

Apply Machine learning approach to predict the Breast Cancer Tumor Type

BY

Tanjed Ahmed khan
ID: 161-15-6849

Md. Shahidul Islam
ID: 161-15-7079

S.M Aman Ullah
ID: 161-15-6875

This Report Presented in Partial Fulfillment of the Requirements for
The Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Zerin Nasrin Tumpa

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Dewan Mamun Raza

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
NOVEMBER 2019

APPROVAL

This Project/internship titled **Apply Machine learning approach to predict the Breast Cancer Tumor type**, submitted by Tanjed Ahmed Khan, ID No: 161-15-6849, Md. Shahidul Islam, ID No: 161-15-7079, S.M Aman Ullah, ID No: 161-15-6875 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 07 December 2019.

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Abdus Sattar
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Farah Sharmin
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



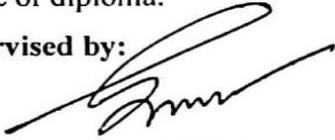
Dr. Md. Saddam Hossain
Assistant Professor
Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

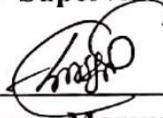
We hereby declare that, this project has been done by us under the supervision of **Zerin Nasrin Tumpa, Lecturer, Department of CSE Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Co- Supervised by:



Dewan Mamun Raza
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Tanjed Ahmed Khan

Tanjed Ahmed Khan
ID: -161-15-6849
Department of CSE
Daffodil International University

Md. Shahidul Islam

Md. Shahidul Islam
ID: -161-15-7079
Department of CSE
Daffodil International University

S.M. Aman Ullah

S.M Aman Ullah
ID: -161-15-6875
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

This work would not have been possible without the support of our almighty Allah. We are especially indebted to our madam Ms. Zerin Nasrin Tumpa, Lecturer of Department of Computer Science and Engineering at Daffodil International University (DIU), who have been supportive of our career goals and who worked actively to provide us with the protected academic time to pursue those goals. As Our teacher and mentor, she has taught us more than we could ever give her credit for here. She has shown us, by her example, what a good scientist (and person) should be.

We are grateful to all of those with whom we have had the pleasure to work during this and other related research or projects. Each of the members of Dissertation team who has provided us extensive personal and professional guidance and taught us a great deal about both scientific research project and also in general life. We would especially like to thank A.B.M. Khorshed Alam the chairman of Bangladesh Breast Cancer Awareness Society (BBCAS). As a specialist and an expert he gives us many information and give us an opportunity to discuss with him from his very busy time.

While undertaking this research project, nobody was more important to us than our family members. We would like to thank our parents in whatever we do, whose support and encouragement is with us. They are the eternal role models and they give us the inspiration that never ends.

ABSTRACT

Maximum people around the world use mobile phones as well as those in under resourced communities. This availability of the mobile phone creates a huge field of mobile health education technology which can gift a promising mechanism for rising cancer hindrance. Recent years, breast cancer has become one of the leading causes of female death in the over populated countries and not only over populated countries but also around the world. Moreover, the occurrence of breast cancer is increasing in the early age groups in this Asian area compared with the western countries. Hence, it's high time to raise awareness to prevent this cancer or detect it in the early stage to decrease the female death rate. At present, machine learning is delivering exciting outcomes in the health care sector. If it can be applied to breast cancer detection and if it is possible to publish it via web application and mobile application then the health care sector will go a step further. In order to do so we propose a Web application and Mobile application which will help to predict breast cancer tumor type by using algorithmic result and will also provide suggestion regarding precautions and to enrich awareness.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	I
Declaration	II
Acknowledgement	III
Abstract	IV
CHAPTER	
CHAPTER 1: INTRODUCTION	1-4
1.1 Introduction	1
1.2 Motivation	1
1.3 Rationale of the Study	2
1.4 Research Questions	2
1.5 Expected Output	3
1.6 Report Layout	3-4
CHAPTER 2: BACKGROUND	5-11
2.1 Introduction	5
2.2 Related Works	5-8
2.3 Research Summary	8
2.3.1 Risk factors	9
2.3.2 Stages of Breast Cancer	10
2.4 Scope of the Problem	11
2.5 Challenges	11

CHAPTER 3: RESEARCH METHODOLOGY	12-20
3.1 Introduction	12
3.1.1 Breast cancer from survey	13
3.1.2 Use case Diagram	15-16
3.2 Research Subject and Instrumentation	17
3.2.1 Machine Learning Intelligence Library	17
3.2.2 Techniques	17
3.3 Data Collection Procedure	18
3.4 Statistical Analysis	19
3.4.1 Data preprocessing	19
3.4.2 Using K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) algorithm	19
3.4.2.1 K-Nearest Neighbors (KNN)	19
3.4.2.2 Support Vector Machine (SVM)	19
3.5 Implementation Requirements	20
CHAPTER 4: EXPERIMENTAL RESULT AND DISCUSSION	21-30
4.1 Introduction	21
4.2 Experimental Results	21
4.3 Descriptive Analysis	21
4.3.1 Training, Testing and the Validation of the model	21
4.3.2 Model performance	21-23
4.3.3 Machine Learning to Web application	23
4.3.3.1 Web application Home page	24
4.3.3.2 About Breast cancer	24

