the project is defined as follows. The project intends on supporting the early identification of breast cancer to increase survival rates. The main focus is given to increase the accuracy of the system by including the most effective factors.

1.7.1 In-Scope

The scope that will be focused on the project is as follows:

- The women in the age range above 25 years are considered - The possibility of breast cancer incidence in women younger than 25 years is considerably low compared to the women in the age range above 25 years.
- Lifestyle factors that have a considerable effect on breast cancer are considered –In this project the lifestyle factors that have a considerable effect on breast cancer risks such as obesity, alcohol consumption and smoking and many more are taken into consideration.
- User-friendly GUI to facilitate requirement– A user-friendly GUI with essential features which will have the functionalities which support data collection, prediction and recommendation of the system is used in the project.

1.7.2 Out-Scope

The factors that are excluded from the project are as follows:

- Prediction of breast cancer risk in males is not included in the project - Due to the lower risk of breast cancer in males, which is lower than 1% and the higher survival rate that lies between 75% - 100%, the prediction of breast cancer rates in males is offered as a future enhancement.
- Women living in foreign countries are not considered for the project – the Only lifestyle of Sri Lankan women is considered for this project as lifestyles vary from different races and countries.

1.8 Research Objectives

1.8.1 Project objectives

Objective 1	**To research into public data regarding breast cancer.**
Description	Carry out deep research regarding breast cancer victims in Sri Lanka and gather information about their lifestyle.
Objective 2	Produce a project plan by selecting suitable research methodologies.
Description	Analyze and select the most suitable research methods and software development methodology to build the system in an efficient way.
Objective 3	Identify the requirements.
Description	Gather all the functional and non-functional requirements.
Objective 4	Identify tools and algorithm.
Description	Analyze and evaluate the algorithms and select suitable technology tools to implement the system.
Objective 5	Identify architectural design requirements.
Description	Evaluate and come up with a suitable design specification for the system
Objective 6	Implement the system.
Description	Select appropriate programming methodologies and implement the breast cancer possibility predicting system. Proper development practices (training the model) should be followed to develop the system.
Objective 7	Test and evaluate the system.
Description	Evaluate the system using proper test cases to make sure the system meets all the requirements and fulfils the sole purpose with higher accuracy.

Table 1: Project objectives

1.8.2 Research objectives

Objective 1	**To research factors in breast cancer prediction system**
Description	To carry out the specific factors hidden in the development
Objective 2	To research into methods
Description	How the examining techniques that need to use the research step should follow
Objective 3	To examine the accuracy of the predictor
Description	To come up with the solution to increasing the accuracy in the breast cancer predictor.
Objective 4	To examine machine learning models
Description	carry out in-depth research into relevant machine learning models to make a prediction in the breast cancer predictor.

Table 2: Research Objectives

1.9 Resource Requirement

1.9.1 Software Requirements

- Operating system – The application will be created on top of both Android and iOS operating system. So, the mobile phone which is going to be used by the user should run either one of these. The application requires the minimum Android OS version of 7.0 and an iOS version of 10.0
- Weka / Tenserflow – These are the tools which will be used to implement algorithms, analyze the dataset, training and testing the model to find the possibility rate
- Python – This is going to be the base language to build the model needed for our application. Because python is a language which has a lot of libraries which is beneficial for further more processes.
- React Native / Flutter – The application will be developed in either one of the cross platforms to create a UI which will be interacting with the user in our application to gather data.
- Google drive – To keep backup files related to the project
- Github – This will be used to work in groups even though the members are separated. This will be the easiest way to keep backup of our project-related source codes too.
- MS Office – To create project-related documents like reports, and documentation.

- Adobe Photoshop –To create interactive icons and wallpapers which are must needed components for the user interface.

1.9.2 Hardware Requirements

- A Smartphone – The smartphone should have at least 2GB of RAM to run the application.
- Android - requires Android 7.0 or later. Compatible with Android phones and tablets.
- iOS - requires iOS 10.0 or later. Compatible with iPhone, iPad, and iPod touch.

1.9.3 Data Requirements

- Breast cancer victim dataset – from Census and Statistics Department of Sri Lanka
- General public census dataset – from Census and Statistics Department of Sri Lanka

Chapter 02: Literature Review

2.1 Literature review of the domain

2.1.1 Introduction for breast cancer

Health Care is a field where many researchers are looking forward to doing their researches to serve the patients, doctors and clinicians time and make their life easier. For the past few years, many types of research have done in this health field, for example, surveys have been done in disease section like diabetes, eye checking, chronical disease, etc. Apart from above Cancer which is a deadly disease that has been caught many researcher's eyes mainly in the past few years. Cancer is the second most cause of death globally and it is responsible for 9.6 million deaths for the year 2018 all over the world. Nearly 70% of deaths from cancer happen in low- and middle-income countries (World-Health-Organization, 2018).

Breast cancer is the second most common cancer occurs (World-Health-Organization, 2018) and most prominent cancer among females in the world which accounts for 23% of the female cancers in the global level (Kawthar Al-Ajmi, Artitaya Lophatananon, Martin Yuille, William Ollier, Kenneth R. Muir, 2018)with a death rate of 570000 reported in 2015 (Yi-Sheng Sun, Zhao Zhao, Zhang-Nv Yang, Fang Xu, Hang-Jing Lu, Zhi-Yong Zhu, Wen Shi, Jianmin, 2017). In 2018, 2.09 million breast cancer cases were recorded and 627000 deaths occurred due to breast cancer worldwide (World-Health-Organization, 2018). There is a higher possibility for breast cancers to be cured with early diagnosis and proper treatment process. Early diagnosis is a key factor for obtaining a good outcome in the treatment process (Polat, 2018). The best effective method is to support early diagnosis (S.Kharya,D.Dubey,S.Soni, 2013), in order to support the treatment process to reduce the deaths due to breast cancer saving the women's lives around the world.

2.1.1.1 What is breast cancer?

Breast cancer is a type of cancer that is developed in the breast tissue. Breast cancer starts with the mutation of breast tissue creating a tumour that is often felt as a lump or which can be identified using an x-ray screening (Anon., 2019). Breast cancer is developed in breast cancer cells. Naturally, the cancer forms in Duct or Lobules in the breast. Lobules are gland which produces breast milk and ducts are the pathways that transport milk to nipples. Most of cancer occurs in fat tissue or Fibrous Connective tissues (Jaime Herndon & Kimberly Holland, 2019) It is a disease that is more prominent in females irrespective of their social and economic levels while the rate of breast cancer in males fall below 1% (Diana Ly1, David Forman2, Jacques Ferlay2, Louise A. Brinton1, and Michael B. Cook1, 2013).

2.1.1.2 Types of breast cancer

Breast cancer types differ from each other by the specific cells in the breast that are affected. breast cancer types can be divided into two main sections. Those are Non-Invasive and Invasive (B Weigelt

HM Horlings B Kreike MM Hayes M Hauptmann LFA Wessels D de Jong MJ Van de Vijver LJ Van't Veer JL Peterse, 2008).

Non-invasive cancer also known as "In-Situ", which means the cancer cell does not spread beyond the milk duct and lobules in the breast. This type of cancer has a higher survival rate than invasive breast cancer. Non-invasive breast cancer can be split into two categories according to the spreading of cancer in breast tissue as Ductal carcinoma in situ(DCIS) and Lobular carcinoma in situ (LCIS).DCIS can be detected through mammogram while LCIS can be detected through bioscopy (Anon., 2018).

Invasive cancer grows in normal tissues in the breast. Most breast cancer types are invasive. Cure rates of invasive cancer are lesser than non-invasive breast cancer. This cancer type can be divided into five main categories as follows,

1.Invasive Ductal Carcinoma.

2.Invasive Lobular Carcinoma.

3.Infammatory Carcinoma.

4.Paget's disease.

5.Rare Types: Male breast cancer, Lymphoma, Scarcoma.

Inflammatory breast cancer (IBC)is a rare and aggressive kind of cancer. IBC tends to grow and spread quickly. IBC starts with breast swelling and redness instead of lumps. Breast cancer swelling starts due to cancer cells clogging the blood vessels that transport lymph. Symptoms of IBC are redness of breast, swelling, warmth, orange peel appearance, skin changes, swelling of lymph nodes, inversion od nipple and aching or burning in the breast. These cancer treatments are chemotherapy, surgery, radiation therapy etc. But the treatment method differs according to the stage of cancer (Anon., 2012).

Paget's disease is a rare type of cancer. In Paget's, cancer cells collect around the nipple area. This cancer starts to spread from the nipple surface to the areola. The causes of that disease include the nipple area becoming scaly, itchy and irritated. The main symptoms of this disease are itching, tingling, pain and sensitivity, scaling and thickening of the skin, flattening of the nipple, and abnormal discharge from the nipple. This cancer can detect through a mammogram, physical examination of the breast, ultrasound or MRI and biopsy. Because Paget's disease of the nipple is rare, doctors often mistake it for eczema, infection or injury, or some other skin condition. For many people, it can take several months to get a correct diagnosis. (Anon., 2016)

Invasive Lobular Carcinoma which is also known as infiltrating lobular carcinoma is a type of breast cancer that is developed in the lobules (milk-producing glands) of the breasts and have the potential to grow and spread to the other parts of the breast tissue and to the other parts of the body (Anon., 2019). It represents 10% - 15% of all the breast cancers diagnosed which is the second-most-common type of breast cancer diagnosed (Anon., n.d.). Invasive lobular carcinoma is most commonly found in pre-

menopausal women but there is a possibility that it can occur in women of any age. This type of cancer is very rare in men (Anon., 2018).

Invasive Lobular Carcinoma can be divided into multiple subtypes base on the nature of cells observed under the microscope. If the cancer cells grow differently to the most typical form of ILC, the subtypes are defined as solid, alveolar and subglobular. When cancer cells itself look different form classic ILC cells, the types of cancer are defined as pleomorphic and signet ring cell (Anon., 2013).

Invasive lobular carcinoma is a disease that is hardly seen on mammograms as cancer spread to the surrounding breast tissue instead of forming a lump (breastcancer.org). The first signs of Invasive Lobular Carcinoma are hardening or thickening of a breast that can be felt by an individual rather than a lump. Other symptoms of ILC include irritation in skin or dimpling, pain in breast, thickening, redness or scariness of the breast skin or the nipple, swelling of part of the breast or all, pain in the nipple, under-arm lump and nipple discharge other than breast milk (Anon., 2019; Staff, 2018).

Diagnosis of ILC is a process involving a combination of procedures which includes a physical examination and imaging tests. Some of the tests used for diagnosing ILC include Ultrasound, Breast MRI, Mammography, Fine needle aspiration biopsy, core needle biopsy, incisional biopsy and excisional biopsy (Anon., 2019) etc.

There are multiple treatments for ILC used depending on the nature of cancer in the individual. The first treatment considered is the surgery and the type of surgery required is decided to depend n the size of cancer compared to the breast, area of the breast affected by cancer, the nature of spreading of cancer in the breast. The types of surgeries used include breast-conserving surgery, mastectomy and surgery to the lymph nodes. After surgery, additional treatment is done if required. The additional treatment methods used to treat ILC include radiotherapy, hormone therapy, chemotherapy and targeted therapy and the method of treatment used to depend on the individual situation of the patient (Anon., 2016).

Invasive Ductal Carcinoma, also known as Infiltrating Ductal Carcinoma (BC org) is a type of breast cancer that begins in the milk ducts of the breast and spreads to the other parts of the breast tissue and beyond (to the other parts of the body). IDC is the most common type of breast cancer representing 70 – 80 per cent of all the breast cancers and it is the type of breast cancer that is most common among men. The occurrence of IDC is very high in older women compared to that of younger women which is almost two-thirds of all the IDC occurrences (Invasive ductal carcinoma, 2018)

The symptoms of Invasive Ductal Carcinoma include lumps in breast, thickening of breasts, swelling of breasts, breast pain, nipple pain, swelling or lump under the arm, discharge from the nipple, inward turning of nipple, pain in breast or armpit and changes in the breast skin (Signs and Symptoms of IDC, 2018),(Johns Hopekins University, n.d.),(Invasive Ductal Carcinoma (IDC), 2019). A wide range of treatments is available for IDC such as mammography, ultrasound scan, Breast MRI, Fine Needle

Aspiration, Core Needle Biopsy, Incisional biopsy and Excisional biopsy (Tests for Diagnosing IDC, 2013) etc.

IDC can be treated using different treatment methods including surgery, chemotherapy, radiation therapy, hormonal therapy and targeted therapy (Invasive Ductal Carcinoma (IDC), 2019). The first treatment method considered is surgery which involves removal of the breast tumour itself. However hormonal therapy and chemotherapy are considered in situations where cancer has spread to several lymph nodes. Targeted therapy which is also known as systematic therapy is used to treat IDC at early stages and in advanced stages (Invasive Ductal Carcinoma (IDC), 2019). The type of targeted therapy used depends on the treatment history and the characteristics of the cancer of an individual (Treatment for IDC, 2019).

2.1.2 Breast cancer prediction

Usually, people visit doctors especially oncologists to check whether they are diagnosed with breast cancer when they observe any symptoms related to breast cancer. Oncologists diagnose breast cancer by undertaking thorough medical checkup history, physical examinations of both the breasts and by checking for swelling or hardening of any lymph nodes in the armpit. Imaging tests like Mammogram, Magnetic Resonance Imaging (MRI) of the breast, Ultrasound of the breast, X-Ray of the breast. Tissue biopsy, Sentinel node biopsy will be taken by the oncologist to check whether the patient was diagnosed with breast cancer. This is the procedure to predict breast cancer and to classify the tumour type whether benign or malignant. (Ch. Shravya, K. Pravalika, Shaik Subhani, 2019).

2.2 Literature review of the technology

2.2.1 Algorithm Selection

The World is developing rapidly, as a result, the revolution of the Information Technology field is now on rapid growth. There are many new Inventions of various concept that came up to help humans in various situation. Machine learning is one of those concepts which was started around in the 1950s (Foote, 2019)Machine learning simply means a machine which is technically computer devices learn from its own experience as like how humans learn from their experience which can be used to identify various insights of a given situation and analyzes the future outcomes.

This project is based on machine learning techniques. Because of that, selecting algorithms which are compatible to produce accurate results is a very important part of it. In machine learning, there are two types.

1. Supervised learning
2. Unsupervised learning

Unsupervised learning is giving away information to the system that is neither classified nor labelled and allowing the algorithm to analyze the given information without providing any directions.

Unsupervised learning problems can be further grouped into clustering and association problems. (Desai, 2018)

- A clustering problem is where you want to discover the inherent groupings in the data such as grouping customers by purchasing behaviour.
- An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X and tend to buy Y

K-means for clustering problems and Apriori algorithm for association rule learning problems are some popular algorithms used in unsupervised learning. (Brownlee, 2016)

Supervised learning is a type of machine learning where the system is trained with a training dataset which includes the input data and desired output data. So, this model will get the ability to predict the output for the given inputs in the future. Basically, it is like teaching a model on how to react in the future. To predict those outputs, the data should go through the algorithms. Some of the algorithms are listed down:

- Linear regression
- Logistic regression
- Artificial Neural networks
- Linear discriminant analysis
- Decision trees
- Similarity learning
- Bayesian logic
- Support Vector Machine (SVM)
- Random forests

The system we are going to implement is based on supervised machine learning techniques. For that, some of the algorithms above mentioned should be selected.

Similar projects which have already implemented based on this topic are most probably about predicting breast cancer using the images of the cancer cells. Using image processing and pattern recognition techniques they have predicted whether the cancer tumour type is benign or malignant. For that, they have used classification algorithms such as Logistic regression, k-nearest neighbour, Support Vector Machines, Naïve Bayes classifier, Decision trees, Sequential Minimal optimization etc.

The system that we are going to implement is mostly going to use numerical and characteristic data. So we are planned to use Support Vector Machine (SVM). It is the most suitable and popular algorithm to build prediction models and to produce accurate results than other algorithms. After the implementation of the system using algorithms, another model will be implemented using Artificial Neural Networks to predict the results more accurately.

2.2.2 Back End Technology Selection

Machine Learning is based on algorithms, where there needs to be some programming language to implement these algorithms. Almost all programming languages can be used for machine learning purposes. Some of the Programming languages for Machine learning are Python, C++, Java, JavaScript, C#, R, Julia, Go, TypeScript, Scala, etc.. (Chand, 2019). As there are several programming languages available for machine learning, the best programming language should be chosen prior to the development, considering the nature of the system to be developed. The best programming language for the machine learning approach is the language that has per build libraries and has advanced support of data science and data models (Chand, 2019). Python is one of the best choices for implementing machine learning models as its robust machine learning model with its powerful library collections. Some of the most popular python libraries are mentioned below.

- Scikit-learn: Scikit learn was introduced as a Google Summer of Code project. It is a robust python machine learning library which features machine learning algorithms such as spectral clustering, mean shift, random forests, cross-validation and more. This ML library provides several supervised and unsupervised ML algorithms and Scikit learn is more focused on data modelling than data manipulating. This library contains a lot of efficient tools for machine learning and statistical modelling (KUNAL JAIN, 2016).
- Pandas: Pandas (Python Data Analysis Library) is an open-source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language (Anon., 2019). Pandas is designed for mainly data manipulation, aggregation, visualization and reading. **Invalid source specified.**
- Keras: It is high level API of TensorFlow which is used to build and train Deep Neural Networks. Keras is an open-source neural network library which facilitates statistical modelling, image manipulation and text manipulation. **Invalid source specified.**
- TensorFlow: It is an AI library which facilitates the creation of large-scale neural networks with multiple layers with the use of data flow graphs. It is also used in building deep learning model and in deployment of machine learning powered applications. It is much efficient in data classification, understanding, discovering, predicting, perception and creation. Tech giants like Intel, Google, Twitter, DeepMind, Coca-Cola and Google use TensorFlow for their operations. **Invalid source specified.**

  Python is a widely-used programming language for machine learning as it is perfect for both the programmers and newcomers to quickly plugin for data science as containing simple syntax (Kharkovyna, 2019). And, another main reason is the wide range of library availability for python. Python libraries provide base-level items, so the developer doesn't have to rewrite the same code

again and again and python allows more data scientists to learn the language in a much shorter time compared to other languages. Especially python language is most flexible, platform-independent, Easy to understand and reliable (Luashchuk, 2019).

2.2.3 Front End Technology Selection

When selecting the best technology to create the web app the main competition was with,

- Android Studio
- Flutter
- React Native

Since the app was intended to support both android and IOS the technology had to support cross-platform. Even though both Flutter and react-native supports cross-platform android studio only supports the Android platform. React Native was built by Facebook and flutter was designed by Google, before flutter came into use React native was already being used to create apps so React native is widely used by most of the companies even now. React Native uses javascript which is a widely recognized web language in the web community. But flutter, on the other hand, uses dart which is easy to understand but rarely used by the programmers. When it comes to UI components and Development of API both technologies must have access to native modules, for this react native thoroughly depends on third-party libraries and flutter doesn't require third party libraries it has an inbuilt component which has direct access to these native modules. Since to use flutter dart programming language must be used and because of the newness of this language most IDE's and text editors do not support so when selecting a text editor for your flutter project you have to be concerned whether it supports dart language or not. Although the flutter community is rapidly growing the react-native have been there in the development society more than flutter because of this react native community and resources have grown since the launch but because flutter is fairly new to the community it is far more behind in resources with compared to react-native. So as for the above facts, it is pretty clear that the most suitable technology that should be used in order to fulfil the requirements that have been mentioned at the start of the project is reacting native, it is further explained using the table below by showing the limitations of each promising technology and why react-native was chosen,

	Limitation	Conclusion
Android Studio	Does not support cross-platform Software fragmentation. Multiple device size compatibility	Since the app should cover both IOS and android platform this technology was not selected
Flutter	Fairly new to use	Because this technology is new to use, it is considered as a technology that is

	Inability to create a project using most of the IDE's because some of them does not support dart language	still developing it may not have all the necessary functions required.
React Native	Have to use third party libraries when using API's and UI components	Even though this technology uses third party libraries it has all the components that are required to create an app compared to the above technologies

Table 3: Comparison between Front end technologies

2.3 Literature review of existing systems

2.3.1 Estrogen receptor status prediction for breast cancer using artificial neural network

There is a cordial relationship between the Estrogen receptor (**ER**) and breast cancer (G. Ball, S. Mian, F. Holding, R.O. Allibone, J. Lowe, S., 2002). There are various researches have been done to identify informative genes that are highly joint with ER status. In this paper done the research to predict breast cancer by getting ER-positive and ER-negative status through an artificial neural network (ANN) (G. Bloom, Yang IV, D. Boulware, K.Y. Kwong, D., 2004). Several data mining methods have been intended for analyzing breast gene microarray data (J. Budczies, W. Weichert, A. Noske, B.M. Müller, C.Weller, T. Wittenberger, H.P. Hofmann, M. Dietel, C.Denkert and V. Gekeler, 2011).

