

**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

[Chapter 01: Introduction 2](#_Toc25655166)

[1.1 Problem Background 2](#_Toc25655167)

[1.1.1 Breast cancer 2](#_Toc25655168)

[1.1.2 Breast cancer and lifestyle 2](#_Toc25655169)

[1.2 Related Work 3](#_Toc25655170)

[1.2.1 Gail model 3](#_Toc25655171)

[1.2.2 Breast Cancer Surveillance Consortium (BCSC) 3](#_Toc25655172)

[1.3 Problem Definition 4](#_Toc25655173)

[1.4 Research Motivation 4](#_Toc25655174)

[1.5 Research Question 4](#_Toc25655175)

[1.6 Research Aim 5](#_Toc25655176)

[1.6.1 Rich Picture Diagram 5](#_Toc25655177)

[1.7 Project Scope 6](#_Toc25655178)

[1.7.1 In-Scope 6](#_Toc25655179)

[1.7.2 Out-Scope 6](#_Toc25655180)

[1.8 Research Objectives 6](#_Toc25655181)

[1.8.1 Project objectives 6](#_Toc25655182)

[1.8.2 Research objectives 7](#_Toc25655183)

[1.9 Resource Requirement 7](#_Toc25655184)

[1.9.1 Software Requirements 7](#_Toc25655185)

[1.9.2 Hardware Requirements 8](#_Toc25655186)

[1.9.3 Data Requirements 8](#_Toc25655187)

[1.9.4 Functional Requirements 8](#_Toc25655188)

[1.9.5 Non-Functional Requirements 8](#_Toc25655189)

[References 9](#_Toc25655190)

[Bibliography 10](#_Toc25655191)

# 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 early stages of Breast Cancer, patients can use fewer radical treatments which causes less side effects as well. (Taylor, 2014)

## Breast cancer

Cancer cells are usually formed from a lump or mass called tumor 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 tumor could be really small. A small painless lump is commonly found as a physical sign. (Amrican Cancer Society, 2017-2018)

## 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. (Medscape, 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 in to 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 life styles.

### 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 the 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 between 35 to 75.
* Cannot calculate the risk of previous diagnosis of breast cancer.
* User should have not under gone 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 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 diagnose 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 (Diagnosis Treatment, 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 lost his or her life. Even after identifying cancer at an early stage the patient will have to go through 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 (side-effects, 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 at 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, 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, patient’s life style 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 medical industry, we decided to analyze the risk factors of 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 user’s lifestyle data using machine learning algorithms?

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

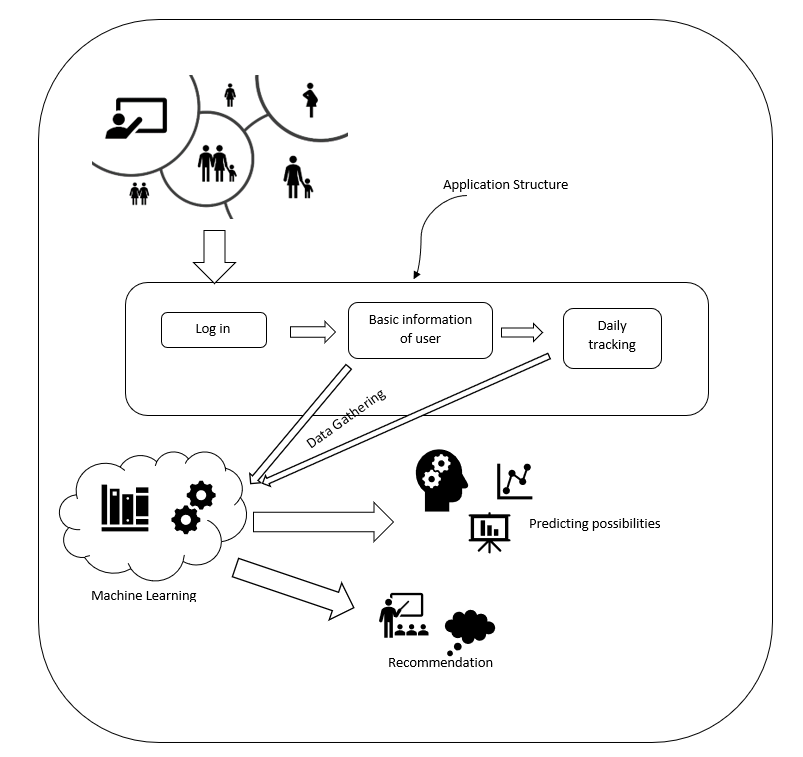
## 1.6 Research Aim

*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 has 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 the public use. The system database will be a hosted in a cloud.