4.3.3.3 Web application Scholars' Quote page	25
4.3.3.4 Web application Knowledge section	25
4.3.3.5 Web application Registration page	26
4.3.3.6 Web application Login page	26
4.3.3.7 Web application Screening page	27
4.3.4 Machine Learning to Mobile Application	28
4.3.4.1 Mobile Application Sign In page	28
4.3.4.2 Mobile Application Sign Up page	28
4.3.4.3 Mobile Application Screening test page	29
4.3.4.4 Mobile Application Result of Screening	30
4.4 Summary	30

CHAPTER 5: SUMMERY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH 31

5.1 Summary of the Study	31
5.2 Conclusions	31
5.3 Recommendations	31
5.4 Implication for Further Study	31

REFERENCES 32-33

APPENDICES

LIST OF FIGURES

FIGURES	PAGE NO
Figure 2.1: Breast Cancer Risk factors chart	09
Figure 2.2: Common Risk factor	09
Figure 2.3: Ages for Breast Cancer chart	10
Figure 3.1: Block diagram for the propounded methodology	12
Figure 3.2: Participants Countries	13
Figure 3.3: Participants Gender	13
Figure 3.4: Participants Occupation	14
Figure 3.5: Invaded Participants Breast Cancer Discover Chart	14
Figure 3.6: Awareness of Breast Cancer Symptoms Chart	15
Figure 3.7: Diagram for general user	15
Figure 3.8: Diagram for cancer patient	16
Figure 3.9: K-Nearest Neighbors (KNN) classification for Breast cancer	17
Figure 3.10: Support Vector Machine (SVMs) Topology	18
Figure 3.11: Three major part of Methodology	20
Figure 4.1: Confusion Matrix for Breast cancer using SVM classifier	23
Figure 4.2: Weighted average performance parameters of SVM algorithm for dataset	23
Figure 4.3: Web application Home page	24
Figure 4.4: About breast cancer	24
Figure 4.5: Web application Scholars' Quote page	25
Figure 4.6: Web application Knowledge section	25
Figure 4.7: Web application Registration page	26

Figure 4.8: Web application Login page	26
Figure 4.9: Unregister user	27
Figure 4.10: Web application Screening page	27
Figure 4.11: Mobile Application Sign In page	28
Figure 4.12: Mobile Application Sign Up page	28
Figure 4.13: Do's or Don'ts	29
Figure 4.14: Enter Age	29
Figure 4.15: Choose the shape	29
Figure 4.16: Choose the Margin	29
Figure 4.17: BI-RADS Assesments	29
Figure 4.18: Choose the Density	29
Figure 4.19: Mobile application Information for doing Screening test (from mammogram report)	29
Figure 4.20: Submit Information	30
Figure 4.21: Result view	30
Figure 4.22: Mobile application Result of screening test	30

LIST OF TABLES

TABLES	PAGE NO
Table 1: Patient's Stages	10

CHAPTER 1

INTRODUCTION

1.1 Introduction

Breast cancer is one of the most common malignancy in female. Day by day it's going in terrible condition along with Prostate cancer, pancreatic cancer, Colon cancer, Bronchus cancer and others also [1]. It is found that younger age group in Bangladesh are suffering most by the carcinoma of breast comparing with the western countries [2]. The burden on the developing nations rising day by day due to increase in the life expectancy and lifestyle changes as well. In NICRH (National Institute of Cancer Research & Hospital), breast cancer is the widespread among the female patients which is the most common cancer as a whole and Bangladesh is also in the reach of its hand [2]. In Bangladesh it is revealed that 22.5 per 100000 women develop breast cancer during her entire lifetime having the death rate of 0.98% which is 137th globally [3]. Hence, if we can detect breast cancer in early stage then it is possible to decrease the female death in Bangladesh. At few decades, the exploration of Data science and Machine Learning Techniques are approaches in medical fields or healthcare systems. Machine Learning approaches great performances for correct diagnosis of breast cancer and classifies the patients and detect them into Malignant [4] or Benign [4] groups. In essential aspects detection from complicated breast cancer datasets, Machine Learning [5] is an extensively recognized as the methodology which is preference in Breast Cancer pattern classification and Forecast Modelling. Almost 40% of Machine Learning studies on breast cancer prediction had been centered on predicting patient survivability. Nowadays, mobile based health solution seems very effective since almost every types of people are using smart phone and varieties of mobile applications in their daily life. We try to combine Machine Learning with Smart mobile application and also Web application in one platform. This new genre, mobile based and web based predictive application can be an effective solution for the female in Bangladesh which will help them to understand the possibilities of having Breast cancer with self-check up through the application as well as the suggestion about the nutrition according to their current health status.