In this research use the Artificial Neural network (ANN) to produce a prediction. First divide data set into 3 subsets for training, test and validation purposes (L.J. Lancashire, O. Schmid, H. Shah and G. Ball, 2005). A 3-layered ANN with backward propagation of errors to supervised learning, learning was constructed for predicting the training samples. Then categorized microarray data of breast cancer patient and analyzed it. After analyzing highly ranked genes by the model have revealed a significant association with the ER status associated with breast cancer. An ANN prediction model was constructed based on the most relevant genes. Using the validation set (GOPAL K. DHONDALAY, DONG L. TONG, GRAHAM R. BALL, 2011).

This model has shown the value of identifying the most relevant gene subset from breast cancer gene microarray data by using 9 out of 100 genes. ANN shows it is a powerful predictor to differentiate ER status (GOPAL K. DHONDALAY, DONG L. TONG, GRAHAM R. BALL, 2011).

2.3.2 Intelligent breast cancer prediction model using data mining techniques

In this model predict breast cancer by using data mining techniques. A characteristic selection method INTERACT is using for breast cancer diagnosis, and the support vector machine (SVM) is used to build the classification model. Data mining techniques are used to create a prediction model. By using those

models find the relationship between diseases and their symptoms (J. M. Moguerza and A. Mu˜noz, 2006).

SVM is applied to build the category model. Through the experiments, the accuracy of the diagnostic model with feature selection is improved (Runjie Shen∗ Yuanyuan Yang Fengfeng Shao, 2014). Meantime, nine features are chosen out as the relevant factors for building the diagnostic model. (I. Guyon and A. Elisseeff, 2003).

2.3.3 Breast Cancer Prediction Based on Backpropagation Algorithm

Azmi and Cob (2010) two researchers had done research on Breast Cancer to classify the tumour from a symptom that causes breast cancer disease and to develop more cost-effective and easy-to-use systems for supporting clinicians. Here to develop this system they have used a machine-learning algorithm named 'Back Propagation' which is used in the Artificial Neural Network (ANN).

For the Breast Cancer research that was done by Azmi and Cob (2010) dataset was taken from the UCI Machine learning repository to feed the model to get a better prediction. To select the best algorithm model which gives the correct accuracy they have tried three main algorithms and from that according to the accuracy percentage that was got, they have chosen the back-propagation algorithm as the most suitable algorithm, which leads to developing a better system. The following table shows the accuracy of the percentage of three algorithms that were tried in the research model.

The classifier	**The classifier**
Backpropagation	96.63
Support Vector Machine	96.19
Decision Tree	92.38

Table 4: Performance comparison

When Comparing to the System We are implementing this system that does not give early prediction give awareness for the people who are not affected to the breast Cancer disease as it only relies on the fact of the breast cancer cell.

2.3.4 A Novel Breast Cancer Prediction System

New technologies like data mining and knowledge discovery have developed allowing researchers and developers to discover knowledge and find hidden patterns in large data sets. Using these new technologies, the two types of research (Tee, I.C. and Gazala, A.H., 2011) has decided the Support system for Breast cancers disease.

The Research Which was done by two researchers (Tee, I.C. and Gazala, A.H., 2011) on breast Cancer was a Decision Support system that was designed to help oncologists and medical professionals in eliminating suffering and death happen due to Breast Cancer. By using the SEER database data set they had created a prototype which predicts a Breast Cancer Survivability rate of a Patient after five years

from the day of the diagnosis done. Classification and regression approach was taken when building this model. (Tee, I.C. and Gazala, A.H., 2011)

The Decision Support System that was built by Tee and Gazala is a Web-based Decision Support system designed using data mining and knowledge discovery technologies. As a development of every machine learning model this also was done by first getting proper dataset, then pre pre-possessing data Which they have broken the dataset into 3 parts, and finally the model's implementation for both classification and Regression. The model for classification part had done using a decision tree algorithm and the Regression part had done using Generalized Linear Regression algorithm.

2.3.5 Analysis of Efficiency of Classification and Prediction Algorithms (Naïve Bayes) for Breast Cancer Dataset

Breast Cancer Which is Commonly Identify Among the Women are most of the time effect because of the tumour Which is Consist of two types of Cell one is Benign Which is Non-Cancerous cells and the Other is Malignant is Cancerous cell that affects the body mostly (Lekha, A., et al, 2015).

By using the dataset of Wisconsin University database (Lekha, A., et al, 2015) researched for a prediction and classification System Which classify and predict if a tumour is either benign or malignant in Breast Cancer Disease, for this they used the Naïve Bayes algorithm to build the machine learning model for both prediction and classification. Even though the prediction system works greatly the limitation of this system is that it only gives more accuracy for large instances of the dataset, when the dataset is in a small number the error rate is somewhat high.

Leka et al (Lekha, A., et al, 2015) Carried Out the experiment relates to the prediction and Classification of Breast Cancer using Tumor in MATLAB (matrix laboratory) Which is a 4th Generation High-level language and a multi-paradigm numerical computing environment.

2.3.6 Artificial Neural Network for Prediction of Breast Cancer

Singhal and Pareek two researchers have done research on building systems that predict breast cancer at the earliest. By taking the Wisconsin Breast Cancer (Diagnosis) Dataset they have trained and tested their model. The approach that was taken to build the model was using Artificial Neural Network methods Feed Forwarding and Back Propagation algorithms. (Singhal, P. and Pareek, S., 2018)

Singhal and Pareek research of predicting Breast Cancer was conducted using Dev.-C++ software and implemented using C language. Using the ANN algorithm of Feed Forwarding and Back Propagation the model of machine learning was built for this system. The accuracy level of this technology in the system was up to 98 % with a low error rate. (Singhal, P. and Pareek, S., 2018)

2.3.7 Prediction of breast cancer using supervised machine learning techniques

There are two types of tumours in breast cancer. To identify the patient's condition regarding breast cancer, the patient should undertake various imaging tests in hospitals. Diagnosis of this disease manually takes long hours and lesser availability of systems. This is a system invented to differentiate the breast cancer tumour type whether it is benign or malignant. This classification will reduce unwanted costs for a patient on their medical tests. Because benign tumour patients should undertake lesser treatments than a malignant tumour patient. So, this will avoid a benign patient to take needless treatments.

To classify the tumour type, classification techniques of machine learning was used from the past data collected from UCI. Initially, Dimensionality Reduction was used to differentiate the independent and principal variables. The dataset they have used is multidimensional and the attribute count is high. So, they have used Principal Component Analysis (PCA), one of the Feature Projection Dimensionality Reduction technique to derive two principal components of the database. Logistic Regression, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) Algorithms were used to build the machine learning model. The system was built in Spyder, the scientific python development environment. According to the test data, among these three algorithms, SVM algorithm has provided the best accuracy rate of 92.7% (Ch. Shravya, K. Pravalika, Shaik Subhani, 2019).

2.3.8 Breast Cancer Prediction Using Data Mining Method

This is a system which was invented to predict breast cancer using machine learning data mining methods. Even though there are many technologies used in predicting the breast cancer, this system was built on top of four data mining methods applied in. They are Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naïve Bayes Classifier and AdaBoost tree. So, this research led them to find which way of implementation led them to the most accurate results. The datasets they have used are Wisconsin Breast Cancer Database (1991), and Wisconsin Diagnostic Breast Cancer Database (1995) because most of the similar researches were built on top of these databases. So, it was easy to compare whether this system is better than the others. This system was trained using patient's clinical data such as clump thickness, uniformity of cell size, marginal adhesion etc. Principal Component Analysis is also implemented to identify the principal attributes. According to the test data, the model implemented using the SVM algorithm showed a higher accuracy rate of 97.1% for WBC database and 97.99% for WDBC database. (Haifeng Wang and Sang Won Yoon, 2015).

2.3.9 A Novel Approach for Breast Cancer Detection using Data Mining Techniques

This system was implemented to investigate the performance of different classification techniques to detect breast cancer. For that Weka was used, an ensemble of tools for data classification, regression, clustering, association rules and visualization. Weka version 3.6.9 was used. The database used for this system is taken from the UCI machine learning repository. This system was implemented using 3

various classification techniques to compare the results. They were IBK(KNN), Best First tree (BF Tree) and Sequential Minimal Optimization (SMO). Among these techniques, the model built on top of SMO algorithm (A new algorithm for training SVM. The advantage of SMO is, it can be implemented simply and analytically) has shown a higher accuracy rate of 96.19%. (Vikas Chaurasia, Saurabh Pal, 2014)

2.3.10 Breast Cancer Prediction Using Data Mining Techniques

Early-stage detection of cancer can help to cure cancer and also it will help to prevent it from recurrence. Data mining is the process of extracting promising patterns and information from the gathered data. According to a study conducted by **Invalid source specified.** about data mining algorithms, it has mainly focused on the accuracy of predicting the possibility of recurrence of the sickness among the patients on the basis of important parameters. In this research the two main algorithms that have been taken into consideration were,

1. Clustering algorithm
2. Classification algorithm

The study looks into four clustering algorithms which are K means, EM, PAM and Fuzzy c-means and four classification algorithms which are SVM, C5.0, Naive Bayes and KNN. In the clustering algorithm, the data which is being used is separated into groups of clusters or sub-clusters. The k-means clustering works by clustering the data into k subclasses defined as for centroids. EM clustering on the other hand which stands for Expectation-maximization is a model that uses unobserved variables to decide the parameters using maximum-likelihood. The PAM model is very similar to K-means except for the fact the in this model grouping into a given number of sub-clusters is based on K-medoids. In fuzzy-clustering data, elements may belong to multiple clusters. In the classification algorithm, it classifies different objects on the basis of training a dataset whose outcome is known. In KNN classification objects are assigned to a class based on most common k nearest neighbours. In the SVM classification first, the data is separated into different points and then grouped into different classes. The objective of this algorithm is to find a plane that has the maximum margin (having the maximum margin will provide some reinforcement to give an idea of how the classification will work in the future). The Naive Bayes model is decided based on the Bayes rule; it will calculate the probability of a dataset that can belong to a class on this rule, the last classification model is C5.0 is a decision tree that will create observation branches with the motive of improving the prediction accuracy.

At the end of the study, it has come to a conclusion that classification algorithm predictors are better than clustering predictors because both SVM model and C5.0 decision tree have an 81% accuracy rate and fuzzy c-means have come up with the lowest accuracy of 37%.

	Limitations

Classification algorithm	• Takes more time to compare • Testing is slow • Requires large storage space.
Clustering algorithm	• Inability to detect if a dataset contains clusters • Inability to detect the correct number of clusters • Inability to detect special patterns

Table 5: Limitations of Classification and Clustering algorithms

Chapter 03: Methodology

3.1 Research Methodology

For a successful and quality project completion, the resources need to be managed in an efficient way by having an appropriate timeline and a financial plan.

Research Methodology	Research Approach	Based on how the research circle commence the research approach is decided. If particular research is started on the basis of an existing hypothesis or a theory and then continued on to extract data based on that, it can be distinguished as a deductive approach but inductive approach, on the other hand, is more concerned on with generating a new theory based on the data that has been already collected. So according to the project that we are carrying out, which is predicting the possibility of a woman being diagnosed with breast cancer based on lifestyle factors which is a theory that is already there so the deductive approach was chosen as the most suitable approach. Since deductive research approach is more aimed at testing a hypothesis it is most suitable to work with quantitative data.
	Strategy of Research	The sole purpose of the strategy of the research is to define how the research questions were handled. In order to make the research accurate and successful out of mediums such as interviews, case studies, surveys, experiments etc. mainly surveys were chosen as it is following a quantitative strategy. Also, mediums such interviews were used to make the project error-free.
	Research philosophy	This refers to the assumptions we make throughout research consciously or unconsciously about the development of knowledge. (Saunders, et al., 2019)

	Time management	Two ways of time management methods are Cross-Sectional studies and Longitudinal studies. Out of these two Longitudinal studies was chosen as the research is based on the same sample of data throughout the research during different points in time. Also, as the behaviour will change from individual to individual.
Data Collection Methodology	Choice of data collection	In this case, both quantitative and qualitative data were used because to make the prediction more accurate sensitive data had to be covered. Since the research is based on both qualitative data and quantitative data, mixed-method was chosen out of Mono method, multi-method and mixed-method. So, by carrying out surveys, interviews mainly the required as quantitative data were collected and statistical data as qualitative data. So the mixed method of both quantitative research and qualitative research was the ideal option to make the data collection a success.
	Techniques used for data collection	Data will be collected through various means. These techniques include statistical survey data, web-based questionnaires, one to one interviews with experts, journal articles, organizational reports and related research papers.

Table 6: Research Methodologies

3.2 Project Management Methodology

The completeness of the project is mostly dependent on how the project was conducted and managed, among many promising methods to follow such as waterfall, PRINCE2, CPM and CCPM the agile method was selected. Following graph shows how an agile method proceeds to analyze a project,

Reasons for selecting the agile method

- Categorize the project management life cycle into different phases.
- It supports flexibility so whenever a change is required or need to be updated agile method will allow making necessary changes
- Reduced risk (it will prevent project failure because of the flexibility)
- Project predictability

Reasons for not selecting the other methods

Method	Reason
Waterfall	Even though this method is popular for use in these kinds of software development projects because of the fact that its linear, sequential and inflexibility developers have to develop each phase with complete confidence because once a phase is completed there is no turning back. Since there is a higher risk in using this method it was not chosen.
CPM	In this method, the time estimation is a bit difficult and also just like the waterfall method it cannot effectively handle sudden changes it is very difficult to draw the CPM chart over again if anything changes plus it consume time to select the correct path.
PRINCE2	This method only supports the waterfall approach and the documentation is heavy with compared to other methods, so this was ruled out from the options.

Table 7: Reason for not selecting methodologies other than Agile

3.3 Risk Management

Risk	Mitigation	Level	Frequency
Finding a suitable dataset. Since the project is about making a prediction on breast cancer all the datasets related to cancer wasn't suitable the	In most projects datasets are widely available but when it comes to predicting breast cancer it deals with a lot of sensitive data so it was hard to find datasets that could be used in the process but through	High	Low

dataset needed to be specific about breast cancer.	consulting expertise and using the internet it was confirmed that the required dataset was available and can be collected.		
Updates and changes. In the development period, several changes in the development may occur. The changes had to be taken in	The reason for not choosing the waterfall method and following the agile method is because of this. In the agile method, the development will take place as several phases and you can go back each phase whenever required and make changes if necessary.	High	Medium
Limited knowledge on the domain. Since the problem domain was related to the medical field in order to continue on to come up with an accurate solution for this problem a broad knowledge and understanding of concepts were required.	Since breast cancer is the second most common cancer in the world considerable amount of people have done researches regarding this topic so it was possible to gain basic knowledge that is required to conduct the project by going through research papers and having one to one interviews with expertise	High	Low
Limited development time. Meeting all the requirements and producing an end product within the given time period.	By dividing the whole project into phases and then allocating team members for each phase correctly according to their knowledge gave a head start to the project and required confidence to complete the project in the given time.	High	High

Table 8: Risk Management

3.4 Work Plan

3.4.1 Activity Schedule

Task	Start	End
1. Project initiation	1st October	7th December
2. Literature review	1st October	7th January
3. Requirement gathering	1st December	31st December

4. Designing the system	14th December	16th January
5. Selecting tools and techniques	24th December	16th January
6. Prototype implementation	1st January	28th February
7. Testing and evaluation	15th February	12th March
8. Documentation and evaluation	21st March	28th March

Table 9: Activity tasks

3.4.2 Detailed Activity Schedule

Task	Start	End
1. Project initiation	1st October	7th December
1.1. Problem definition	1st October	31st October
1.2. Problem domain	1st October	13th November
1.3. Research questions	15th October	15th November
1.4. Project scope	7th November	23rd November
1.5. Research aim, objectives	24th October	30th November
1.6. PID review	1st December	7th December
2. Literature review	1st October	7th January
2.1. LR of the domain	1st October	10th November
2.2. LR of related research	1st October	30th November
2.3. LR of existing technologies	1st November	31st December
2.4. LR of existing systems	15th November	7th January
3. Requirement engineering	1st December	31st December
3.1. Gathering functional and non-functional requirements	1st December	8th December
3.2. Stakeholder analysis	7th December	20th December
3.3. Questionnaire findings	1st December	27th December
3.4. UML diagrams	16th December	31st December
4. Designing the system	14th December	16th January
4.1. Design goals	14th December	22nd December
4.2. Design high level architecture	23rd December	16th January
4.3. Design low-level diagrams	20th December	7th January
5. Selecting tools and techniques	24th December	16th January
5.1. Comparing different tools and technology	24th December	7th January
5.2. Selecting the most suitable tools and technology	2nd January	16th January
6. Prototype implementation	1st January	28th February

6.1. Implementing Core functions	1st January	6th February
6.2. Enhancing the application	18th January	14th February
6.3. Reviewing work	7th February	20th February
6.4. Mentor feedback	23rd February	28th February
7. Testing and evaluation	15th February	12th March
7.1. Test plan	15th February	28th February
7.2. Unit testing	1st March	12th March
7.3. Integration testing	1st March	12th March
8. Documentation and evaluation	21st March	28th March
8.1. Creating a draft project report	21st March	26th March
8.2. Mentor review on the final product	24th March	28th March

Table 10: Detailed activity schedule

3.4.3 Work breakdown structure

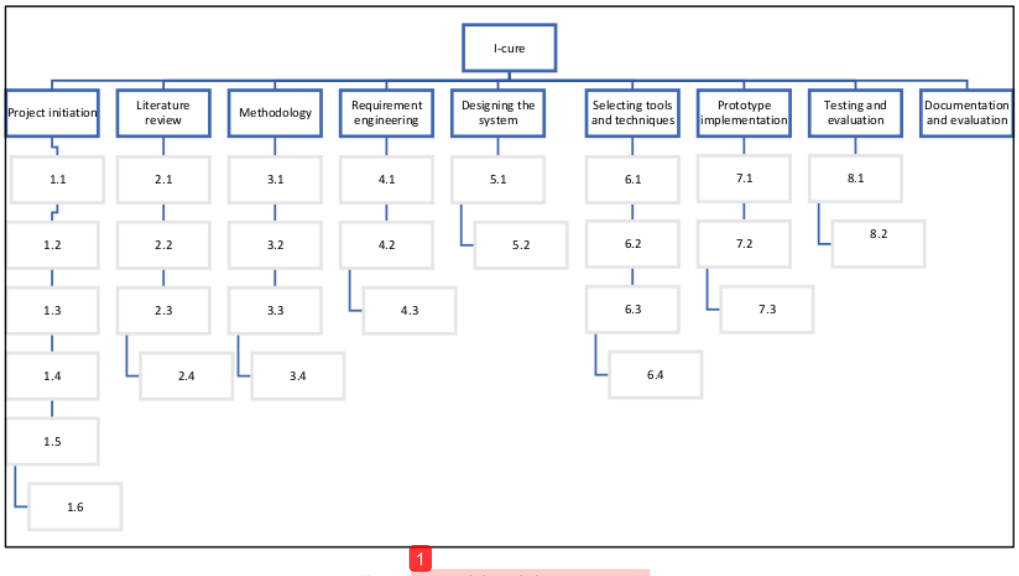


Figure 2: Work breakdown structure

3.4.4 Gantt Chart

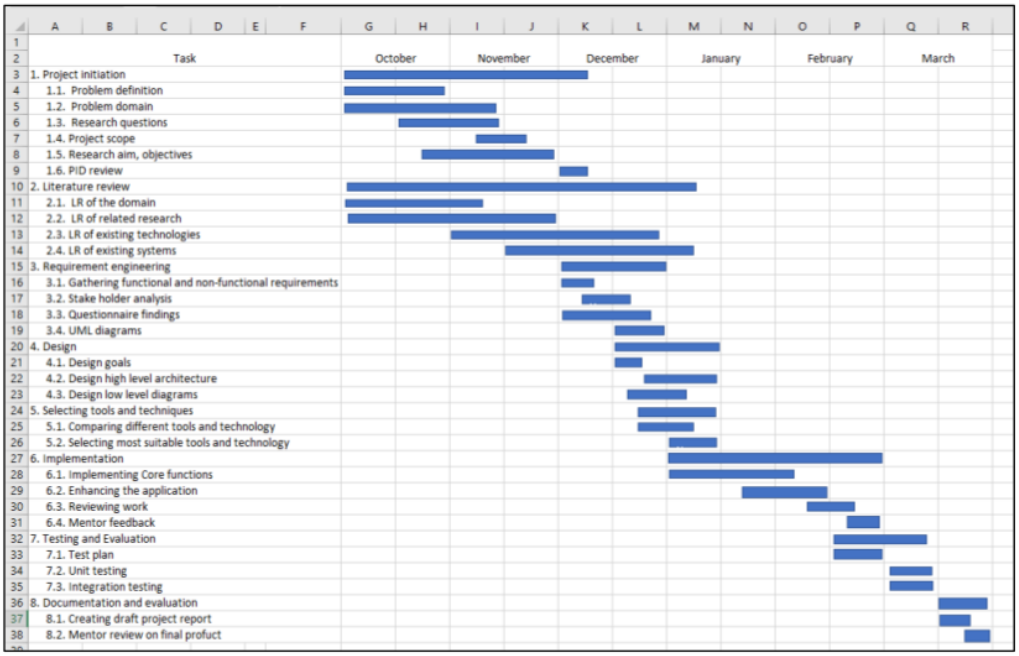


Figure 3: Gantt Chart of the work plan

Chapter 04: Requirement Engineering

4.1 Stakeholder analysis

Demonstration of the relationships between the stakeholders and system using onion model and explanation of the role played by stakeholders are described.