### 1.6.1 Rich Picture Diagram



## 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. 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 in 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 (Jessica Katz, 2019) compared to the women in the age range above 25 years.
* Lifestyle factors that has a considerable effect on breast cancer are considered –In this project the lifestyle factors that has a considerable effect on breast cancer risk such as obesity, alcohol consumption and smoking and many more (Dieterich, Max et al, 2014) 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% (Can., 2019) and the higher survival rate that lies between 75% - 100% (Can., 2019), 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 – Only lifestyle of Sri Lankan women is considered for this project as lifestyles varies 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 a 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 the 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 fulfills the sole purpose with higher accuracy. |

## 1.8.2 Research objectives

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

## 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 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 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 an easiest way to keep backup of our project related source codes too.
* MS Office – To create project related documents like reports, and documentations.
* 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

### 1.9.4 Functional Requirements

|  |  |  |
| --- | --- | --- |
| No | Requirement Description | Priority |
| FR1 | The user should be able to add information into application | High |
| FR2 | The user should able to add data in to tracking up daily | High |
| FR3 | The user should able to create account in application | High |
| FR4 | The user should able to delete account | High |
| FR5 | The user should able to be checking information which includes breast cancer prevention information. | High |

### 1.9.5 Non-Functional Requirements

|  |  |  |
| --- | --- | --- |
| No | Requirement | Description |
| NFR1 | Accuracy | Generated prediction by system should be accurate. |
| NFR2 | Reliability | When the user selects any option system should surely display output according to option. |
| NFR3 | Performance | The system should take only 3 second to generate 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 | Username and password authentications, Data should be encrypted and Decrypted. |

# References

American Society of Clinical Oncology. (2019, 07). *Breast Cancer: Statistics*. Retrieved from Cancer.Net: https://www.cancer.net/cancer-types/breast-cancer/statistics#:~:targetText=The%20average%205%2Dyear%20survival,are%20diagnosed%20with%20this%20stage.

Amrican Cancer Society. (2017-2018). Breast Cancer Facts & Figures. *Cancer.org*, 33.

Breast Cancer Surveillance Consortium. (2016). Breast Cancer Surveillance Consortium Risk Calculator.

cancer, I. A. (2018 September 12). *Latest global cancer data.* Geneva: World health organization.

*Diagnosis Treatment*. (2019). Retrieved 11 19, 2019, from Mayoclinic: https://www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594

Dib, Z. Z. (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, 9 11). What are the risk factors of Breast Cancer?

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

Erbil, N. N. (2015). Breast cancer risk assessment using the Gail model: a Turkish study.

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

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)*.

*Medscape*. (2019, 11 19). Retrieved from what-is-the-gail-model-for-breast-cancer-risk-assessment: https://www.medscape.com/answers/1945957-180245/what-is-the-gail-model-for-breast-cancer-risk-assessment

Muhammad Sufyian Bin Mohd Azmi, Zaihisma Che Cob. (2018 December 13). *Breast cancer prediction based on backpropagation Algorithm.* Malaysia: IEEE. Retrieved from IEEE Xplore Digital Library.

*side-effects*. (2019, 02 05). Retrieved 11 19, 2019, from Centers for disease control and prevention: https://www.cdc.gov/cancer/survivors/patients/side-effects-of-treatment.htm

Tao, Z. S. (2015). Breast cancer: epidemiology and etiology. *Cell biochemistry and biophysics,* , 333-338.

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

Wacholder, S. H.-C. (2010). New England Journal of Medicine. *Performance of common genetic variants in breast-cancer risk models*, 986-993.

World Health Organization. (2019). Retrieved from World Health Organization: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/#:~:targetText=Breast%20cancer%20is%20the%20most,all%20cancer%20deaths%20among%20women.

# Bibliography

American Society of Clinical Oncology. (2019, 07). *Breast Cancer: Statistics*. Retrieved from Cancer.Net: https://www.cancer.net/cancer-types/breast-cancer/statistics#:~:targetText=The%20average%205%2Dyear%20survival,are%20diagnosed%20with%20this%20stage.

Amrican Cancer Society. (2017-2018). Breast Cancer Facts & Figures. *Cancer.org*, 33.

Breast Cancer Surveillance Consortium. (2016). Breast Cancer Surveillance Consortium Risk Calculator.

cancer, I. A. (2018 September 12). *Latest global cancer data.* Geneva: World health organization.

*Diagnosis Treatment*. (2019). Retrieved 11 19, 2019, from Mayoclinic: https://www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594

Dib, Z. Z. (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, 9 11). What are the risk factors of Breast Cancer?

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

Erbil, N. N. (2015). Breast cancer risk assessment using the Gail model: a Turkish study.

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

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)*.

*Medscape*. (2019, 11 19). Retrieved from what-is-the-gail-model-for-breast-cancer-risk-assessment: https://www.medscape.com/answers/1945957-180245/what-is-the-gail-model-for-breast-cancer-risk-assessment

Muhammad Sufyian Bin Mohd Azmi, Zaihisma Che Cob. (2018 December 13). *Breast cancer prediction based on backpropagation Algorithm.* Malaysia: IEEE. Retrieved from IEEE Xplore Digital Library.

*side-effects*. (2019, 02 05). Retrieved 11 19, 2019, from Centers for disease control and prevention: https://www.cdc.gov/cancer/survivors/patients/side-effects-of-treatment.htm

Tao, Z. S. (2015). Breast cancer: epidemiology and etiology. *Cell biochemistry and biophysics,* , 333-338.

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

Wacholder, S. H.-C. (2010). New England Journal of Medicine. *Performance of common genetic variants in breast-cancer risk models*, 986-993.

World Health Organization. (2019). Retrieved from World Health Organization: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/#:~:targetText=Breast%20cancer%20is%20the%20most,all%20cancer%20deaths%20among%20women.