1.2 Motivation

Bangladesh is an over populated country from its largeness that's why in our country some scarcity contains. People are not concern about their simple diseases because some lack of awareness. In here, peoples are not detected their diseases at the first attempt by symptoms that's why the result

of which is gone to the critical moment. The impact of breast cancer is increasing day by day which is mostly suffering the woman in our country and around the world. After researching on this topic, we can see that our women's are just know about the name of this disease but they are not so much aware of it and hesitate to talking about this topic. We have visited Bangladesh Breast Cancer Awareness society and discuss about that topic with a specialist. His organization doing campaign and inform to the people about breast cancer information and gives instructions. But people can't know that in which level or which grade they are surviving. So, from this study is to ensure to predicting the level of Breast Cancer according to the symptoms.

1.3 Rationale of the study

The effect of this research project will always continue and also the effect made by this research project is long term. Earlier works which was related with this similar topic they has been done so far of classification and get the accuracy with comparison. Based on this research we get some information and know that other country like India, Ireland they take some initiative about Breast Cancer issues to apply on real life device but they also not so much enrich. But as like our country none of them are used in real life device accuracy detection and suggest anything which for a user can get this simple question answer that, how can it will be reduce? They just inputting patient's symptoms and use reminder to remind a patient about their schedule list. This is the main gap between the study and the studies have been accomplished before. Throughout this research project that issues will be focused. The most important thing is that, by this following study it will be used in the Real life implementation such devices and web based for people to assume their Breast cancer types, level, condition very easily.

1.4 Research Questions

We have developed some research questions which has been referred to the problems of our project focuses on and outlined the tasks that have to be completed. The research questions are as follows:

1. Which Machine Learning algorithms or Data Mining techniques will be used to accomplishing the satisfied and optimal classification accuracy and why?
2. How the data will be linked with Breast cancer detection level.
3. In real life applications devices and which data are used in earlier prediction work and in this study which extra features are added.

1.5 Expected Output

Many researchers have done a lot of research on this and many successful results have been found. But they did not research and apply it to any platform, they only detected through research. Some researcher have only a way of classifying, but no means of identifying possible restriction features. As the output of our research, we want to detect by using machine learning methods that produce better results. Besides, these detection results will be covered by various platforms like web and smart applications. So that the user can apply the information himself to get the desired result.

1.6 Report Layout

Chapter 1: Introduction

In this section, explained and focused all of the basic things about our research project. Where many information about breast cancer has been reviewed and how much Bangladeshi and other lower-middle country peoples can be affected it also explained here and provided some information about that also. The main point is here that how we got this idea to bring the matter to the fore we describe in this section. Over the above, some questions which are related with our research project that will be come in fore. The most importantly about this research project expected outcome will be discussed in this part.

Chapter 2: Background

The section of background is the explanation of previous literature works and where related works also come in front which had been done by researchers. Because of their previous works we can find the way to thought about that issue broadly and uniquely. We can find out through this what they have done in their previous work, which algorithms are properly worked and give them best accuracy in result, how was their research success rate, how much they can implement and besides that what couldn't they able to implement such kind of things really helps to us in our every steps to made this research project widely. We can learn many things and get that what is the scope of our research project in the same area which will be focused here and also discussed about that what kind of challenges we will have to face to get this this research done.

Chapter 3: Research Methodology

Basically, the title of this section says its main goal is research methodology. In fact, if we consider this part to be the root of the research, it is not wrong to think so. Because in this section described that most specifically how the research project will be get done through via which procedure used. In this section firstly showing our survey results and its description also given after that part also showing about our diagram of application system and given hints there for understand easily. In shortly, discussed about the research subject and instrument such as steps of methodology diagram, machine learning library's which is uses for apply machine learning methods properly, describe about technique. We know that, Data plays an important role on research because without data a research can't discover and represent a proper and unique thing. So, mainly focused in data and discussed here about data such as data collection process, which algorithm uses for analyzing this data, which data parameter uses on a particular case as like preprocessing data and how data are classified and so on. In there also have include the statistical analysis and implementation requirements.

Chapter 4: Experimental result and discussion

This is the most important section in this research project where discussed about all the evaluations also showing all the mathematical terms. In this section every experimental results and analysis provided with its description. Where every experiment and analysis is related with mathematical term which helps to create methods model and getting proper result in this research project. Web application and Smart application system result view and description provided and get to know that how it can works in real life. At the end of this section there given a short note which summarized this section.