4.1.1 Onion model

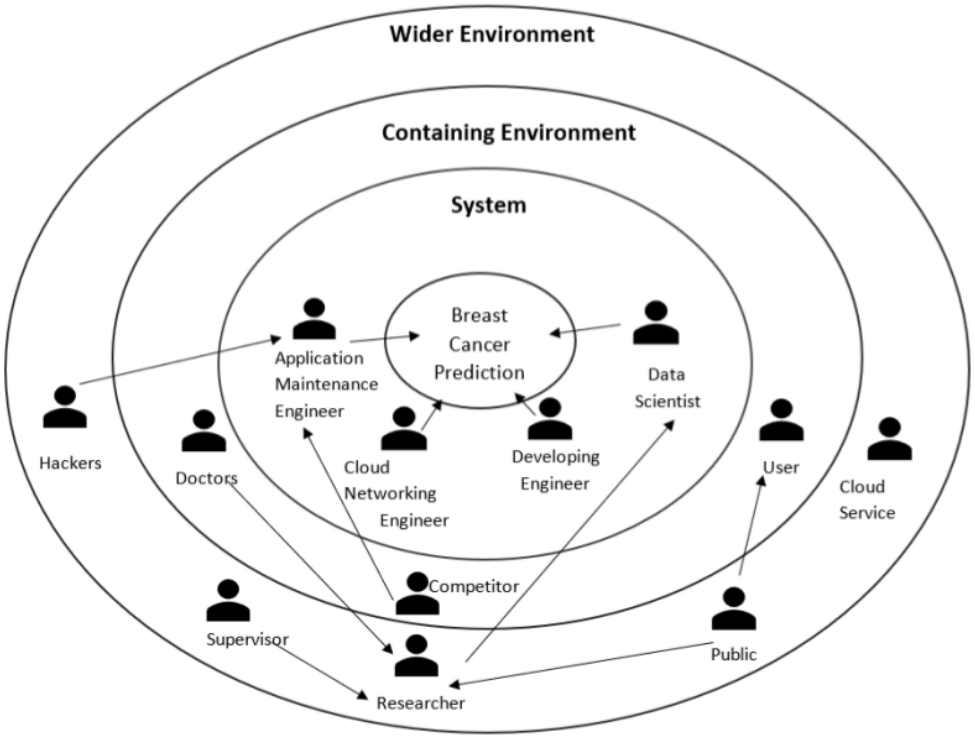


Figure 4: Onion model

4.1.2 Description of stakeholder roles

Stakeholder	Role	Benefits
Application Maintenance Engineer	Technical Support	Managing database and the application development
Cloud Networking Engineer	Operational Maintenance	Collaborating with stakeholders to improve network performance

Developing Engineer	Operational Maintenance	Responsible for designing, creating and testing the system
Data Scientist	Operational Maintenance	Responsible for pre-processing of data which is needed to develop the machine learning model
Doctor	Data Gathering Support	To get information and expert advice related to the problem domain
Competitor	Development of Product	Enables to identify specific and unique traits that are appealing to customers through the product
User	Functional Beneficiary	Uses the machine learning model developed using the tool
Supervisor	Advise Support	More concerned with orchestrating and controlling the work of a group
Hackers	Negative Stakeholder	Intend to damage the system and interrupt the whole system and its related data
Researcher	Qualitative research of the problem domain	Analysis of existing system and identifying the difference of the technologies used between various system related to the problem domain
Cloud Service Provider	Web Support	Provide computing, storage and network facilities required for the system functionality

Table 11: Stakeholder roles

4.2 Requirement engineering methodology

The requirement gathering process for the development of the system is an important part of the project as it is the base that the project depends on. This section of the document explains about various methodologies and practices followed in information gathering process for the system development. Observation of existing systems, brainstorming and questionnaires are some examples of information gathering methods followed in the process.

4.2.1 Observation of existing systems

Observation of existing systems is one of the most efficient methods to gather information required for the system. Analysis and investigation of existing systems are helpful in identifying current trends and finding insights for the development of a new system.

Advantages	Disadvantages
• The clear idea of what the existing system is and how it differs from the system that we are trying to implement. • The technologies used and what are the problem faced when implementing them. • Limitations of the system and methods of overcoming them.	• Requires more knowledge in the field of investigation and require more time to identify the insights and idea of the implementation of the existing system. • Need to read more research and journal articles to get a clear knowledge of the systems
Findings	
Most of the existing breast cancer prediction models require mammography and other test results but does not use a person's daily lifestyle patterns and habits in the breast cancer risk prediction process.	

Table 12: Analysis of requirement elicitation through observation of existing systems

4.2.2 Brainstorming

Brainstorming is another methodology to gather requirements necessary for the project. This is conducted among the members of the group to get a conclusion of the project as what to do? How to do? etc. Here it is critically and logically analyzing the idea of each member considering the flexibility and feasibility of the ideas given.

Advantages	Disadvantages
• Ability to identify proper domain for the project by analysis and discussion of related problems.	• Consuming huge amounts of time to finalize and confirm an idea considering the feasibility and flexibility factors of the ideas presented.
Findings	
The project problem domain, the solution required to solve it, the technologies to implement and the feasibility issues were identified.	

Table 13: Analysis of requirement elicitation through Brainstorming

4.2.3 Literature review

The foundation of a project starts from going through several literature reviews as it is the main source to gather information required for the design and development of a project. The literature review of this project is produced for the problem domain, existing systems and technologies used.

Advantages	Disadvantages
• Ability to get a clear idea of the problem domain, technologies used, existing works in research etc.	• Should go through almost all the literature review related to the project to get an idea of the system that implemented in it and differentiate it.

	• Time-consuming.
Findings	
Identification of the problem domain (getting a clear idea of what breast cancer is) and its effect on women, existing work and their effect on prediction and early identification of breast cancer. Almost all the research done regarding the breast cancer prediction is based on breast cancer cell using machine learning algorithms like ANN (Artificial Neural Network), Naïve Bayes etc.	

Table 14: Analysis of requirement elicitation through literature review

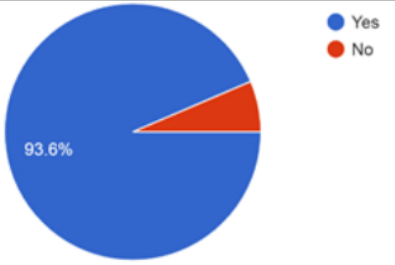
4.2.4 Questionnaire

The Questionnaires related to this project was done using google forms**.** This is an indirect and economical way of accumulating information from the targeted audience.

Advantages	**Disadvantages**
• Wide coverage of people to attach with the project Idea. • The rapidity of the responses help in time-saving • Questionnaire ensures anonymity to its respondents, so the responder is more comfortable and confident to send the response.	• Limited responses. • Sometimes the data obtained can be unreliable
Findings	
Getting responses from the responders who have breast cancer-related people in their families and identifying public opinion on gathering and utilizing their lifestyle information in the cancer prediction process.	

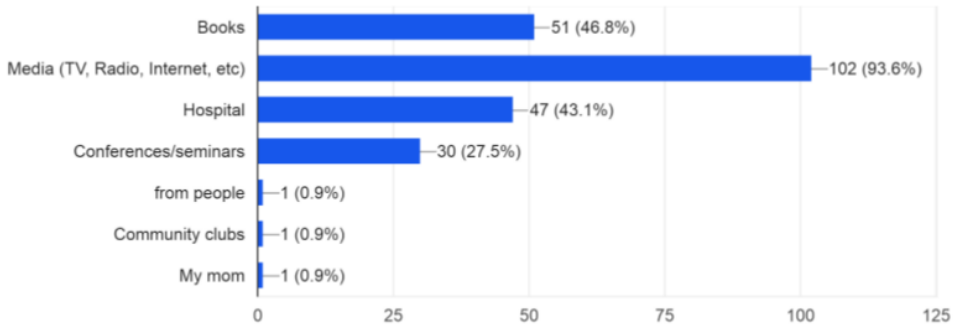
Table 15: Analysis of requirement elicitation through a questionnaire

4.3 Questionnaire findings

Question	Have you heard of breast cancer
Aim of question	To identify the knowledge about breast cancer in people of Sri Lanka
Observations	
 Of all the responders responded, 93.6% of responders had heard about breast cancer while 7.4% of responders had not heard about breast cancer before.	

Conclusion
Majority of people living in Sri Lanka are aware of breast cancer while still there is a small percentage of the population who are not aware of breast cancer which is a deadly disease. There still is a requirement for measures to be taken to increase breast cancer awareness among people as breast cancer is a curable disease if it is identified at earlier stages.

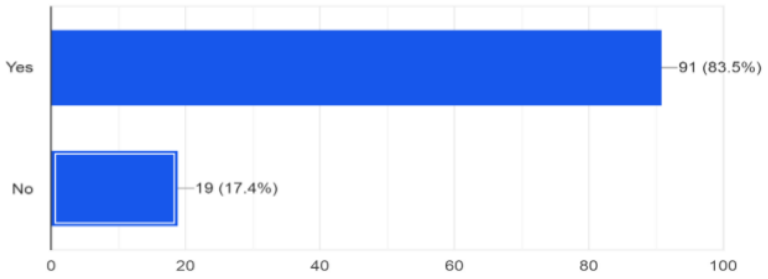
Question	What is your source/s of information
Aim of question	To identify all the sources and their availability to acknowledge people about breast cancer in Sri Lanka
Observations	



It was observed that there are different sources available for increasing breast cancer awareness in Sri Lanka. 93.6% of responders responded media (such as TV, radio, internet etc..), 46.8% responded books, 43.1% responded from hospitals, 27.5% responded from conferences/seminars while only 0.9% responded each for from people, community clubs and from parents.

Conclusion
There are different sources and programs conducted to increase breast cancer awareness in public in Sri Lanka. It can be concluded that mainstream media such as tv and radio, different knowledge sources found on the internet (websites, journals, videos etc..) and books are most effective in increasing breast cancer awareness.

Question	Did you know that breast cancer is the most common cancer that is frequently identified among women
Aim of question	To analyze the level of knowledge of breast cancer occurrence in people of Sri Lanka
Observations	



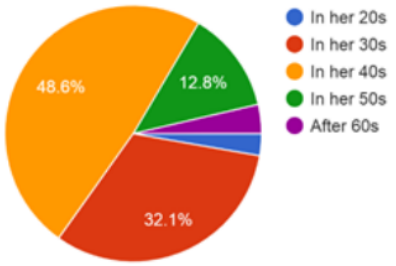
83.5% of responders were aware of higher occurrence of breast cancer in women while only 17.4% wasn't unaware of the fact.

Conclusion

Breast cancer is the most common type of cancer identified in women around the world. Most of the people in Sri Lanka are aware of the higher occurrence of breast cancer in women while only a small percentage doesn't have the knowledge regarding breast cancer occurrence.

Question	When do you think a woman is at a greater risk of developing breast cancer
Aim of question	To identify the opinion in people about the age-related breast cancer occurrence

Observations



From the responses received, it was clearly observable that most of the responders have an opinion that breast cancer risk is high in people in age range 40-49, while only 3.7% of responders think that breast cancer occurrence is high in 60 years and above. Only 2.8% of responders think breast cancer risk is high in people in their 20's. The corresponding responses for greater risk of breast cancer in ages 30's and 50's was 32.1% and 12.8% respectively.

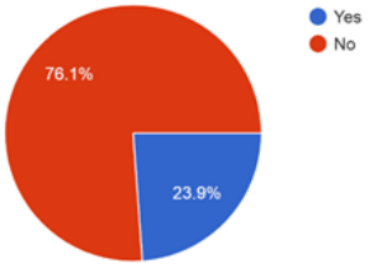
Conclusion

According to the observation, it can be concluded that majority of the people living in Sri Lanka thinks that breast cancer risk is higher in people in age ranges 30's and 40's. Only a small percentage of people thinks breast cancer risk is higher in age ranges 20's and after 60's. There is a requirement to increase public awareness of age-related breast cancer occurrence as breast cancer risk increases with the increasing age

Question	Have you heard of a product that will predict the possibility of humans being diagnosed as cancer patients on beforehand

Aim of question	To identify the knowledge about breast cancer predicting systems in people of Sri Lanka

Observations



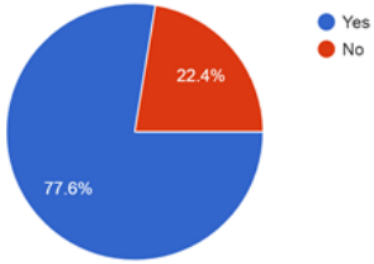
It was clearly observable that from all the responders responded, only 23.9% of responders had any knowledge of breast cancer predicting systems while most of the responders didn't have any knowledge of breast cancer predicting systems.

Conclusion

There are multiple numbers of machine learning models and tools developed for predicting breast cancer risks considering genetical and non-genetical factors. But many people living in Sri Lanka are unaware of any of those tools which are used for breast cancer prediction. Only a small number of people are aware of breast cancer prediction tools which are useful in early identification of breast cancer.

Question	Is it okay to track the daily lifestyle of a person to predict the cancer possibility of that person
Aim of question	To get the opinion in people on collecting and using the information of their daily lifestyle to predict their cancer possibility

Observations



Of all the people responded, most of the responders were okay with tracking and collecting their daily lifestyle information for the purpose of predicting cancer risk while only a 22.4% of responders were not satisfied.

Conclusion

This question is used to obtain the opinion of people on tracking and collecting their daily lifestyle information for cancer prediction. Majority of the people agree with tracking and collecting their daily lifestyle patterns and information for disease prediction by multiple parties. There is a small number of people who are not satisfied with the collection of their daily lifestyle information. As the majority of the people agrees with collecting and using

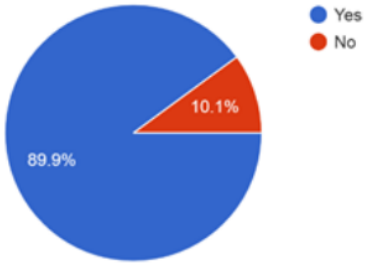
their information for cancer prediction, there is a clear advantage in producing disease prediction systems using daily lifestyle information.	
Question	Would it be helpful to have a mobile application to predict the possibility of having breast cancer
Aim of Question	To find the preference in people to use a mobile application to predict breast cancer possibility in them
Observations	
	From all the people responded, 89.9% of responders were happy to use a mobile application which will predict the breast cancer possibility in them while only 10.1% of responders were not satisfied with a mobile application to predict breast cancer possibility.
Conclusion	
There are multiple health tracking and disease prediction mobile and web applications are available for use. Although a fewer number of people are not satisfied with a mobile application for breast cancer prediction, the majority is satisfied with a mobile application for prediction of breast cancer risk in them. Therefore, a mobile application which could predict breast cancer risk in an individual considering genetical and non-genetical factors would be effective in increasing breast cancer awareness and early detection of breast cancer.	

Table 16: Questionnaire findings

4.4 Context diagram

Proper knowledge of the system environment, it's boundaries, the component of the system and their interactions are important in the system design process. Context diagrams are used for visualizing this high-level view of the system and to obtain a better understanding of it.

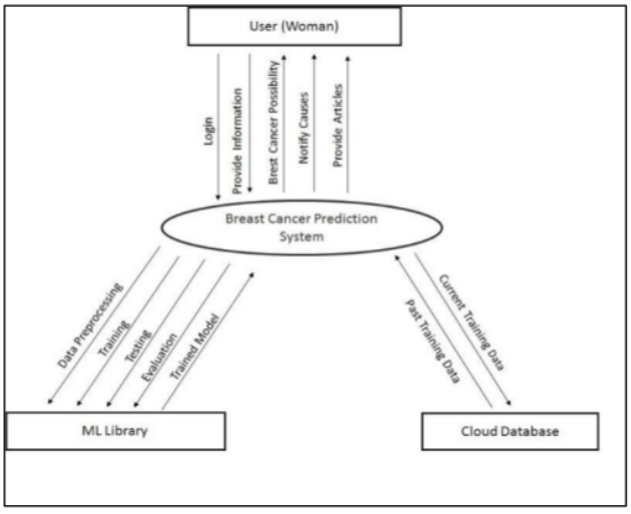


Figure 5: Context diagram

4.5 Use case diagram

The interaction of the user with the system is clearly shown by the below use case diagram.

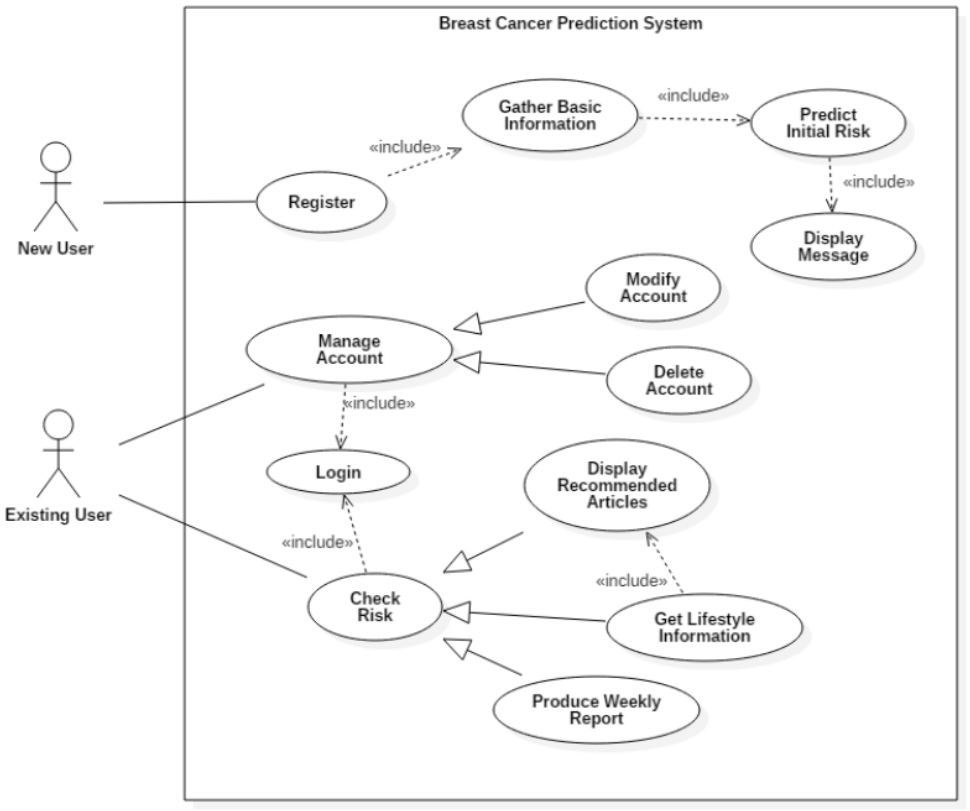


Figure 6: Use case diagram

4.6 Use case descriptions

4.6.1 Use case description for user registration use case

Use Case No	001
Priority	High
Use Case Name	Register
Primary Actor	New User
Secondary Actor	N/A
Pre - Conditions	The user should not exist in the database The network connection should be active The username should be unique
Post Conditions	The user is registered in the system The user is prompted with the privacy policy agreement

Extended Use Cases	N/A	
Included Use Cases	Initial Data Gathering	
Main flow	User	System
	2. Enter email 5. Enter password	1. Prompt for email 3. Validate Email 4. Prompt for password 6. Validate password 7. Prompt for user agreement acceptance 8. Add user to the database 9. Navigate to the basic information form
Alternate flow	N/A	
Exceptional flows	The email entered is not a valid email The password entered does not match the valid password conditions The user is already registered on the system The user has declined user agreement	

Table 17: Use case description for user registration use case

4.6.2 Use case description for user login use case

Use Case No	002	
Priority	High	
Use Case Name	Login	
Primary Actor	Existing user	
Secondary Actor	N/A	
Pre - Conditions	The user must exist in the database The network connection should be available	
Post Conditions	The user is logged in to the system The user is navigated to the main menu	
Extended Use Cases	N/A	
Included Use Cases	N/A	
Main flow	User	System

	2. Enter username/email 4. Enter password	1. Prompt for username/email 3. Prompt for password 5. Authenticate user 6. Navigate user to the main menu
Alternate flow	The user is imported from the device backup storage in case of data loss	
Exceptional flows	The username entered is not found on the database The password entered is not correct Internet connection is not available	

Table 18: Use case description for user login use case

4.6.3 Use case description for initial breast cancer risk prediction use case

Use Case No	003	
Priority	High	
Use Case Name	Predict initial risk	
Primary Actor	New Actor	
Secondary Actor	N/A	
Pre - Conditions	The user should be registered in the system The network connection should be available The user has entered the required information	
Post Conditions	The user is provided with the initial breast cancer prediction The user is provided with the feedback	
Extended Use Cases	N/A	
Included Use Cases	Display Message	
Main flow	User	System
	4. Confirm prediction message	1. Process obtained information 2. Produce breast cancer risk prediction 3. Display breast cancer risk percentage 7. Display feedback message
Alternate flow	N/A	
Exceptional flows	The user has not entered the required information A network connection is not available	

Table 19: Use case description for initial breast cancer risk prediction use case

4.6.4 Use case description for weekly report production use case

Use Case No	004
Priority	High
Use Case Name	Produce Weekly Report
Primary Actor	Existing User
Secondary Actor	N/A
Pre - Conditions	The user is logged in to the system The user must select the weekly report option The network connection should be available The user has entered the required information
Post Conditions	The user is provided with a weekly report
Extended Use Cases	N/A
Included Use Cases	N/A
Main flow	User: 1. Select the weekly report option System: 2. Retrieve weekly report from the database 3. Display the report for the user
Alternate flow	N/A
Exceptional flows	A network connection is not available

Table 20: Use case description for weekly report production use case

4.6.5 Use case description forget lifestyle information use case

UC No	**005**
Priority	High
UC Name	Get lifestyle information
Primary Actor	Existing User
Secondary Actor	N/A
Pre - Conditions	The network connection should be available The user must be available in the database The user must be logged in to the application The user should have entered basic information and should have viewed the initial breast cancer risk

Post Conditions	The confirmation message is displayed A customized message is displayed to the user	
Extended Use Cases	Display recommended articles	
Included Use Cases	N/A	
Primary Path	User	System
	1. The user selects "Enter daily information" 3. Enter daily information (by selecting icons) 4. Press the confirm button	2. Display the daily information pane 5. Store obtained information in the database 6. Process obtained information and identify risk factors affecting the user's breast cancer risk 7. Generate and display a message to the user informing about user's risk factors
Alternate Path	N/A	
Exception Path	A network connection is not available User has not entered the basic information User has not viewed the initial risk The user has pressed cancel button in the middle of entering lifestyle data	

Table 21: Use case description for get lifestyle information use case

4.6.6 Use case description for gather basic information use case

UC No	**006**
Priority	High
UC Name	Gather basic information
Primary Actor	New User
Secondary Actor	N/A
Pre - Conditions	The user has registered in the application A network connection is available
Post Conditions	The basic information form is displayed to the user
Extended Use Cases	N/A

Included Use Cases	Produce initial risk	
Primary Path	User	System
	1. User registers entering the email and the password 3. Enters basic information 4. Press the confirm button	2. Display the basic information form to the user 5. Store gathered data in the database
Alternate Path	The user quits the application after registering without confirming email and resumes the process after confirming email at a different time.	
Exception Path	A network connection is not available The user has entered incorrect data	

Table 22: Use case description for gather basic information use case

4.6.7 Use case description for modify account use case

UC No	**007**	
Priority	Low	
UC Name	Modify Account	
Primary Actor	Existing User	
Secondary Actor	N/A	
Pre - Conditions	The user is found on the database The user has entered basic information The user has entered basic information and viewed initial risk prediction	
Post Conditions	The user information stored in the database is modified. A confirmation message is displayed to the user	
Extended Use Cases	N/A	
Included Use Cases	N/A	
Primary Path	User	System
	1. Select view user profile 3. Select manage profile 5. Select modify profile 7. Modify the profile information 8. Confirm modifications to the profile	2. Display user profile 4. Display profile management options 6. Display editable information fields 9. Modify the user profile information and store the

		modified profile on the database 10. Store the new profile information on the local device storage
Alternate Path	N/A	
Exception Path	The user press cancel button without saving modifications to the profile A network connection is not available	

Table 23: Use case description for modify account use case

4.6.8 Use case description for delete account use case

UC No	**008**	
Priority	Low	
UC Name	Delete account	
Primary Actor	Existing User	
Secondary Actor	N/A	
Pre - Conditions	The user is found on the database The network connection should be available	
Post Conditions	The user is logged out of the account The user account is deleted from the database	
Extended Use Cases	N/A	
Included Use Cases	N/A	
Primary Path	User	System
	1. Select view profile 3. Select delete profile 5. Press confirm button	2. Display user profile 4. Prompt user for confirmation 6. Delete user profile from the database 7. Logout user and redirects user to the register /login page
Alternate Path	N/A	
Exception Path	User press cancel in the profile delete confirmation A network connection is not available	

Table 24: Use case description for delete account use case

4.6.9 Use case description for display recommended article use case

UC No	009	
Priority	Medium	
UC Name	Display recommended articles	
Primary Actor	Existing User	
Secondary Actor	N/A	
Pre - Conditions	The user is found on the database A network connection is available The user has entered basic information and viewed initial prediction	
Post Conditions	Articles recommended according to the user risk factors are displayed to the user	
Extended Use Cases	N/A	
Included Use Cases	N/A	
Primary Path	User	System
	1. The user selects view articles	2. Withdraw user risk factors and breast cancer risk percentage from the database 3. Select the risk factors that affect the person's breast cancer risk 4. Filter articles considering the selected risk factors 5. Generate and display a list of selected articles to the user
Alternate Path	N/A	
Exception Path	The user hasn't viewed the initial prediction, therefore personal risk factors cannot be identified, and article recommendations cannot be generated. The network is disconnected.	

Table 25: Use case description for display recommended article use case

4.6.10 Use case description for display recommended article use case

UC No	010
Priority	Display message
UC Name	Low
Primary Actor	New User

Secondary Actor	N/A	
Pre - Conditions	The user is registered in the system The user has entered basic information and viewed the initial breast cancer risk percentage	
Post Conditions	A customized message is displayed to the user	
Extended Use Cases	N/A	
Included Use Cases	N/A	
Primary Path	User	System
	1. Views initial risk prediction 2. Press OK and confirm after viewing the prediction	3. Generate a personalized message considering the risk factors affecting the person's breast cancer risk 4. Display the message to the user
Alternate Path	N/A	
Exception Path	The user exits the application after viewing the prediction without confirmation	

Table 26: Use case description for display recommended article use case

4.6 Requirement specifications

In the system, functional requirements are identified and prioritized according to the following and prioritized levels.

Priority Level	Description
High	Core requirements of the system
Medium	A necessary but deferrable requirement which makes the system less usable but still functional.
Low	Requirements which are not required for the system functioning.

Table 27: Priority level descriptions

4.7.1 Functional requirements

The below table shows the functional Requirements of the system

No	Requirement Description	Priority
FR1	The user should be able to create an account in the application	High
FR2	The user should be able to login to the application	High
FR3	The user should be able to obtain the initial risk prediction	High
FR4	The user should be able to obtain a weekly report	High
FR5	The user should be able to add his/her basic information to the application	High
FR6	The user should able to add data into tracking up daily	Medium
FR7	The user should be provided with personalized articles collection regarding breast cancer	Medium
FR8	The user should be able to modify his/her account information	Low
FR9	The user should able to delete the account	Low

Table 28: Functional requirements

4.7.2 Non-functional requirements

The below table shows the Non-functional Requirements of the system

No	Requirement	Description
NFR1	Accuracy	The prediction generated by the system should be accurate.
NFR2	Reliability	The system should interact and produce the correct output for user interactions.
NFR3	Performance	The system should take only 3 seconds to generate the outcome.
NFR4	Usability and user-friendliness	Display button for a different selection.
NFR5	Reusability	User can update data
NFR6	Maintainability	Generate monthly survey
NFR7	Security	User should be authenticated, and user information should be encrypted for security.

Table 29: Non-functional requirements

Chapter 05: Design

5.1 Design goals

Design goal	Description
Correctness	The system targets making the machine learning procedure computerized and assists tenderfoots with having the option to perform machine learning. In the event that the system can't fabricate machine learning models, at that point it won't be considered as right. Furthermore, in the event that the beginner designers think that its hard to utilize the system, at that point additionally it won't be considered as right.
Performance	Since in average machine learning, constructing a few ML models takes a ton of time and computational force, having the option to consequently assemble a few models simultaneously will successfully improve execution. Additionally having the option to utilize figuring power proficiently will improve execution.
Scalability	The essential engineering and the primary spine of the system should work extensively well in times when enormous datasets, complex calculations are given. Furthermore, it ought to likewise have the option to include new calculations and preprocessing ventures without numerous changes.
Reusability	The segments of the system ought to be worked in light of seclusion, with the goal that it very well may be reused in augmentations or different improvements
Adaptability	The parts of the system ought to be worked in a manner that can without much of a stretch embrace to new client highlights. The modules ought to be assembled with the end goal that new expansions can be effectively connected deeply item. It additionally ought to be assembled with the end goal that expelling a module from the system doesn't break the whole system.

Table 30: Design goals

5.2 Rich picture

The application is essentially a system which is going to predict the possibility of getting breast cancer. So, the primary input here will be the user's daily lifestyle data which needs to be updated every day to ensure accurate results. User's data will be put through the machine learning process and the results of possibility percentage and suggestion messages will be displayed. The model will be trained with real human lifestyle database.

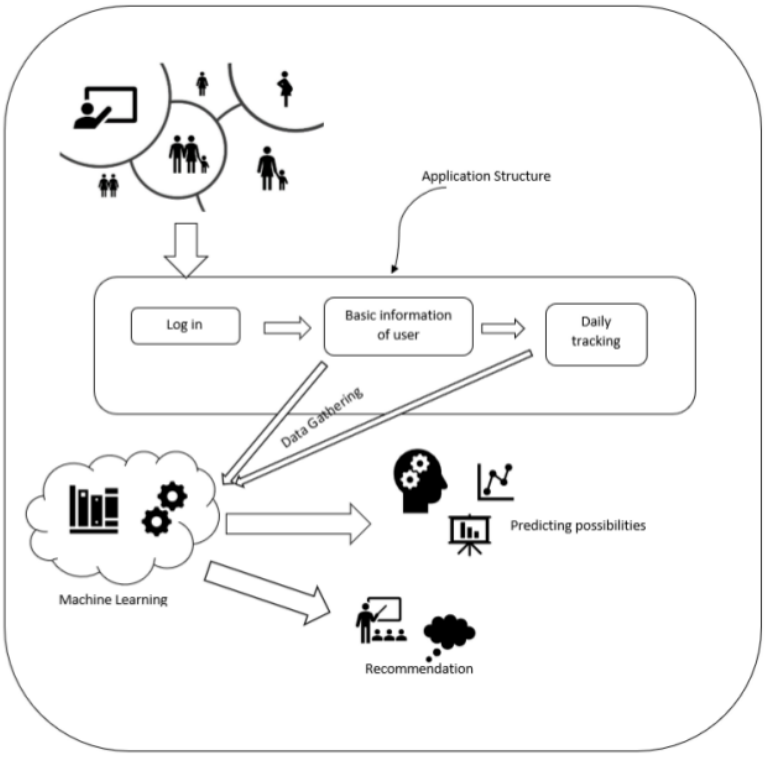


Figure 7: Rich picture diagram

5.3 Domain model

Domain model is used to show the designed structure of the targeted system. Below is the domain model for the Breast Cancer prediction system.

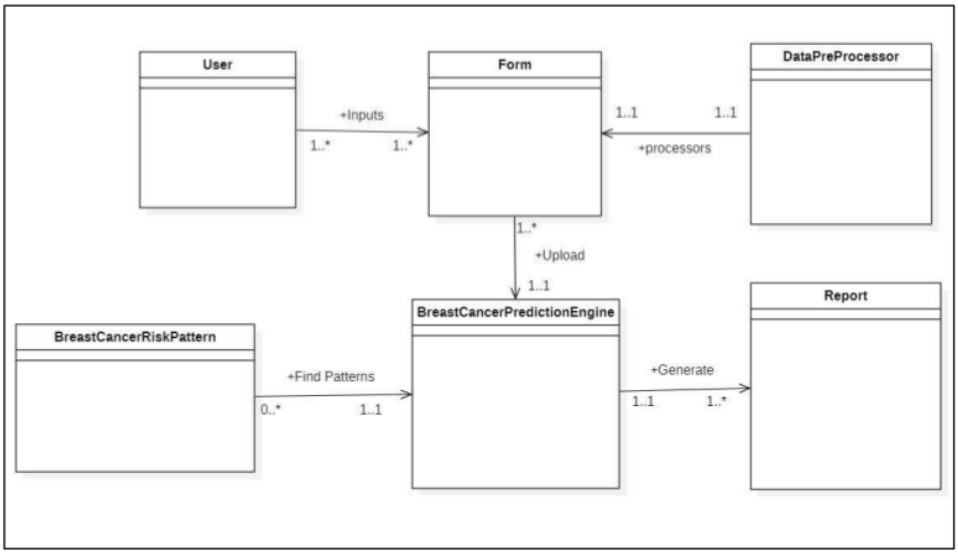


Figure 8: Domain model

5.4 Class diagram

The Class diagram is a static structural diagram which shows the classes of the system, their attributes and operations. Below is the class diagram for the Breast cancer prediction system.

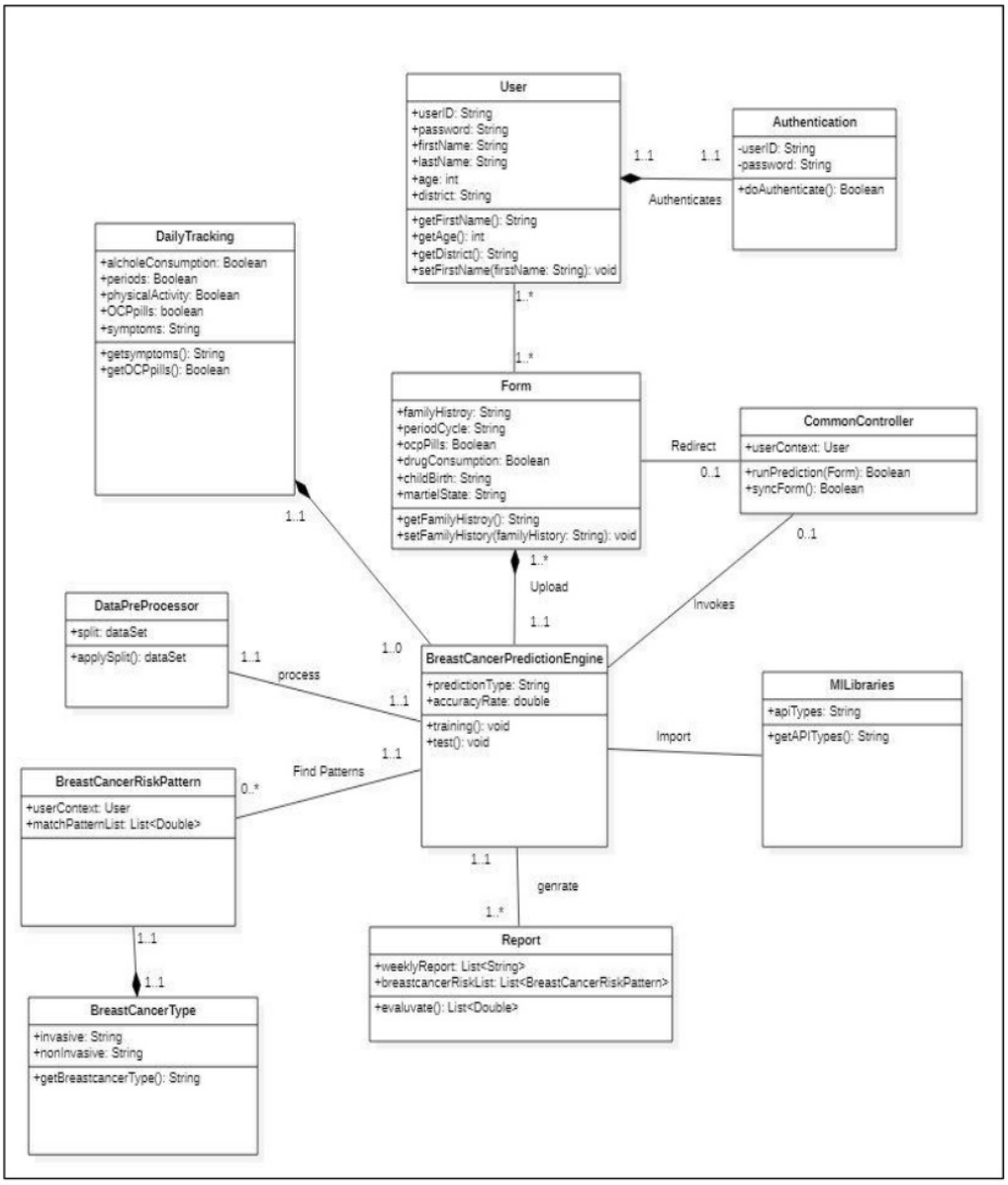


Figure 9: Class diagram

5.5 Activity diagram

The activity diagram is used to show the flow of the functionalities that take place inside a system requirement. The below is the activity diagram for the system.