Chapter 5: Summery, Conclusion, Recommendation and Implication for future research

In very shortly, the whole research project described or summarized in this section and reported all those things which is available or not available. This section describes more about recommendation and this projects further implication. At present, machine learning is delivering breakthroughs in other sectors as well as in the health sector and, as a result, it can be said that the ultimate result of our research project can helps to think others for more advanced research in this field and Healthcare sector with also the stream of awareness will go a step further by this research project we hope.

CHAPTER 2

BACKGROUND

2.1 Introduction

There are a lot of other countries like ours, there is a means of becoming breast cancerous where survival rate is low. On the other hand people generally takes it as a cursed and dislike this person actually who needs treatment. To get rid of this curse, we need to expand our research work on these issues. The way our previous researchers applied their algorithms and get good detection result to move forward. In before, they work same type of works which has been done in different way which has helped us a lot for sorting that which algorithms to work with.

2.2 Related Work

Many researchers are proposed their works in different themes and motivated by many ways to detect or identifying this topic related issues. Many researchers applies different methods and the results have been varied. So, this research work is not the very first time which has been done and scholars are have been done almost same sort of work such as-

In a study [6] breast cancer is discussed about the people of India. They proposed that the main reason of Breast Cancer in India are treatment in advanced stage of the disease, lack of awareness. They also included that the development of breast cancer among the 30-50 years are much higher than 50-70 years in recent time and the survival rate is very poor. The motive of their paper is to generate awareness about self-breast examination in women and promote the idea of installing mobile health apps to prevent breast cancer. Their main objectives were to develop questionnaires of discussion based on FGD model (Focus Group Discussion), UTAUT model (Unified Theory of Acceptance and Use of Technology) [7]. They also discussed about the statistical result. Their main focus was to outcome about reason to use the breast cancer checking mobile app for the doctors and the patients.

According to Improving Outcomes from Breast Cancer in a Low-Income Country: Lessons from Bangladesh [8] they have reviewed about 238 Breast Cancer case records at the tertiary care government hospital, Khulna Medical College Hospital found that the cases were confirmed through tissue diagnosis and clinically the cases ware over two-year period and most of the cases were found in Stage III+ and one-fourth women had lumpectomy surgery and removed tissues

without any additional treatments. They also added that about 34% women don't know about the treatments of the breast cancer cases amongst 245 rural Bangladeshi women. They also stated a record about 238 new cases 2007-2008 in Khulna medical hospital where 4% cases were in stage (I) and (II), about 87% were in stage (III) and 9% were found in stage(IV). In this paper they also demonstrated about the obstacles for the women of the Bangladeshi rural and urban area.

Hossain et al. [9] in this study they reviewed about Epidemiology, Risk factors, Pathology, Diagnosis and Treatment of Breast Cancer in Bangladesh perspective. They highlight and differentiate to the developing countries with Bangladesh about lack of breast cancer awareness, inadequate access. Most of the women are unaware of breast cancer and cannot diagnose early where early detection has a vital impact on overall results of treatment. They introduced an ambitious female-based primary healthcare strategy focusing on education, screening and early detection of breast cancer in Bangladesh, keeping in mind all these problems.

Yue et al. [10] in this study they focused on Machine Learning Techniques and their Applications in Breast Cancer diagnosis and prognosis [11]. They included Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Decision Trees (DTs), and K-Nearest Neighbors (k-NNs) all those Machine Learning Techniques in their work. They uses their primary data which was drawn from the Wisconsin Breast Cancer Database (WBCD). This is the benchmark database for evaluating the outcomes through various algorithms. They got different accuracy from different Machine Learning algorithm. They get from ANNs- 98% accuracy, in SVMs- 97.2% accuracy, in K-NNs- 98.25% accuracy.

Asri et al. [12] in this study they uses different algorithms for make a better decision in Diagnosis and Analysis and get the performance comparison between different Machine Learning. They included Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB) and k Nearest Neighbors (K-NNs) those Machine Learning Techniques in their work. From their work they get the best accuracy which was SVM- 97.13% with lowest error rate. They uses on their work the Wisconsin Breast Cancer (original) datasets.

Mihaylov et al. [13] they discussed about the main factors, used to predict survival time is the originally developed tumor-integrated clinical feature, which combines tumor stage, tumor size, and age at diagnosis. In their studies they uses different types of Machine Learning techniques which was provide the promising outcomes and analysis of their research work shows an advantage of the Linear Support Vector Regression (LSVR), Lasso Regression (LR), Kernel Ridge

Regression (KRR), K-Neighborhood Regression (KNR), and Decision Tree Regression (DTR). They use two datasets on their work which shows good performance and no data losses. On their study the cross-validation for accuracy gets the best performances.