5.5.1 Activity diagram for the register use case

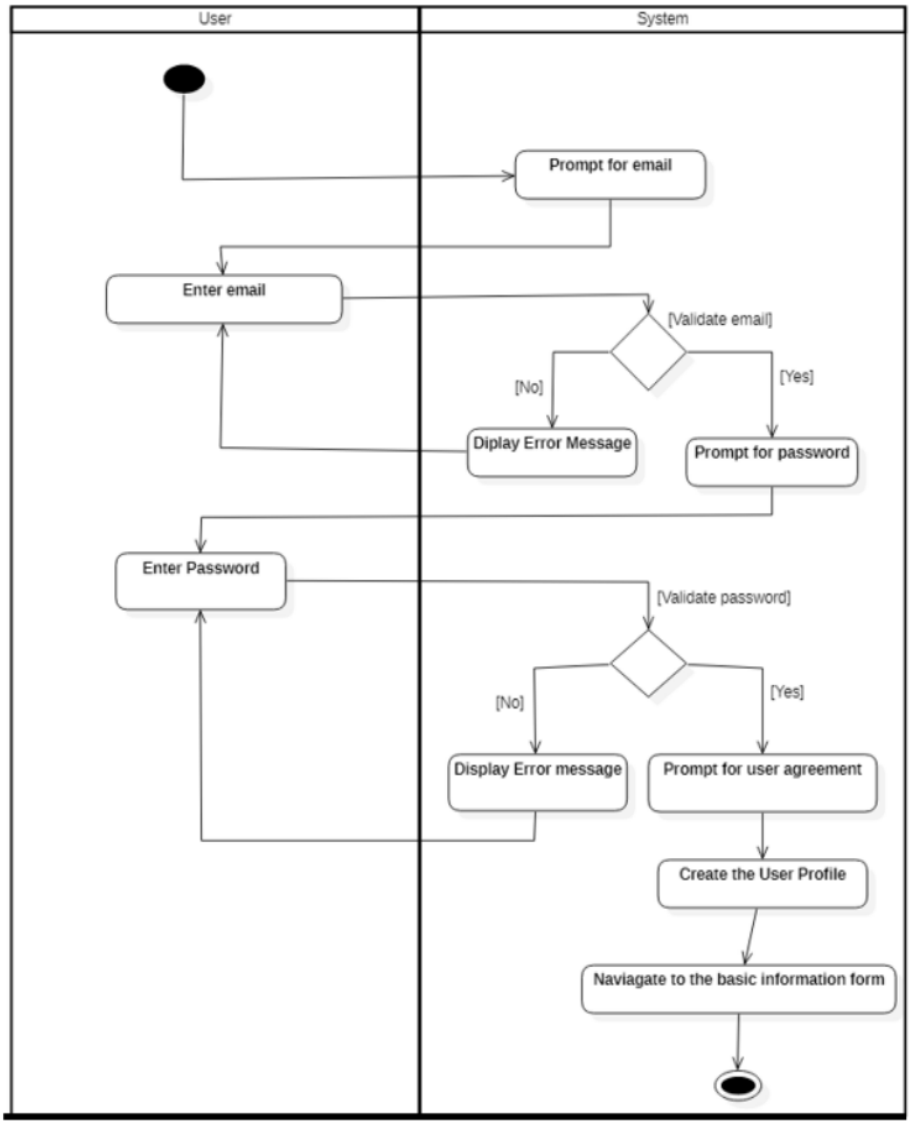


Figure 10: Activity diagram for the register use case

5.5.2 Activity diagram for the Gathering basic information use case

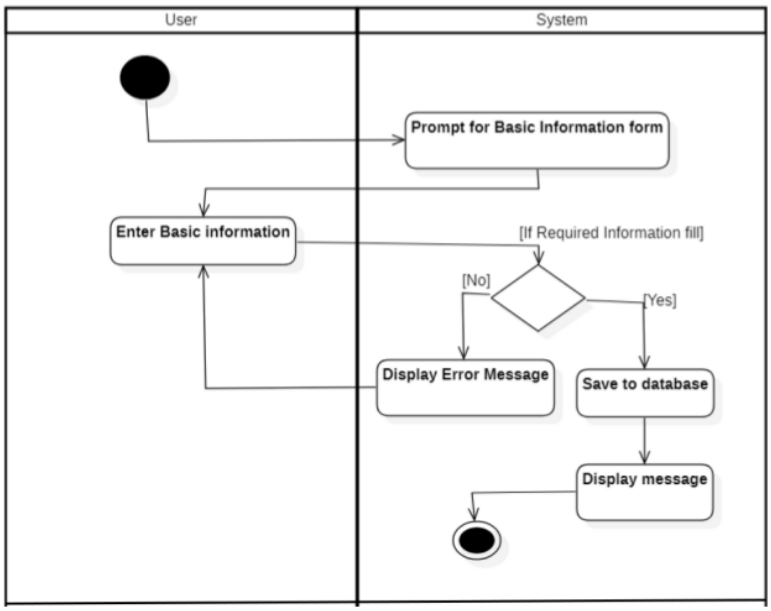


Figure 11: Activity diagram for the Gathering basic information use case

5.5.3 Activity diagram for initial prediction use case

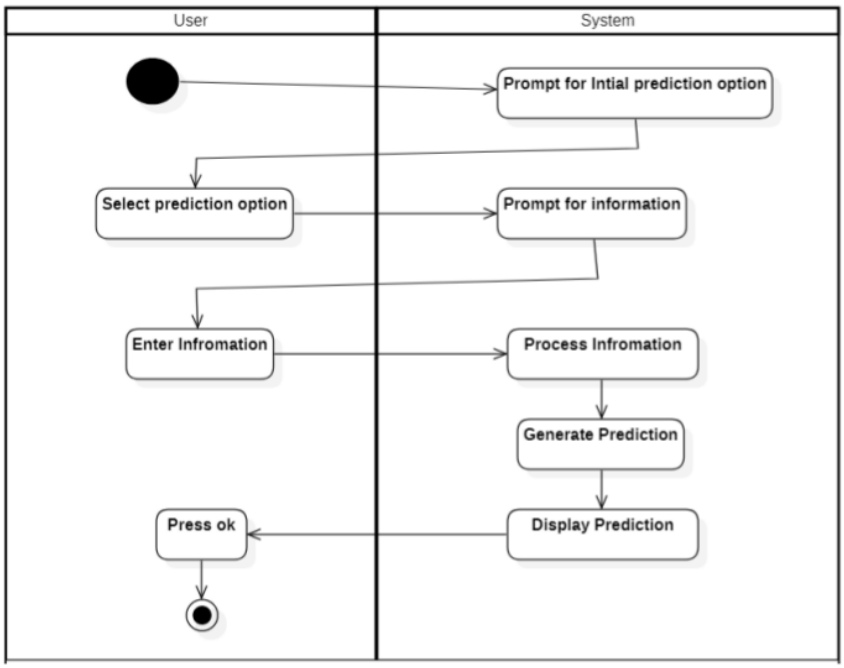


Figure 12: Activity diagram for initial prediction use case

5.5.4 Activity diagram for display personalized message use case

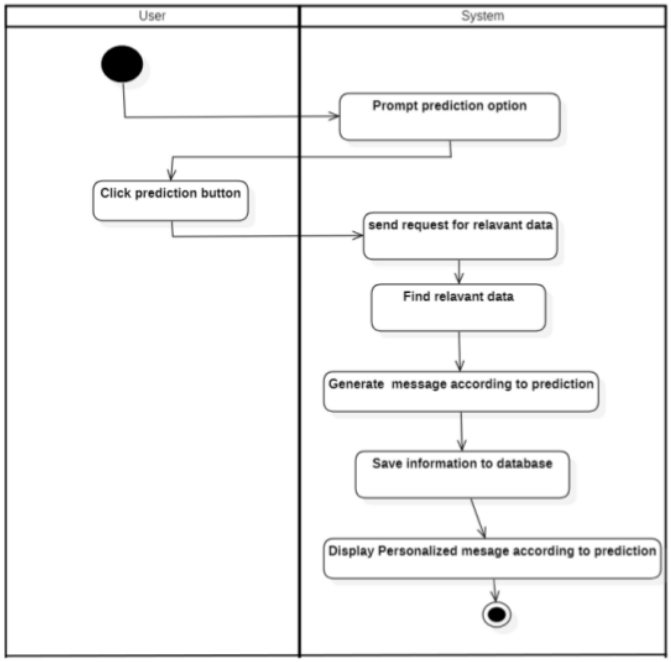


Figure 13: Activity diagram for display personalized message use case

5.5.5 Activity diagram for view article use case

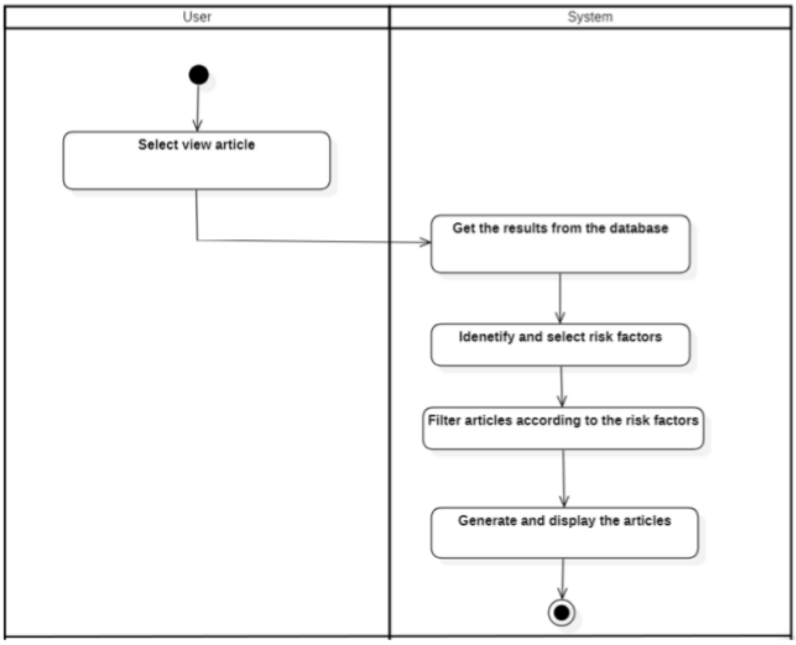


Figure 14: Activity diagram for view article use case

5.5.6 Activity diagram for the login use case

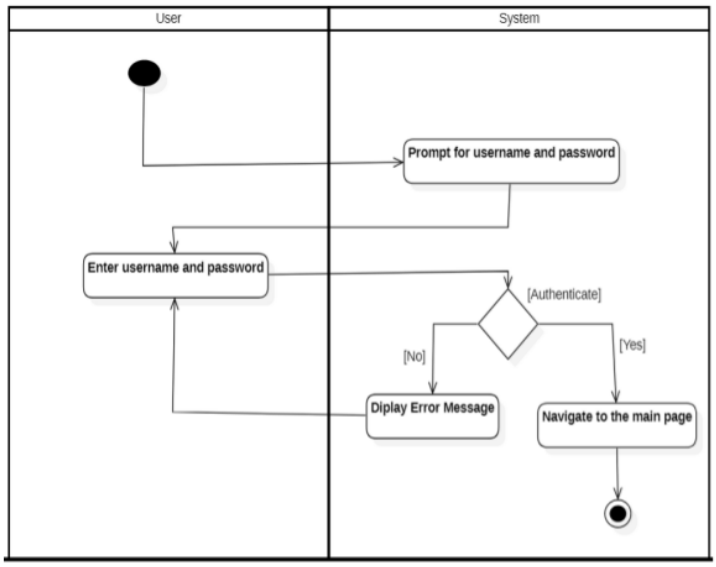


Figure 15: Activity diagram for the login use case

5.5.7 Activity diagram for modify account use case

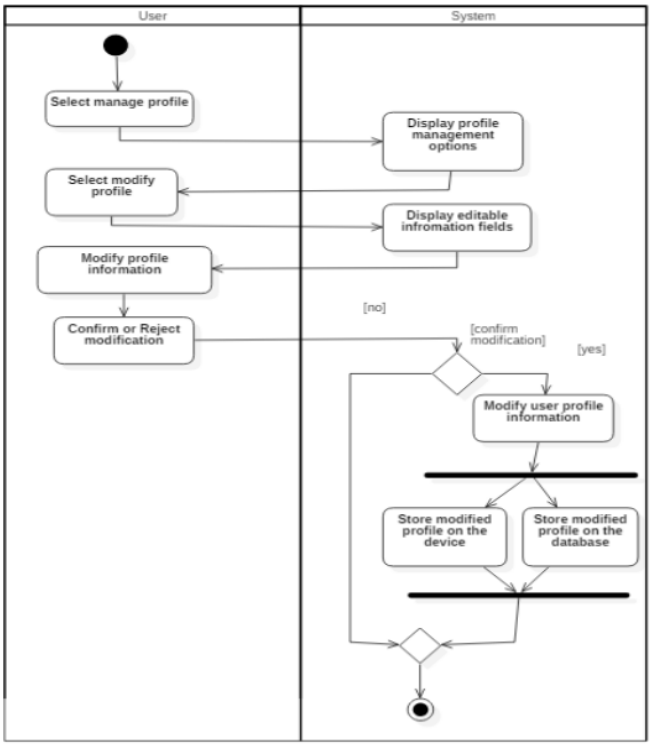


Figure 16: Activity diagram for modify account use case

5.5.8 Activity diagram for delete account use case

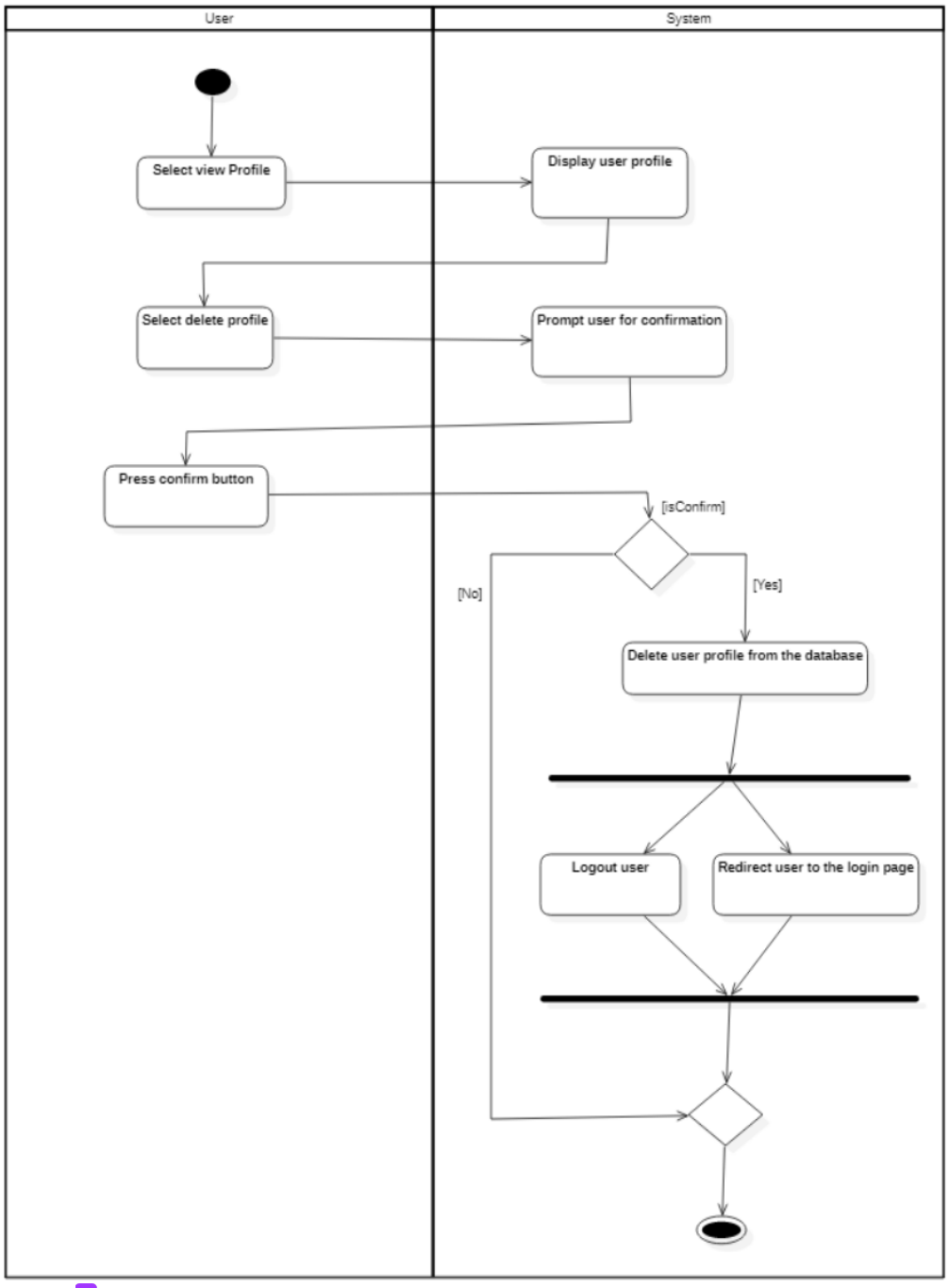
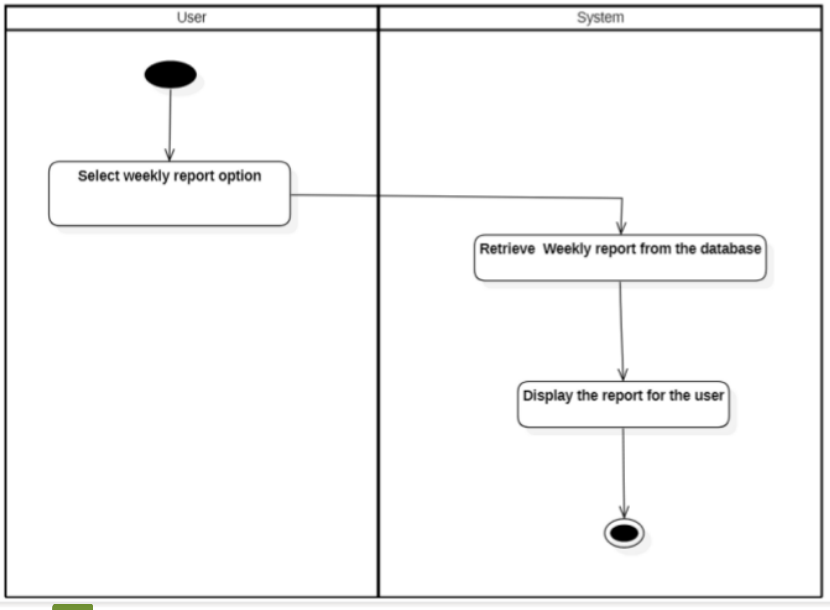


Figure 17: Activity diagram for delete account use case

5.5.9 Activity diagram for produce weekly report use case



190

Figure 18: Activity diagram for produce weekly report use case

3

5.5.10 Activity diagram for enter daily information use case

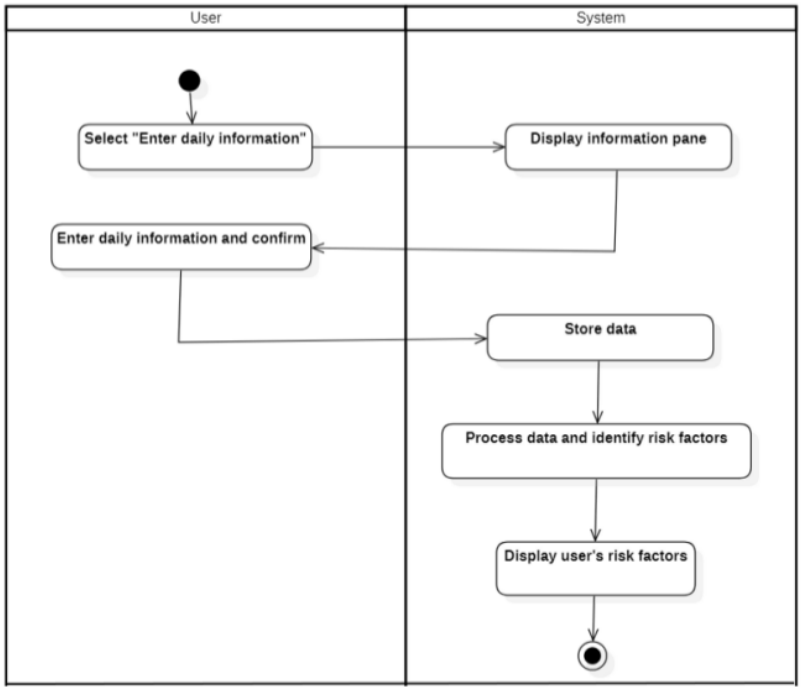


Figure 19: Activity diagram for enter daily information use case

5.6 Sequence diagram

The Sequence diagram for the system is as below which shows the interaction of the system object in a time sequence.

5.6.1 Sequence of registering into the system as a new user

This diagram shows the sequence about signing up into the system as a new user. The user can use a Gmail account to sign up into the system. Alternative paths are also shown in the diagram.