Liu et al. [14] they proposed different Machine Learning methods for automatically identify cancer cells and after the higher detection accuracy to remove internal subjective human factors in the process. They work in this paper, the Logistic Regression (LR) algorithm of Sklearn Machine Learning library is used to classify the data sets of breast cancer (diagnosis). Two features are selected for showing the result which are maximum texture and maximum perimeter and their classification accuracy is 96.5%.

Park et al. [15] they experimented the Machine Learning method via using a custom-built microscope-compatible micro-indentation system. They use pathological images on this indentation and each indentation labeled as either it is “Normal” or it is “Cancerous”. They applied Support Vector Machine (SVM) method. They obtained the classification accuracy outcome from entire breast tissue samples and the accuracy outcome is $76.20\% \pm 9.28\%$.

Chaurasia et al. [16] they make comparison between different kinds of Machine Learning methods which are actually well performance based supervised learning classifiers and find the best classifier in Breast Cancer datasets. Such as Naïve Bayes (NB), SVM-RBF kernel, RBF Neural Networks, Decision Trees (J48) and simple CART. In their research they show that Support Vector Machine (SVM-RBF) kernel is more accurate than other classifiers. From SVM method they get the accuracy outcome 96.84%. In their work they also use Wisconsin Breast Cancer (original) datasets.

Sharkas et al. [17] used the Discrete Wavelet Transform (DWT), the Contour-let Transform {CT} and the Principal Component Analysis (PCA) methods for prominence expulsion. It is able to detect Normal and Abnormal Tissues from that system and it classifies the Malignant and Benign Tumors. Their success rate was achieved almost 98%.

Ragab et al. [18] after Discrete Wavelet Transform (DWT) method they also make a comparison between Support Vector Machines (SVM) and Artificial Neural Networks (ANN) for classifying Normal, Abnormal Tissues and Benign, Malignant tumors was introduced. In that comparison they get the detection rate or accuracy which was 96% for ANN and 98% for SVM.

Wichakam et al.[19] used the Deep Convolutional Neural Network (DCNN) and Support Vector Machine (SVM) from Machine Learning methods. The sensitivity achieved was 98.44%. In this work they use the INbreast Dataset.

Ahmad et al.[20] they implemented three Machine Learning techniques in their work which are Decision Tree (DT) (C4.5), Support Vector Machine (SVMs), and Artificial Neural Network (ANNs) for develop their predictive models. The main thing is in their work that they would try to compare the performance between these three well known Machine Learning algorithms with their dataset where data will through via it's Sensitivity, Specificity and Accuracy. They collect their datasets data from an Iranian Center for Breast Cancer (ICBC) program and patients were registered from 1997 to 2008. In their dataset it had been contained 1189 records, Where the Predictor variable was 22 and one outcome variable was also included.

Delen et al.[21] used Artificial Neural Networks (ANNs), Decision Trees (DT) and Logistic Regression (LR) and they analyze a largest dataset which is the SEER cancer incidence database for develop the prediction models of breast cancer survival.

2.3 Research Summery

On the basis of early related work they applied different kinds of techniques like Support Vector Machine (SVMs), K-Nearest Neighbors (KNNs), Artificial Neural Network (ANNs), MLP and some people also apply Convolution Neural Network (CNN). Where Support Vector Machine (SVMs) algorithm work well in every study for doing prediction and detection and also diagnosis and prognosis.