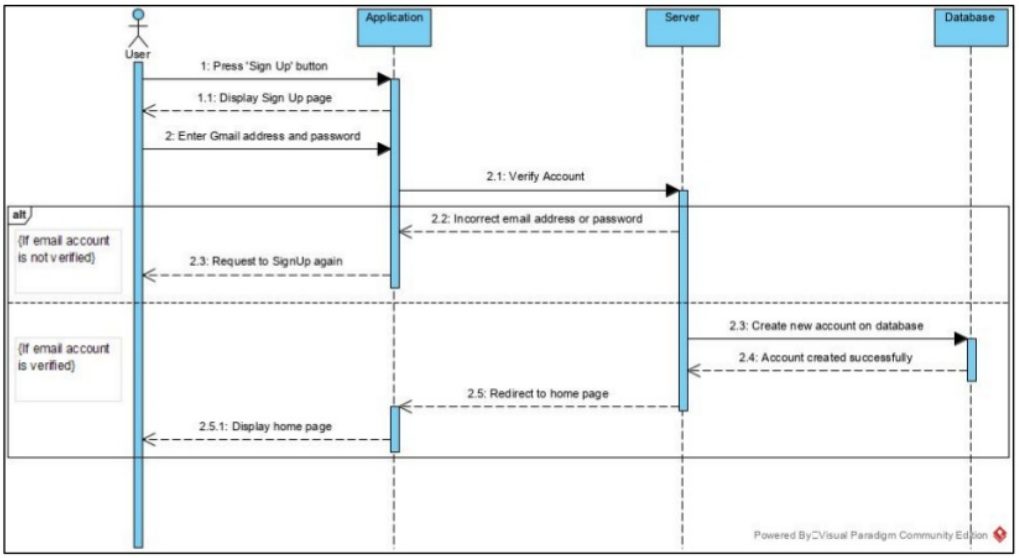


Figure 20: Sequence of registering into the system as a new user

5.6.2 Sequence of gathering initial information

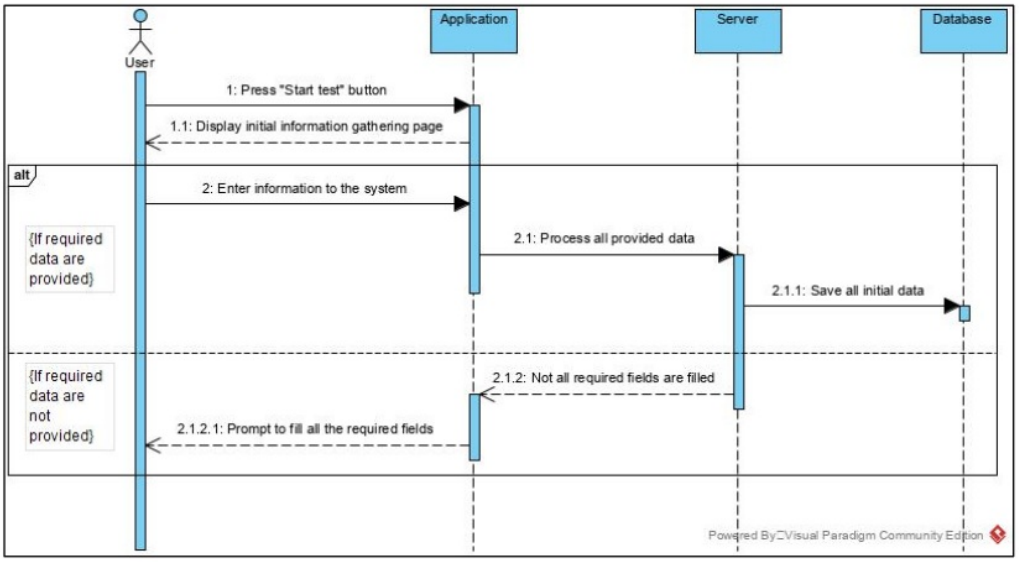


Figure 21: Sequence of gathering initial information

This diagram shows the sequence about gathering initial information from the user after signing up for the initial prediction process. Alternative paths are also shown regarding providing of required data.

5.6.3 Sequence of possibility prediction and displaying results

This diagram shows the flow of how the predicting process happens and displaying results occur in the system after the user entering the initial information

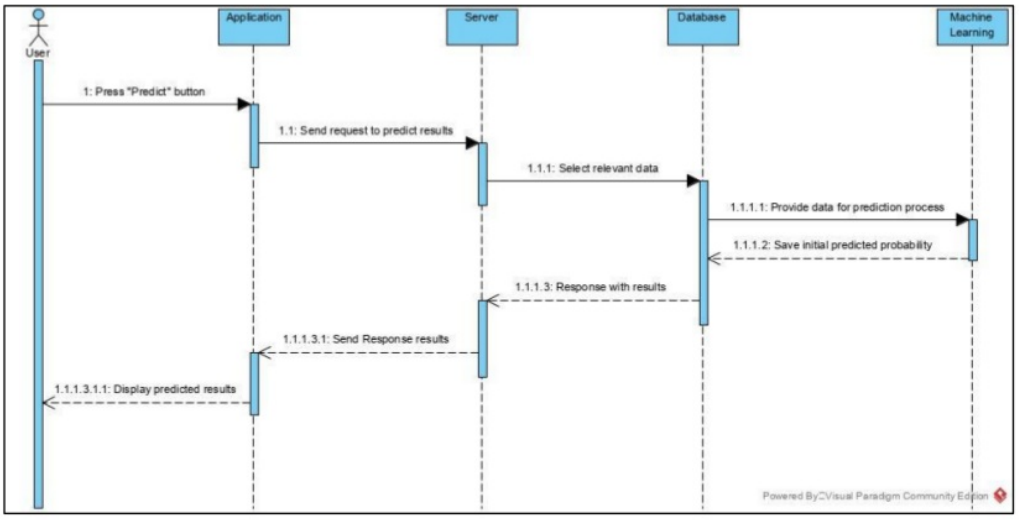


Figure 22: Sequence of possibility prediction and displaying results

5.6.4 Sequence of displaying recommended articles

This diagram shows how the user can read more about the most influencing factor that results in his possibility prediction. Alternative paths are also shown.

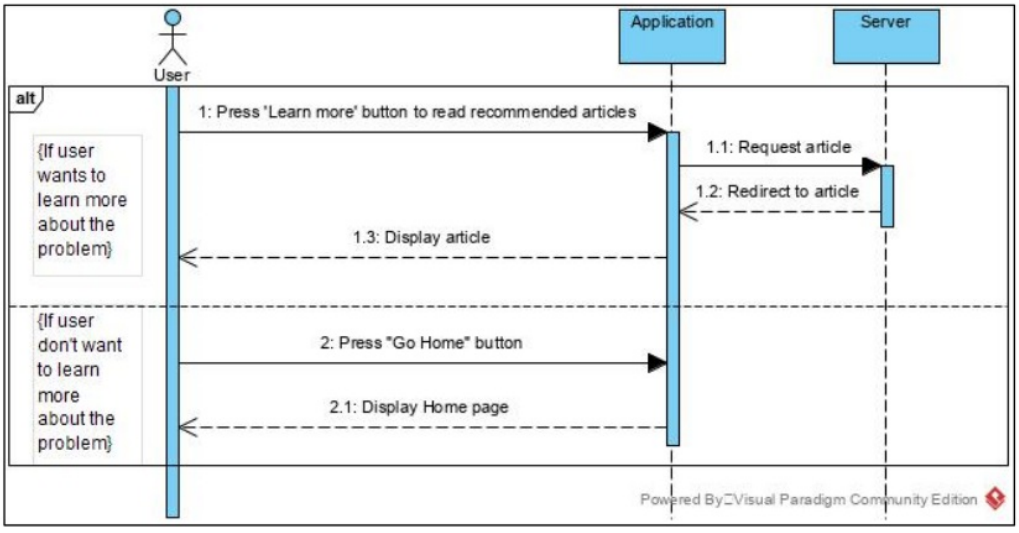


Figure 23: Sequence of displaying recommended articles

5.6.5 Sequence of logging in for an existing user

This diagram shows the flow of how an existing user logs in to the system. This application contains the user's most private and sensitive details. So, every time when the user opens the application it will request the password from the user.

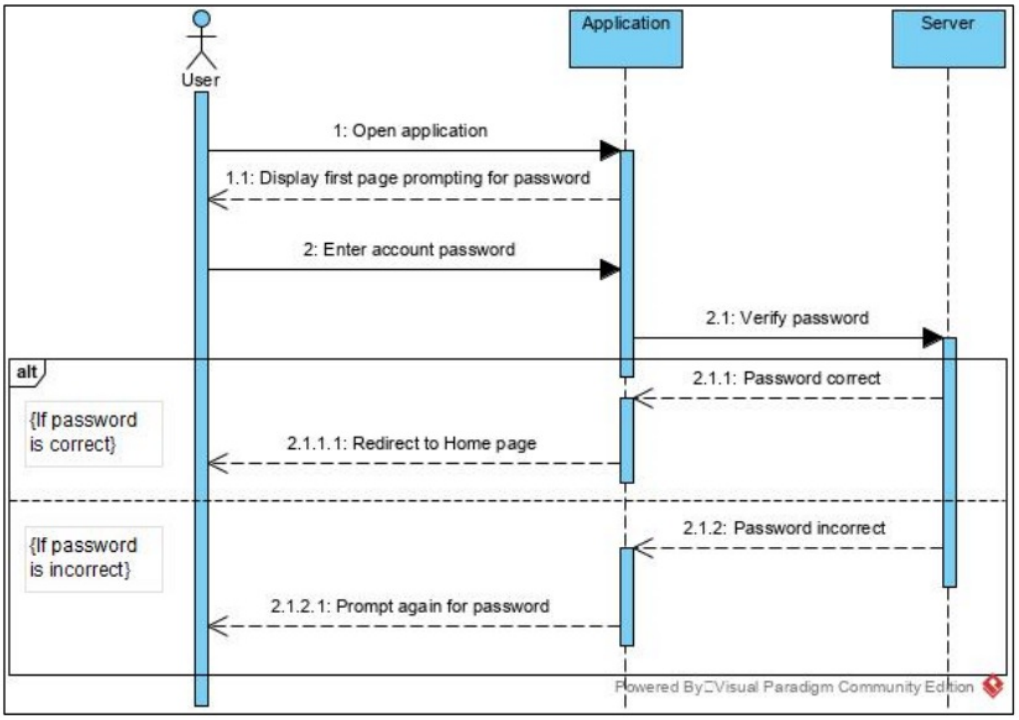


Figure 24: Sequence of logging in for an existing user

5.6.6 Sequence of daily tracking and prediction

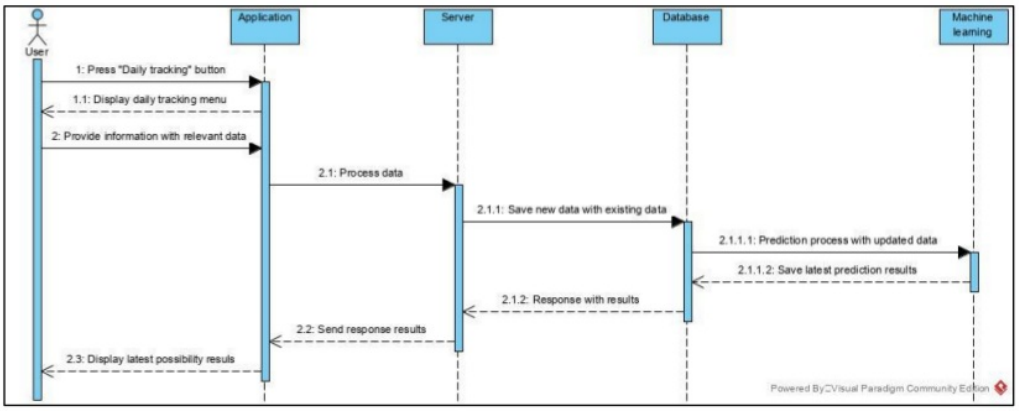


Figure 25: Daily tracking and prediction

This diagram shows the flow of gathering the daily real-time data from the user and saves to the database. Then it will be used to predict the latest possibility percentage.

5.6.7 Sequence of modifying account

This diagram shows the flow about modifying an account. It has 2 main options, modify information and Delete account. Alternate flows are also shown in the diagram.

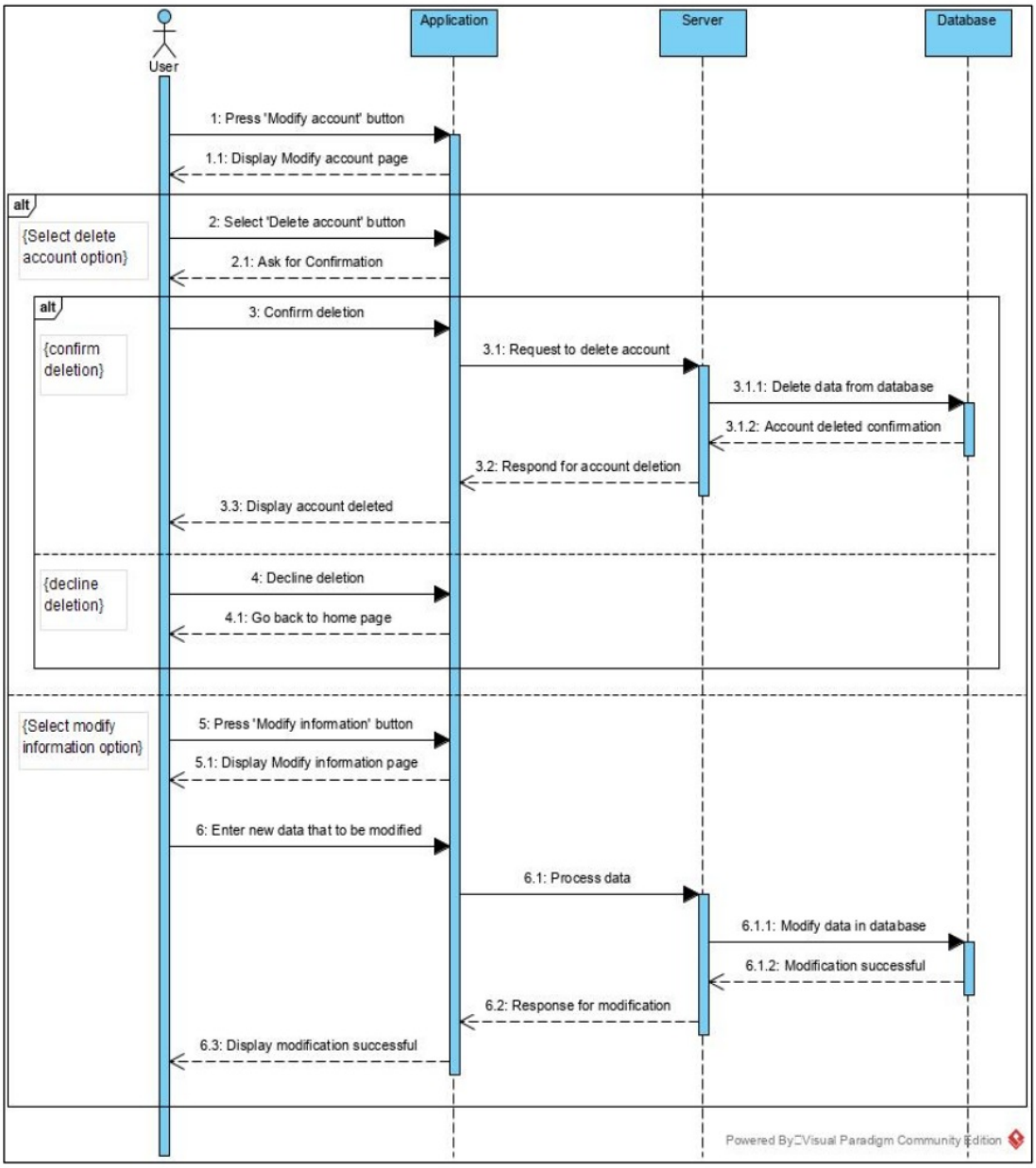


Figure 26: Sequence of modifying account

5.7 High-level system architecture

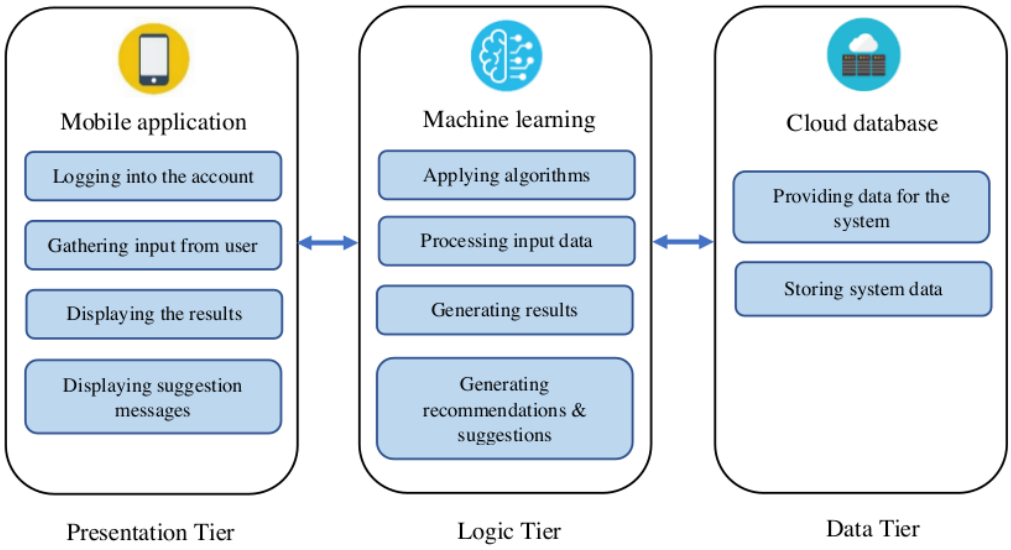


Figure 27: 3-tier architecture

Presentation tier	
Logging in to the account	Every user must have an account and have to be logged in to that account in order to stay connected with the system and keep the prediction updated. When creating the account all the sensitive data that is required to make the prediction will be gathered from the user. The accuracy of the prediction depends on how the user answers the question and updates the system after logging in to the system.
Gathering input from the user	The most important part is gathering data from the user. Since the system is thoroughly depended upon the input obtained from the user the more information you get from the user the more accurate the results going to turn out. The system will ask questions to get values for the parameters in the system and then the prediction will be made based upon the values that were gathered from the user.
Displaying the results	After gathering the data, the prediction part will take place. The user will be expecting to see the correct results after giving out all the details about him/her to the system. After the calculation of the prediction rate, the results will be displayed accordingly.
Displaying suggestion messages	Soon after the system makes the prediction it will take the possibility percentage rate into consideration and decides whether the outcome is positive or negative, if the outcome is negative it will search for the reason why the result came out

	as negative. After identifying the specific weakness or the reason for the results, the system will then suggest things to do on how to make the results more positive.
Logic tier	
Applying algorithms	Applying the algorithm is the most crucial part. It has to be done after identifying the pattern of the user's provided data and real time data of daily life style.
Processing input data	After getting the data, it will go through a certain process to calculate the possibility. This process is programmed only to get the required data which is used to obtain the final results and the data which isn't much of use will be left alone and not taken into consideration. The main process of input data is to calculate the percentage rate. For this process to be a success dataset of cancer patients were taken and after identifying the pattern, the required data on making the prediction was chosen.
Generating results	After processing the data the results will be generated according to the user's input. The logical part of calculating the results will take place first and then after generating the results according to that logic the results will be displayed to the user.
Generating recommendations and suggestions	As mentioned above in the mobile application part the recommendations and suggestions will be generated based upon the specific reason on why the results came out as negative. Based on that factor the system will recommend suggestions on how to overcome that factor.
Data tier	
Providing data for the system	To make the prediction the system will be requiring data about the user, so these required data will be obtained from the database which was created from all the details given by the user. This database must be updated by the user whenever needed otherwise the accuracy of the results will be at a lower level.
Storing system data	There will be a database used to store all the data about the user and it will contain all the up to date details about the user. User will have to manually input all the data about them and the system will store that data in the database and those data will be used in the system to make the prediction and for the recommendations.

Table 31: Description for high level architecture

5.8 UI wireframes

As the system mainly depends on the UIs for the user's input and output, wireframes are designed for the entire system. This application will be a mobile application. The main UI flow starts with a user logging into the system. Google accounts will be used to manage a user's accounts. Inputs will be taken from the users through a questionnaire. Then, the answers will be analyzed and output will be released.