2.3.1 Risk factors

From this case study and survey, there have found that most common risk factors-

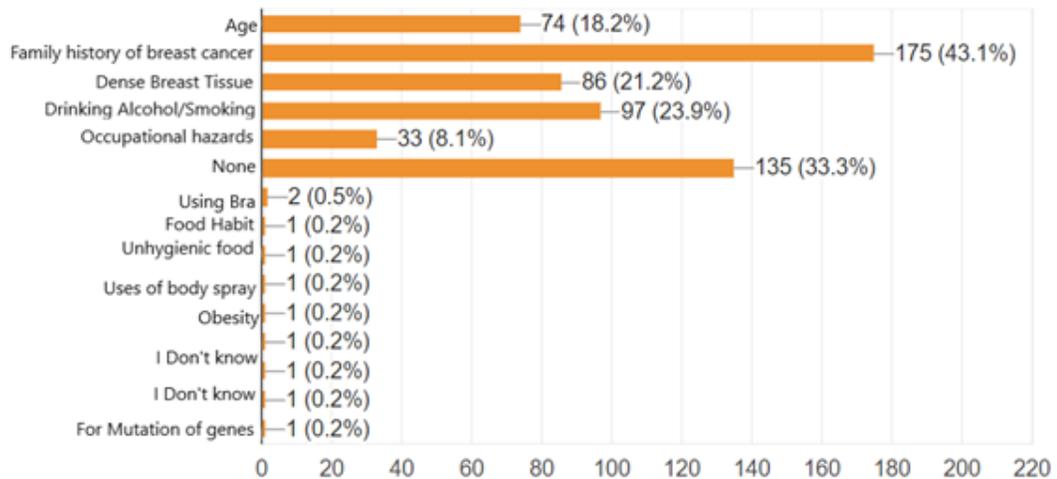


Figure 2.1. Breast Cancer Risk factors chart

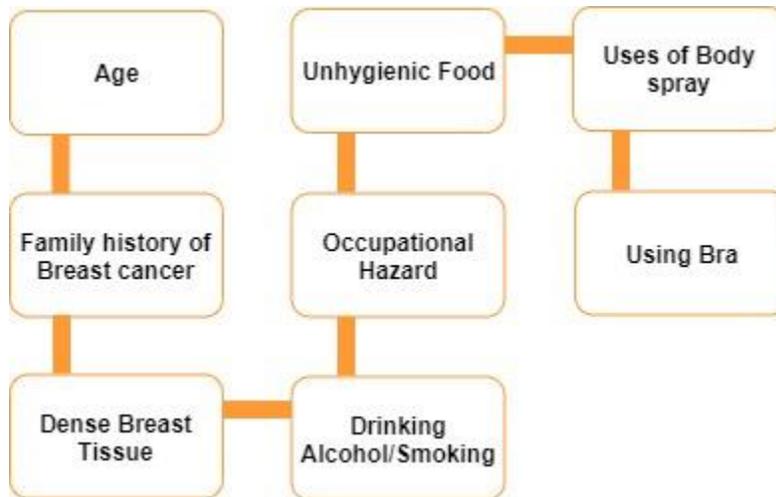


Figure 2.2. Common Risk factors

After doing a survey we got that results of percentages which define that how much people are knows about breast cancer risk factors and we also searches to know about any new risk factors from the people. By using this percentages of results we found that common risk factors which is in Figure 2.

2.3.2 Stages of Breast Cancer

Stages are the process that physicians use to assess the size and location of a patient’s cancer. This information is required for the determination of the optimal form of treatment. Breast Cancer can be classified into stage 0 to stage IV according to the size and nature of spread. Various studies were taken to determine the age variation. According to a study of Chattagram maa-O Shishu Hospital Medical College states that most of the malignant lesions are above the age of 40 years. About 50 patients took part in that study and the outcome is shown in the graph below.

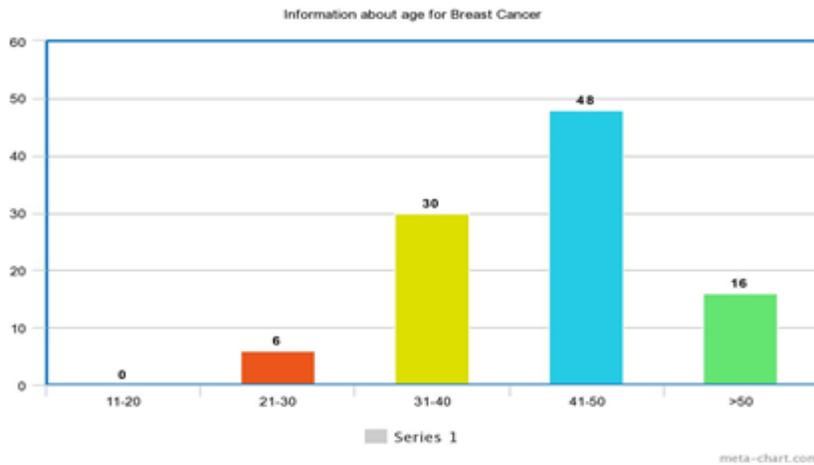


Figure 2.3. Ages for Breast Cancer chart

Source: Study on Clinical Presentation of Breast Carcinoma of 50 Cases (Chattagram Maa-O-Shishu Hospital Medical College Journal Volume 13, Issue 2, May 2014)

Some information also came to front about the stages of the affected patients from this study. This study shows that 76% of the total patients were having advanced carcinoma breast (Stage III, IV) at the time of reporting. Displayed table shows the data of the outcome.

Table 1. Patient’s Stages

Source: H. D. N. M. M. B. A. A. J. B. Mosammat Mira Pervin, Study on Clinical Presentation of Breast Carcinoma of 50 Cases, Chattagram Maa-O-Shishu Hospital Medical College Journal, 2014, PP. 09.

Stages	Number of patients	Percentages (%)
Stage-I	5	10%
Stage-II	7	14%
Stage-III	27	54%
Stage -IV	11	22%

Almost all the studies reached the outcome as we do through our survey. So we can assume that women above 40 years are in the most critical situation.

The current state will fall to pieces if early attention is not given. So, necessary steps must be taken by policy makers and health professionals to be aware the general female population about breast cancer.

2.4 Scope of the problem

In our project has been working on some of the main parameters that make it easier to classify algorithms. And the precept scope of our research project works is to outline the optimum algorithm in terms of classification to find the actual types of Breast Tumor which is associated with Breast cancer. The major goal of this research project is executed with the comparative overall performance analysis between KNN and SVM. After that Machine learning prediction result it will push to the web application section and Smart application section for directly using by user.

2.5 Challenges

Data is a very valuable thing in a research project and finding data is a matter of hard work. The Health related work for analysis the most faced problem is Data. For us, nothing has happened in terms of data collection, and it was a challenge for us. When we did the survey, we had to go through a lot of frustration at first to get the data, but later it changed dramatically and we were successful in getting the data. Another challenge arose when we went to establish machine learning with the part of web application. We want the user to be aware of a successful outcome according to their information and according to the machine learning analysis prediction. Eventually we reached our goal and were able to complete the project.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The main part of the work of a research project is its methodology. This is because, through methodology, a research project is able to move forward step by step and achieve perfection. It is possible to discover new things through the correct modes of methodology and proper planning. As a result, we have also worked on breast cancer datasets through various procedures. Actually Breast cancer dataset can be related to a number of things, in which there are two main ways for prediction or analysis: report based or image based. With regard to health systems, it is impossible to obtain image-based data. So we worked with the mammogram report based dataset and integrated the machine learning method into the web and smart applications with the user in mind.

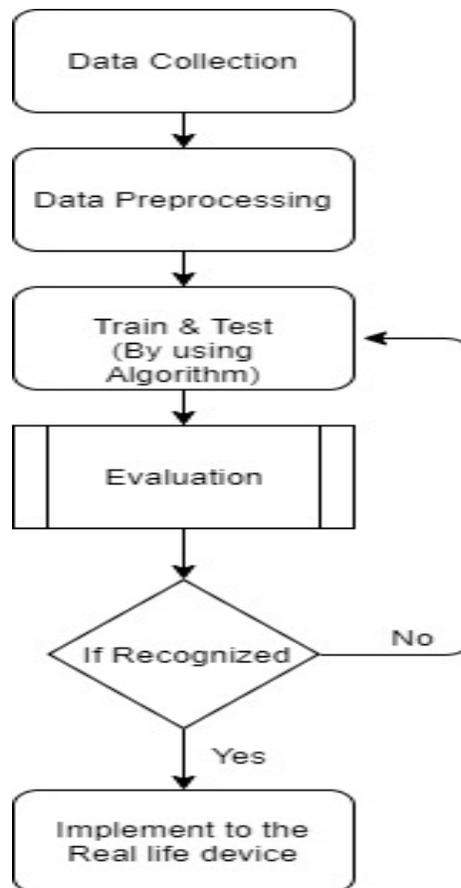


Figure 3.1. Block diagram for the propounded methodology

3.1.1 Breast cancer from survey

An online survey was arranged which aimed to assess the level of knowledge about breast cancer, risk factors and early warning signs. The study emphasize on the participants occupation, age and also the family history of breast cancer.

Around 406 responses were found where the participants are from seven countries and most of the participants are from Bangladesh (98.5%), India (0.5%) and Nepal, Scotland, Nigeria, Russia (0.2%). Whereas, Female participants are 55.7% and Male participants are 44.3% and there most ages are 16-24 (80.3%).