Figure 28: First page

Figure 29: Initial data gathering page

Figure 30: Results and recommendation page

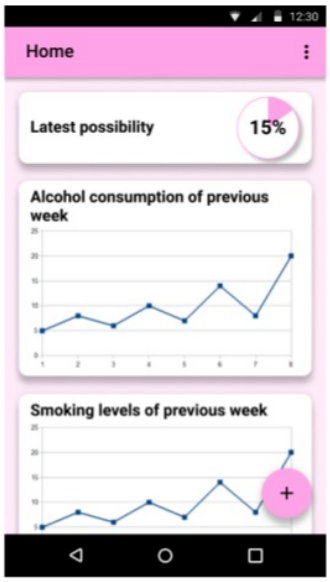


Figure 31: Home page

References

American Cancer Society, 2017-2018. Breast Cancer Facts & Figures. *Cancer.org*, p. 33.

American Society of Clinical Oncology, 2019. *Breast Cancer: Statistics*. [Online]
Available at: https://www.cancer.net/cancer-types/breast-cancer/statistics#:~:targetText=The%20average%205%2Dyear%20survival,are%20diagnosed%20with%20this%20stage.

Anon., 2012. *BreastCancer.org*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/inflammatory/treating
[Accessed 20 10 2019].

Anon., 2013. *Subtypes of ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/subtypes
[Accessed 21 December 2019].

Anon., 2013. *Tests for Diagnosing IDC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc/tests/diagnosing
[Accessed 5 November 2019].

Anon., 2016. *BreastCancer.org*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/pagets/sympt_diag
[Accessed 21 10 2019].

Anon., 2016. *Treatment for ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/treatment
[Accessed 21 December 2019].

Anon., 2018. *Breastcancer.org*. [Online]
Available at: https://www.breastcancer.org/symptoms/types
[Accessed 20 10 2019].

Anon., 2018. *Invasive ductal carcinoma*. [Online]
Available at: https://breastcancernow.org/information-support/facing-breast-cancer/diagnosed-breast-cancer/primary-breast-cancer/invasive-ductal-breast-cancer
[Accessed 5 November 2019].

Anon., 2018. *Invasive Lobular Breast Cancer*. [Online]
Available at: https://breastcancernow.org/information-support/facing-breast-cancer/diagnosed-breast-cancer/primary-breast-cancer/invasive-lobular-breast-cancer
[Accessed 21 December 2019].

Anon., 2018. *Signs and Symptoms of IDC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc/symptoms
[Accessed 02 November 2019].

Anon., 2019. *The American Cancer Society*. [Online]
Available at: https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html#written_by
[Accessed 18 09 2019].

Anon., 2019. *Diagnosis Treatment*. [Online]
Available at: https://www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594
[Accessed 19 11 2019].

Anon., 2019. *Invasive Ductal Carcinoma (IDC)*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc
[Accessed 25 October 2019].

Anon., 2019. *Invasive Lobular Carcinoma (ILC)*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc
[Accessed 20 December 2019].

Anon., 2019. *Keras Documentation*. [Online]
Available at: https://keras.io/

Anon., 2019. *Medscape*. [Online]
Available at: https://www.medscape.com/answers/1945957-180245/what-is-the-gail-model-for-breast-cancer-risk-assessment

Anon., 2019. *pandas: powerful Python data analysis toolkit*. [Online]
Available at: https://pandas.pydata.org/pandas-docs/stable/
[Accessed 29 12 2019].

Anon., 2019. *side-effects*. [Online]
Available at: https://www.cdc.gov/cancer/survivors/patients/side-effects-of-treatment.htm
[Accessed 19 11 2019].

Anon., 2019. *Signs and Symptoms of ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/symptoms
[Accessed 20 December 2019].

Anon., 2019. *tensorflow.org*. [Online]
Available at: https://www.tensorflow.org/federated

Anon., 2019. *Tests for Diagnosing ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/tests/diagnosing
[Accessed 20 December 2019].

Anon., 2019. *Treatment for IDC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc/treatment
[Accessed 23 December 2019].

Anon., n.d. *Invasive Lobular Carcinoma (ILC)*. [Online]
Available at:
https://www.hopkinsmedicine.org/breast_center/breast_cancers_other_conditions/invasive_lobular_carcinoma.html
[Accessed 20 December 2019].

Azmi, M.S.B.M. and Cob, Z.C., 2010. *Breast Cancer Prediction Based On Backpropagation. Algorithm. IEEE Student Conference on Research and Development*.. [Online]
Available at: https://ieeexplore.ieee.org/abstract/document/5703994
[Accessed 20 10 2019].

B Weigelt HM Horlings B Kreike MM Hayes M Hauptmann LFA Wessels D de Jong MJ Van de Vijver LJ Van't Veer JL Peterse, 2008. *Refinement of breast cancer classification by molecular characterization of histological special types*. [Online]
Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/path.2407
[Accessed 10 9 2019].

Breast Cancer Surveillance Consortium, 2016. Breast Cancer Surveillance Consortium Risk Calculator.

Brownlee, J., 2016. *Machine Learning Mastery*. [Online]
Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

cancer, I. A. f. r. o., 2018 September 12. *Latest global cancer data*, Geneva: World health organization.

Ch. Shravya, K. Pravalika, Shaik Subhani, 2019. Prediction of Breast Cancer Using Supervised Machine Learning Techniques. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(6), pp. 1106-1110.

Chand, M., 2019. *Best Programming Language for Machine Learning*.. [Online]
Available at: https://www.c-sharpcorner.com/article/best-programming-language-for-machine-learning/
[Accessed 29 12 2019].

Desai, R., 2018. *Towards data science*. [Online]
Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

Diana Ly1, David Forman2, Jacques Ferlay2, Louise A. Brinton1, and Michael B. Cook1, 2013. An International Comparison of Male and Female Breast Cancer. *NIH Public Access*, 15 04.

Dib, Z. Z. a. H. A., 29th April 2019. *The worldwide female breast cancer incidence and survival, 2018*, Atlanta, GA: American Association for Cancer Research.

Division of Cancer Prevention and Control, 2018. What are the risk factors of Breast Cancer?. 11 9.

Elham Safarzadeh, S. S. S. a. B. B., 2014. Herbal Medicine as Inducers of Apoptosis in Cancer Treatment. *Advanced Pharmaceutical Bulletin*.

Erbil, N. N. D. Ç. İ. a. B. N., 2015. Breast cancer risk assessment using the Gail model: a Turkish study..

Ewertz, M. a. J. A., 2011. Late effects of breast cancer treatment and potentials for rehabilitation.. *Acta Oncologica*, pp. 187-193.

Foote, K., 2019. *A Brief History of Machine Learning*.. [Online]
Available at: https://www.dataversity.net/a-brief-history-of-machine-learning/
[Accessed 28 12 2019].

G. Ball, S. Mian, F. Holding, R.O. Allibone, J. Lowe, S., 2002. *An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid*, s.l.: s.n.

G. Bloom, Yang IV, D. Boulware, K.Y. Kwong, D., 2004. *Multi-Platform, Multi-Site,Microarray-Based Human Tumor Classification*, s.l.: s.n.

GOPAL K. DHONDALAY, DONG L. TONG, GRAHAM R. BALL, 2011. *ESTROGEN RECEPTOR STATUS PREDICTION FOR BREAST CANCER USING ARTIFICIAL NEURAL NETWORK*, Nottingham: IEEE.

Haifeng Wang and Sang Won Yoon, 2015. *Breast Cancer Prediction Using Data Mining Method*. s.l., s.n.

I. Guyon and A. Elisseeff, 2003. *An introduction to variable and feature selection*, s.l.: The Journal of Machine Learning Research .

J. Budczies, W. Weichert, A. Noske, B.M. Müller, C.Weller, T. Wittenberger, H.P. Hofmann, M. Dietel, C.Denkert and V. Gekeler, 2011. *Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays*, s.l.: s.n.

J. M. Moguerza and A. Mu˜noz, 2006. *"Support vector machines with application,* s.l.: IEEE.

Jaime Herndon & Kimberly Holland, 2019. A Comprehensive Guide to Breast Cancer. *Health line*, 21 02.

Johns Hopekins University, n.d. *Invasive Ductal Carcinoma (IDC)*. [Online]
Available at: https://www.hopkinsmedicine.org/breast_center/breast_cancers_other_conditions/invasive_ductal_carcinoma.html
[Accessed 20 December 2019].

Kawthar Al-Ajmi, Artitaya Lophatananon, Martin Yuille, William Ollier, Kenneth R. Muir, 2018. *Review of non-clinical risk models to aid prevention of breast cancer*. [Online]
Available at: https://link.springer.com/article/10.1007%2Fs10552-018-1072-6
[Accessed 10 August 2018].

Kharkovyna, O., 2019. *Python vs R. Choosing the Best Tool for AI, ML & Data Science*. [Online]
Available at: https://medium.com/datadriveninvestor/python-vs-r-choosing-the-best-tool-for-ai-ml-data-science-7e0c2295e243
[Accessed 28 12 2019].

KUNAL JAIN, 2016. *Scikit-learn in Python*. [Online]
Available at: https://www.analyticsvidhya.com/blog/2015/01/scikit-learn-python-machine-learning-tool/
[Accessed 29 12 2019].

L.J. Lancashire, O. Schmid, H. Shah and G. Ball, 2005. *Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis*, s.l.: s.n.

Lekha, A., et al, 2015. *Analysis of Efficiency of Classification and Prediction Algorithms (Naïve Bayes) for Breast Cancer Dataset. International Conference on Emerging Research in Electronics, Computer Science and Technology*.. [Online]
Available at: https://ieeexplore.ieee.org/document/7498997
[Accessed 20 10 2019].

Luashchuk, A., 2019. *Why I Think Python is Perfect for Machine Learning and Artificial Intelligence*. [Online]
Available at: https://towardsdatascience.com/8-reasons-why-python-is-good-for-artificial-intelligence-and-machine-learning-4a23f6bed2e6
[Accessed 29 12 2019].

Max Dieterich, Johannes Stubert, Toralf Reimer, Nicole Erickson, and Anika Berlingb, 2014 Nov 25. Influence of Lifestyle Factors on Breast Cancer Risk. *Breast Care (Basel)*.

Muhammad Sufyian Bin Mohd Azmi, Zaihisma Che Cob, 2018 December 13. *Breast cancer prediction based on backpropagation Algorithm*, Malaysia: IEEE.

Polat, K., 2018. *A Novel ML Approach to Prediction of Breast Cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier*. [Online]
Available at: https://ieeexplore.ieee.org/document/8567245

Rathi, S., 2019. *Top 8 Python Libraries for Machine Learning & Artificial Intelligence*.. [Online]
Available at: https://hackernoon.com/top-8-python-libraries-for-machine-learning-and-artificial-intelligence-y08id3031
[Accessed 29 12 2019].

Runjie Shen∗ Yuanyuan Yang Fengfeng Shao, 2014. *Intelligent Breast Cancer Prediction Model Using Data Mining Techniques*, China: IEEE.

S.Kharya,D.Dubey,S.Soni, 2013. *Predictive Machine Learning Techniques for*. [Online]
Available at: https://pdfs.semanticscholar.org/ade1/c950bd5e56d442ab93edab8635443efbef3c.pdf

Saunders, M. N., Lewis, P. & Thornhill, A., 2019. Research methods for business students. In: *eigth edition*. Harlow: Pearson.

Singhal, P. and Pareek, S., 2018. *Artificial Neural Network for Prediction of Breast Cancer*. *Second International conference on I-SMAC*.. [Online]
Available at: https://ieeexplore.ieee.org/abstract/document/8653700
[Accessed 20 10 2019].

Staff, M. C., 2018. *Invasive Lobular Carcinoma*. [Online]
Available at: https://www.mayoclinic.org/diseases-conditions/invasive-lobular-carcinoma/symptoms-causes/syc-20373973
[Accessed 21 December 2019].

Tao, Z. S. A. L. C. S. T. Z. Z. a. Z. J., 2015. Breast cancer: epidemiology and etiology. *Cell biochemistry and biophysics*, , pp. 333-338.

Taylor, P., 2014. Early stage breast cancer can be cured in most women. 24 6.

Tee, I.C. and Gazala, A.H., 2011. *A Novel Breast Cancer Prediction System*.. [Online]
Available at: //ieeexplore.ieee.org/abstract/document/5946170
[Accessed 20 10 2019].

Vikas Chaurasia, Saurabh Pal, 2014. A Novel Approach for Breast Cancer Detection. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(1).

Wacholder, S. H. P. P. R. G.-C. M. F. H. D. W. T. M. C. D. H. S. K. P. a. R. B., 2010. New England Journal of Medicine. *Performance of common genetic variants in breast-cancer risk models*, pp. 986-993.

World Health Organization, 2019. [Online]
Available at: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/#:~:targetText=Breast%20cancer%20is%20the%20most,all%20cancer%20deaths%20among%20women.

World-Health-Organization, 2018. *WHO - Cancer*. [Online]
Available at: https://www.who.int/news-room/fact-sheets/detail/cancer

Yi-Sheng Sun, Zhao Zhao, Zhang-Nv Yang, Fang Xu, Hang-Jing Lu, Zhi-Yong Zhu, Wen Shi, Jianmin, 2017. *Risk Factors and Preventions of Breast Cancer*. [Online]
Available at: https://www.ijbs.com/v13p1387.htm

Bibliography

American Cancer Society, 2017-2018. Breast Cancer Facts & Figures. *Cancer.org*, p. 33.

American Society of Clinical Oncology, 2019. *Breast Cancer: Statistics*. [Online]
Available at: https://www.cancer.net/cancer-types/breast-cancer/statistics#:~:targetText=The%20average%205%2Dyear%20survival,are%20diagnosed%20with%20this%20stage.

Anon., 2012. *BreastCancer.org*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/inflammatory/treating
[Accessed 20 10 2019].

Anon., 2013. *Subtypes of ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/subtypes
[Accessed 21 December 2019].

Anon., 2013. *Tests for Diagnosing IDC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc/tests/diagnosing
[Accessed 5 November 2019].

Anon., 2016. *BreastCancer.org*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/pagets/sympt_diag
[Accessed 21 10 2019].

Anon., 2016. *Treatment for ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/treatment
[Accessed 21 December 2019].

Anon., 2018. *Breastcancer.org*. [Online]
Available at: https://www.breastcancer.org/symptoms/types
[Accessed 20 10 2019].

Anon., 2018. *Invasive ductal carcinoma*. [Online]
Available at: https://breastcancernow.org/information-support/facing-breast-cancer/diagnosed-breast-cancer/primary-breast-cancer/invasive-ductal-breast-cancer
[Accessed 5 November 2019].

Anon., 2018. *Invasive Lobular Breast Cancer*. [Online]
Available at: https://breastcancernow.org/information-support/facing-breast-cancer/diagnosed-breast-cancer/primary-breast-cancer/invasive-lobular-breast-cancer
[Accessed 21 December 2019].

Anon., 2018. *Signs and Symptoms of IDC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc/symptoms
[Accessed 02 November 2019].

Anon., 2019. *The American Cancer Society*. [Online]
Available at: https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html#written_by
[Accessed 18 09 2019].

Anon., 2019. *Diagnosis Treatment*. [Online]
Available at: https://www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594
[Accessed 19 11 2019].

Anon., 2019. *Invasive Ductal Carcinoma (IDC)*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc
[Accessed 25 October 2019].

Anon., 2019. *Invasive Lobular Carcinoma (ILC)*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc
[Accessed 20 December 2019].

Anon., 2019. *Keras Documentation*. [Online]
Available at: https://keras.io/

Anon., 2019. *Medscape*. [Online]
Available at: https://www.medscape.com/answers/1945957-180245/what-is-the-gail-model-for-breast-cancer-risk-assessment

Anon., 2019. *pandas: powerful Python data analysis toolkit*. [Online]
Available at: https://pandas.pydata.org/pandas-docs/stable/
[Accessed 29 12 2019].

Anon., 2019. *side-effects*. [Online]
Available at: https://www.cdc.gov/cancer/survivors/patients/side-effects-of-treatment.htm
[Accessed 19 11 2019].

Anon., 2019. *Signs and Symptoms of ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/symptoms
[Accessed 20 December 2019].

Anon., 2019. *tensorflow.org*. [Online]
Available at: https://www.tensorflow.org/federated

Anon., 2019. *Tests for Diagnosing ILC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/ilc/tests/diagnosing
[Accessed 20 December 2019].

Anon., 2019. *Treatment for IDC*. [Online]
Available at: https://www.breastcancer.org/symptoms/types/idc/treatment
[Accessed 23 December 2019].

Anon., n.d. *Invasive Lobular Carcinoma (ILC)*. [Online]
Available at:
https://www.hopkinsmedicine.org/breast_center/breast_cancers_other_conditions/invasive_lobular_ca

rcinoma.html

[Accessed 20 December 2019].

Azmi, M.S.B.M. and Cob, Z.C., 2010. *Breast Cancer Prediction Based On Backpropagation. Algorithm. IEEE Student Conference on Research and Development*.. [Online]

Available at: https://ieeexplore.ieee.org/abstract/document/5703994

[Accessed 20 10 2019].

B Weigelt HM Horlings B Kreike MM Hayes M Hauptmann LFA Wessels D de Jong MJ Van de Vijver LJ Van't Veer JL Peterse, 2008. *Refinement of breast cancer classification by molecular characterization of histological special types*. [Online]

Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/path.2407

[Accessed 10 9 2019].

Breast Cancer Surveillance Consortium, 2016. Breast Cancer Surveillance Consortium Risk Calculator.

Brownlee, J., 2016. *Machine Learning Mastery*. [Online]

Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

cancer, I. A. f. r. o., 2018 September 12. *Latest global cancer data,* Geneva: World health organization.

Ch. Shravya, K. Pravalika, Shaik Subhani, 2019. Prediction of Breast Cancer Using Supervised Machine Learning Techniques. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(6), pp. 1106-1110.

Chand, M., 2019. *Best Programming Language for Machine Learning*.. [Online]

Available at: https://www.c-sharpcorner.com/article/best-programming-language-for-machine-learning/

[Accessed 29 12 2019].

Desai, R., 2018. *Towards data science*. [Online]

Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

Diana Ly1, David Forman2, Jacques Ferlay2, Louise A. Brinton1, and Michael B. Cook1, 2013. An International Comparison of Male and Female Breast Cancer. *NIH Public Access,* 15 04.

Dib, Z. Z. a. H. A., 29th April 2019. *The worldwide female breast cancer incidence and survival, 2018,* Atlanta, GA: American Association for Cancer Research.

Division of Cancer Prevention and Control, 2018. What are the risk factors of Breast Cancer?. 11 9.

Elham Safarzadeh, S. S. S. a. B. B., 2014. Herbal Medicine as Inducers of Apoptosis in Cancer Treatment. *Advanced Pharmaceutical Bulletin*.

Erbil, N. N. D. Ç. İ. a. B. N., 2015. Breast cancer risk assessment using the Gail model: a Turkish study..

Ewertz, M. a. J. A., 2011. Late effects of breast cancer treatment and potentials for rehabilitation.. *Acta Oncologica,* pp. 187-193.

Foote, K., 2019. *A Brief History of Machine Learning*.. [Online]
Available at: https://www.dataversity.net/a-brief-history-of-machine-learning/
[Accessed 28 12 2019].

G. Ball, S. Mian, F. Holding, R.O. Allibone, J. Lowe, S., 2002. *An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid,* s.l.: s.n.

G. Bloom, Yang IV, D. Boulware, K.Y. Kwong, D., 2004. *Multi-Platform, Multi-Site,Microarray-Based Human Tumor Classification,* s.l.: s.n.

GOPAL K. DHONDALAY, DONG L. TONG, GRAHAM R. BALL, 2011. *ESTROGEN RECEPTOR STATUS PREDICTION FOR BREAST CANCER USING ARTIFICIAL NEURAL NETWORK*, Nottingham: IEEE.

Haifeng Wang and Sang Won Yoon, 2015. *Breast Cancer Prediction Using Data Mining Method*. s.l., s.n.

I. Guyon and A. Elisseeff, 2003. *An introduction to variable and feature selection,* s.l.: The Journal of Machine Learning Research .

J. Budczies, W. Weichert, A. Noske, B.M. Müller, C.Weller, T. Wittenberger, H.P. Hofmann, M. Dietel, C.Denkert and V. Gekeler, 2011. *Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays,* s.l.: s.n.

J. M. Moguerza and A. Mu˜noz, 2006. *"Support vector machines with application,* s.l.: IEEE.

Jaime Herndon & Kimberly Holland, 2019. A Comprehensive Guide to Breast Cancer. *Health line,* 21 02.

Johns Hopekins University, n.d. *Invasive Ductal Carcinoma (IDC)*. [Online]
Available at:
https://www.hopkinsmedicine.org/breast_center/breast_cancers_other_conditions/invasive_ductal_car

cinoma.html

[Accessed 20 December 2019].

Kawthar Al-Ajmi, Artitaya Lophatananon,· Martin Yuille, William Ollier, Kenneth R. Muir, 2018. *Review of non-clinical risk models to aid prevention of breast cancer*. [Online]
Available at: https://link.springer.com/article/10.1007%2Fs10552-018-1072-6
[Accessed 10 August 2018].

Kharkovyna, O., 2019. *Python vs R. Choosing the Best Tool for AI, ML & Data Science*. [Online]
Available at: https://medium.com/datadriveninvestor/python-vs-r-choosing-the-best-tool-for-ai-ml-data-science-7e0c2295e243
[Accessed 28 12 2019].

KUNAL JAIN, 2016. *Scikit-learn in Python*. [Online]
Available at: https://www.analyticsvidhya.com/blog/2015/01/scikit-learn-python-machine-learning-tool/
[Accessed 29 12 2019].

L.J. Lancashire, O. Schmid, H. Shah and G. Ball, 2005. *Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis*, s.l.: s.n.

Lekha, A., et al, 2015. *Analysis of Efficiency of Classification and Prediction Algorithms (Naïve Bayes) for Breast Cancer Dataset. International Conference on Emerging Research in Electronics, Computer Science and Technology*.. [Online]
Available at: https://ieeexplore.ieee.org/document/7498997
[Accessed 20 10 2019].

Luashchuk, A., 2019. *Why I Think Python is Perfect for Machine Learning and Artificial Intelligence*. [Online]
Available at: https://towardsdatascience.com/8-reasons-why-python-is-good-for-artificial-intelligence-and-machine-learning-4a23f6bed2e6
[Accessed 29 12 2019].

Max Dieterich, Johannes Stubert, Toralf Reimer, Nicole Erickson, and Anika Berlingb, 2014 Nov 25. Influence of Lifestyle Factors on Breast Cancer Risk. *Breast Care (Basel)*.

Muhammad Sufyian Bin Mohd Azmi, Zaihisma Che Cob, 2018 December 13. *Breast cancer prediction based on backpropagation Algorithm*, Malaysia: IEEE.

Polat, K., 2018. *A Novel ML Approach to Prediction of Breast Cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier*. [Online]
Available at: https://ieeexplore.ieee.org/document/8567245

Rathi, S., 2019. *Top 8 Python Libraries for Machine Learning & Artificial Intelligence*.. [Online]
Available at: https://hackernoon.com/top-8-python-libraries-for-machine-learning-and-artificial-intelligence-y08id3031
[Accessed 29 12 2019].

Runjie Shen∗ Yuanyuan Yang Fengfeng Shao, 2014. *Intelligent Breast Cancer Prediction Model Using Data Mining Techniques*, China: IEEE.

S.Kharya,D.Dubey,S.Soni, 2013. *Predictive Machine Learning Techniques for*. [Online]
Available at: https://pdfs.semanticscholar.org/ade1/c950bd5e56d442ab93edab8635443efbef3c.pdf

Saunders, M. N., Lewis, P. & Thornhill, A., 2019. Research methods for business students. In: *eigth edition*. Harlow: Pearson.

Singhal, P. and Pareek, S., 2018. *Artificial Neural Network for Prediction of Breast Cancer*. *Second International conference on I-SMAC*.. [Online]
Available at: https://ieeexplore.ieee.org/abstract/document/8653700
[Accessed 20 10 2019].

Staff, M. C., 2018. *Invasive Lobular Carcinoma*. [Online]
Available at: https://www.mayoclinic.org/diseases-conditions/invasive-lobular-carcinoma/symptoms-causes/syc-20373973
[Accessed 21 December 2019].

Tao, Z. S. A. L. C. S. T. Z. Z. a. Z. J., 2015. Breast cancer: epidemiology and etiology. *Cell biochemistry and biophysics*, , pp. 333-338.

Taylor, P., 2014. Early stage breast cancer can be cured in most women. 24 6.

Tee, I.C. and Gazala, A.H., 2011. *A Novel Breast Cancer Prediction System*.. [Online]
Available at: //ieeexplore.ieee.org/abstract/document/5946170
[Accessed 20 10 2019].

Vikas Chaurasia, Saurabh Pal, 2014. A Novel Approach for Breast Cancer Detection. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(1).

Wacholder, S. H. P. P. R. G.-C. M. F. H. D. W. T. M. C. D. H. S. K. P. a. R. B., 2010. New England Journal of Medicine. *Performance of common genetic variants in breast-cancer risk models*, pp. 986-993.

World Health Organization, 2019. [Online]
Available at: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/#:~:targetText=Breast%20cancer%20is%20the%20most,all%20cancer%20deaths%20among%20women.

World-Health-Organization, 2018. *WHO - Cancer*. [Online]
Available at: https://www.who.int/news-room/fact-sheets/detail/cancer

Yi-Sheng Sun, Zhao Zhao, Zhang-Nv Yang, Fang Xu, Hang-Jing Lu, Zhi-Yong Zhu, Wen Shi, Jianmin, 2017. *Risk Factors and Preventions of Breast Cancer*. [Online]
Available at: https://www.ijbs.com/v13p1387.htm

SDGP 2020 - G-Force - Report Submission

ORIGINALITY REPORT

34%	16%	12%	30%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Westminster Student Paper	8%
2	irep.ntu.ac.uk Internet Source	2%
3	Submitted to South Bank University Student Paper	1%
4	www.ijitee.org Internet Source	1%
5	Submitted to De Montfort University Student Paper	<1%
6	Soonmyung Paik. "A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer", New England Journal of Medicine, 12/30/2004 Publication	<1%
7	Uma Ojha, Savita Goel. "A study on prediction of breast cancer recurrence using data mining techniques", 2017 7th International Conference on Cloud Computing, Data Science &	<1%

	Engineering - Confluence, 2017 Publication	
8	Submitted to Australian College of Nursing Student Paper	<1%
9	Submitted to King's College Student Paper	<1%
10	Submitted to National College of Ireland Student Paper	<1%
11	Submitted to MCC Training Institute Student Paper	<1%
12	Submitted to South Bank University Student Paper	<1%
13	Submitted to University of Dundee Student Paper	<1%
14	Submitted to PSB Academy (ACP eSolutions) Student Paper	<1%
15	Submitted to University of KwaZulu-Natal Student Paper	<1%
16	Submitted to Asia Pacific Instutute of Information Technology Student Paper	<1%
17	Submitted to University of Sydney Student Paper	<1%
18	"Breast Cancer Diagnosis (BCD) Model Using	

	Machine Learning", International Journal of Innovative Technology and Exploring Engineering, 2019 Publication	<1%
19	ijircce.com Internet Source	<1%
20	Lancashire, L.J.. "Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach", Artificial Intelligence In Medicine, 200806 Publication	<1%
21	Submitted to Deakin University Student Paper	<1%
22	massachusettsfashion.org Internet Source	<1%
23	Submitted to University of Cumbria Student Paper	<1%
24	Submitted to Coventry University Student Paper	<1%
25	Submitted to Hong Kong University of Science and Technology Student Paper	<1%
26	Submitted to East Berkshire College Student Paper	<1%

27	Submitted to Oldham Sixth Form College Student Paper	<1%
28	Ir Cath Tee, Ali H. Gazala. "A novel breast cancer prediction system", 2011 International Symposium on Innovations in Intelligent Systems and Applications, 2011 Publication	<1%
29	Shen, Runjie, Yuanyuan Yang, and Fengfeng Shao. "Intelligent Breast Cancer Prediction Model Using Data Mining Techniques", 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, 2014. Publication	<1%
30	Kapil Juneja, Chhavi Rana. "An improved weighted decision tree approach for breast cancer prediction", International Journal of Information Technology, 2018 Publication	<1%
31	www.breastcancercare.org.uk Internet Source	<1%
32	Submitted to Nottingham Trent University Student Paper	<1%
33	Submitted to Imam Muhammad Bin Saud Islamic University Student Paper	<1%

34	Submitted to BENEMERITA UNIVERSIDAD AUTONOMA DE PUEBLA BIBLIOTECA Student Paper	<1%
35	onlinelibrary.wiley.com Internet Source	<1%
36	Submitted to University of Birmingham Student Paper	<1%
37	www.scirp.org Internet Source	<1%
38	Submitted to Botswana Accountancy College Student Paper	<1%
39	Submitted to University of Northumbria at Newcastle Student Paper	<1%
40	www.breastinvestigators.com Internet Source	<1%
41	Submitted to University of Sunderland Student Paper	<1%
42	Submitted to University of Brighton Student Paper	<1%
43	Submitted to Federation University Student Paper	<1%
44	Somayeh Jalayeri, Majid Abdolrazzagh-Nezhad. "Chemical reaction optimization to disease	<1%

	diagnosis by optimizing hyper-planes classifiers", Soft Computing, 2019 Publication	
45	Muhammad Sufyian Bin Mohd Azmi, Zaihisma Che Cob. "Breast Cancer prediction based on Backpropagation Algorithm", 2010 IEEE Student Conference on Research and Development (SCOReD), 2010 Publication	<1%
46	Submitted to CSU, San Jose State University Student Paper	<1%
47	doaj.org Internet Source	<1%
48	ijarcsse.com Internet Source	<1%
49	Submitted to The Hong Kong Polytechnic University Student Paper	<1%
50	ijsrcseit.com Internet Source	<1%
51	advancedbreastcancer.net Internet Source	<1%
52	Submitted to London School of Marketing Student Paper	<1%
53	www.cs.umb.edu	

	Internet Source	<1%
54	Submitted to University of Stellenbosch, South Africa Student Paper	<1%
55	labtestsonline.org Internet Source	<1%
56	Submitted to University College London Student Paper	<1%
57	Submitted to MVJ College of Engineering, Bangalore Student Paper	<1%
58	Submitted to University of Greenwich Student Paper	<1%
59	Submitted to University of Ulster Student Paper	<1%
60	Submitted to Leeds Beckett University Student Paper	<1%
61	Submitted to University of Derby Student Paper	<1%
62	Submitted to University of Leicester Student Paper	<1%
63	www.estij.org Internet Source	<1%

64	orion.math.uwaterloo.ca Internet Source	<1%
65	Submitted to University of Huddersfield Student Paper	<1%
66	mirdsnp.ccr.buffalo.edu Internet Source	<1%
67	Submitted to Anglia Ruskin University Student Paper	<1%
68	Submitted to Grand Canyon University Student Paper	<1%
69	Submitted to Monash University Student Paper	<1%
70	stackoverflow.com Internet Source	<1%
71	Submitted to City University Student Paper	<1%
72	allegrainnovate.com Internet Source	<1%
73	ejournal.litbang.depkes.go.id Internet Source	<1%
74	Rashmi G D, A Lekha, Neelam Bawane. "Analysis of efficiency of classification and prediction algorithms (Naïve Bayes) for Breast Cancer dataset", 2015 International Conference	<1%

	on Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2015 Publication	
75	www.healthnews.ng Internet Source	<1%
76	Submitted to The Robert Gordon University Student Paper	<1%
77	breastcancernow.org Internet Source	<1%
78	ijcoa.com Internet Source	<1%
79	www.irjet.net Internet Source	<1%
80	espace.library.uq.edu.au Internet Source	<1%
81	Submitted to London School of Commerce - Dhaka Student Paper	<1%
82	Submitted to Kaplan University Student Paper	<1%
83	Submitted to Higher Education Commission Pakistan Student Paper	<1%
84	David A. Omondiagbe, Shanmugam Veeramani, Amandeep S. Sidhu. "Machine Learning	<1%

	Classification Techniques for Breast Cancer Diagnosis", IOP Conference Series: Materials Science and Engineering, 2019 Publication	
85	Submitted to CITY College, Affiliated Institute of the University of Sheffield Student Paper	<1%
86	issuu.com Internet Source	<1%
87	docshare.tips Internet Source	<1%
88	Submitted to Illinois Eastern Community Colleges Student Paper	<1%
89	Submitted to Ballyfermot College Student Paper	<1%
90	www.mdpi.com Internet Source	<1%
91	searchenterpriseai.techtarget.com Internet Source	<1%
92	Submitted to Symbiosis International University Student Paper	<1%
93	epdf.tips Internet Source	<1%

94	Submitted to University of Auckland Student Paper	<1%
95	Submitted to 21390 Student Paper	<1%
96	Submitted to University of Medicine and Dentistry of New Jersey Student Paper	<1%
97	link.springer.com Internet Source	<1%
98	Submitted to William T. Dwyer High School Student Paper	<1%
99	Submitted to Prairie State College Student Paper	<1%
100	www.ijbs.com Internet Source	<1%
101	Submitted to Intercollege Student Paper	<1%
102	Submitted to Softwarica College of IT & E-Commerce Student Paper	<1%
103	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
104	Submitted to Segi University College	

	Student Paper	<1%
105	Submitted to Louisiana Tech University Student Paper	<1%
106	www.diva-portal.org Internet Source	<1%
107	Submitted to The British College Student Paper	<1%
108	bfflco.com Internet Source	<1%
109	Submitted to Ngee Ann Polytechnic Student Paper	<1%
110	Submitted to University of Reading Student Paper	<1%
111	"Bio-Inspired Systems: Computational and Ambient Intelligence", Springer Science and Business Media LLC, 2009 Publication	<1%
112	researchopenworld.com Internet Source	<1%
113	Submitted to Liverpool John Moores University Student Paper	<1%
114	www.hindawi.com Internet Source	<1%

115	Cansu Karakas, Anna Biernacka, Tuyen Bui, Aysegul A. Sahin et al. "Cytoplasmic Cyclin E and Phospho–Cyclin-Dependent Kinase 2 Are Biomarkers of Aggressive Breast Cancer", The American Journal of Pathology, 2016 Publication	<1%
116	www.howtocheckforbreastcancer.com Internet Source	<1%
117	Submitted to Universiti Sains Malaysia Student Paper	<1%
118	staroba.my Internet Source	<1%
119	www.stayprivate.com Internet Source	<1%
120	Submitted to Amity University Student Paper	<1%
121	"Abstract", Breast Cancer Research and Treatment, 2006 Publication	<1%
122	Submitted to University of Portsmouth Student Paper	<1%
123	Submitted to University of Dammam Student Paper	<1%
124	Submitted to The Chartered Institute of Marketing	<1%

	Student Paper	
125	Submitted to Universiti Teknologi Malaysia Student Paper	<1%
126	Submitted to Blinn College Student Paper	<1%
127	Submitted to William Paterson University Student Paper	<1%
128	emedicine.medscape.com Internet Source	<1%
129	Submitted to Kingston University Student Paper	<1%
130	dias.library.tuc.gr Internet Source	<1%
131	www.ajol.info Internet Source	<1%
132	djangostars.com Internet Source	<1%
133	Submitted to Georgia Military College Student Paper	<1%
134	Submitted to DeVry University Onlline Student Paper	<1%
135	www.bcsc-research.org Internet Source	<1%

136	koreascience.or.kr Internet Source	<1%
137	Submitted to Buckinghamshire Chilterns University College Student Paper	<1%
138	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
139	www.c-sharpcorner.com Internet Source	<1%
140	Submitted to University of Surrey Student Paper	<1%
141	Submitted to University of Wales Swansea Student Paper	<1%
142	Submitted to AUT University Student Paper	<1%
143	Submitted to Mount Mercy College Student Paper	<1%
144	Submitted to University of Arizona Student Paper	<1%
145	www.tandfonline.com Internet Source	<1%
146	jamanetwork.com Internet Source	<1%

147	Submitted to Regis University Student Paper	<1%
148	www.theseus.fi Internet Source	<1%
149	W McCool. "Breast health care A review", Journal of Nurse-Midwifery, 1998 Publication	<1%
150	Submitted to University of Stirling Student Paper	<1%
151	Submitted to Lovely Professional University Student Paper	<1%
152	Submitted to Hankuk University of Foreign Studies Student Paper	<1%
153	Submitted to University of Southampton Student Paper	<1%
154	Submitted to University of Northampton Student Paper	<1%
155	Submitted to Flinders University Student Paper	<1%
156	breastscreening.cancer.gov Internet Source	<1%
157	"Advances in Intelligent Systems and Computing II", Springer Science and Business	<1%

Media LLC, 2018
Publication

158	Brahma Dathan, Sarnath Ramnath. "Object-Oriented Analysis, Design and Implementation", Springer Science and Business Media LLC, 2015 Publication	<1%
159	Submitted to CTI Education Group Student Paper	<1%
160	www.omicsonline.org Internet Source	<1%
161	Submitted to University of Keele Student Paper	<1%
162	www.antiessays.com Internet Source	<1%
163	Submitted to CVC Nigeria Consortium Student Paper	<1%
164	Submitted to Softwarica College Of IT & E-Commerce Student Paper	<1%
165	Submitted to CSU, Los Angeles Student Paper	<1%
166	Submitted to Tshwane University of Technology Student Paper	<1%

Submitted to Bolton Institute of Higher

167	Education Student Paper	<1%
168	www.medicalnewstoday.com Internet Source	<1%
169	Submitted to Florida International University Student Paper	<1%
170	Submitted to TechKnowledge Student Paper	<1%
171	appliedresearch.cancer.gov Internet Source	<1%
172	Submitted to Cranfield University Student Paper	<1%
173	"Advanced Intelligent Systems for Sustainable Development (AI2SD'2018)", Springer Science and Business Media LLC, 2019 Publication	<1%
174	Submitted to Bournemouth University Student Paper	<1%
175	Submitted to Florida Atlantic University Student Paper	<1%
176	Jennifer S. Haas, Diana L. Miglioretti, Berta Geller, Diana S. M. Buist et al. "Average Household Exposure to Newspaper Coverage about the Harmful Effects of Hormone Therapy	<1%

and Population-Based Declines in Hormone Therapy Use", Journal of General Internal Medicine, 2007
Publication

177	www.appster.org Internet Source	<1%
178	clok.uclan.ac.uk Internet Source	<1%
179	tcr.amegroups.com Internet Source	<1%
180	Submitted to Middlesex University Student Paper	<1%
181	Submitted to Universiti Kebangsaan Malaysia Student Paper	<1%
182	Submitted to CSU, Fullerton Student Paper	<1%
183	Submitted to University of South Florida Student Paper	<1%
184	Submitted to City University of Hong Kong Student Paper	<1%
185	Gopal K. Dhondalay, Dong L. Tong, Graham R. Ball. "Estrogen receptor status prediction for breast cancer using artificial neural network", 2011 International Conference on Machine Learning and Cybernetics, 2011	<1%

Publication

186	Cheng-Har Yip, Robert A. Smith, Benjamin O. Anderson, Anthony B. Miller et al. "Guideline implementation for breast healthcare in low- and middle-income countries", Cancer, 2008 Publication	<1%
187	Submitted to Trident University International Student Paper	<1%
188	Submitted to Chester College of Higher Education Student Paper	<1%
189	Submitted to Queen Mary and Westfield College Student Paper	<1%
190	Submitted to University of Hong Kong Student Paper	<1%
191	Kawthar Al-Ajmi, Artitaya Lophatananon, William Ollier, Kenneth R. Muir. "Risk of breast cancer in the UK biobank female cohort and its relationship to anthropometric and reproductive factors", PLOS ONE, 2018 Publication	<1%
192	Submitted to University of Central England in Birmingham Student Paper	<1%
193	Submitted to Sheffield Hallam University Student Paper	<1%

194	Submitted to University of Hertfordshire Student Paper	<1%
195	Submitted to National University of Ireland, Galway Student Paper	<1%
196	"The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018)", Springer Science and Business Media LLC, 2018 Publication	<1%
197	Submitted to University of Bedfordshire Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		