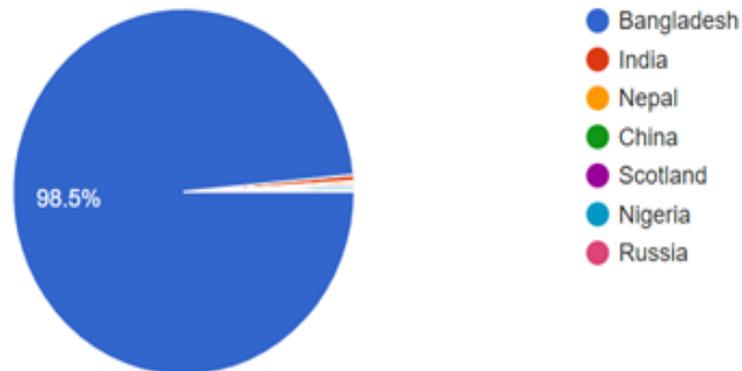


Figure 3.2. Participants Countries

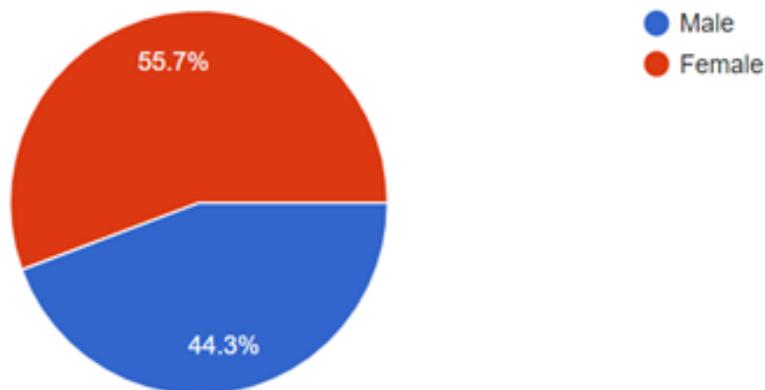


Figure 3.3. Participants Gender

Responses have been reported from 79.6% participants have heard about Breast cancer those who were mostly students (86.9%) and connected with different types of occupations.

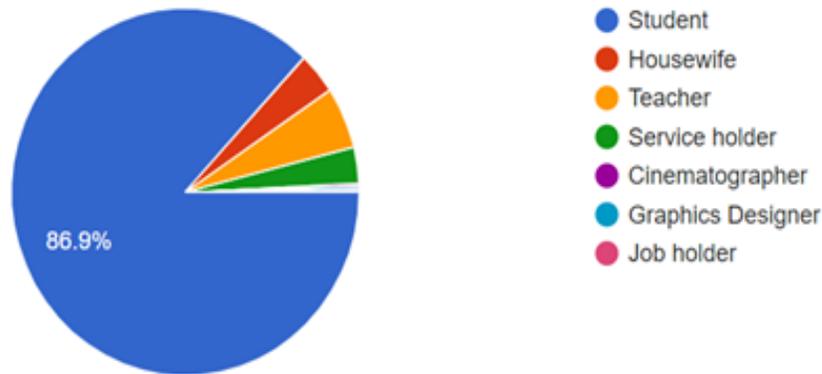


Figure 3.4. Participants Occupation

This Survey shows that 1.5% participants are invaded by Breast Cancer and they discover that via- Found most of as a result of another test or investigation where the percentage were 27.3% and second most is a lump, Had bleeding (discharge or other symptoms) and Had a mammogram where the percentage were same 18.2%.

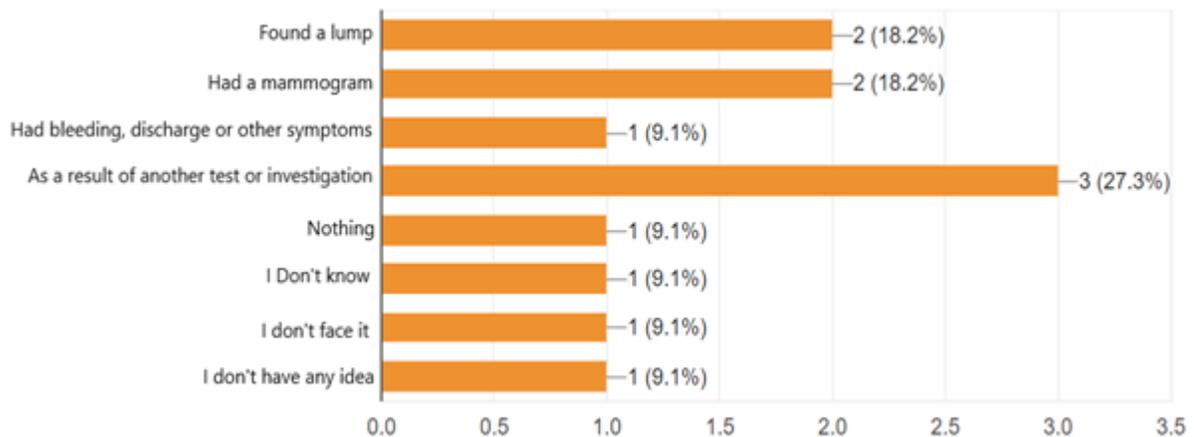


Figure 3.5. Breast Cancer finding sources

The survey we have been through the most common symptom of breast cancer is found that “a new lump or thickening in one breast or armpit that is different to the rest of the breast (48.3%)” and other most symptoms percentages are found as, Size and shape: if one breast changes in size or shape (40.4%), Rashes on or around the nipple (26.4%), Nipple changes: if a nipple becomes inverted (pulled in), changes shape or points differently (26.4%), Discharge or fluid (not milky) from one or both nipples (24.6%), Pain or discomfort in the breast or armpit that is not related to your periods (23.6%), Skin changes: puckering, dimpling, inflammation or redness of the skin

(17.2%), Swelling under the armpit or around the collarbone (6.7%). In this survey it also shown that, 0.2% participants they don't know about that symptoms and none percentages (31%).

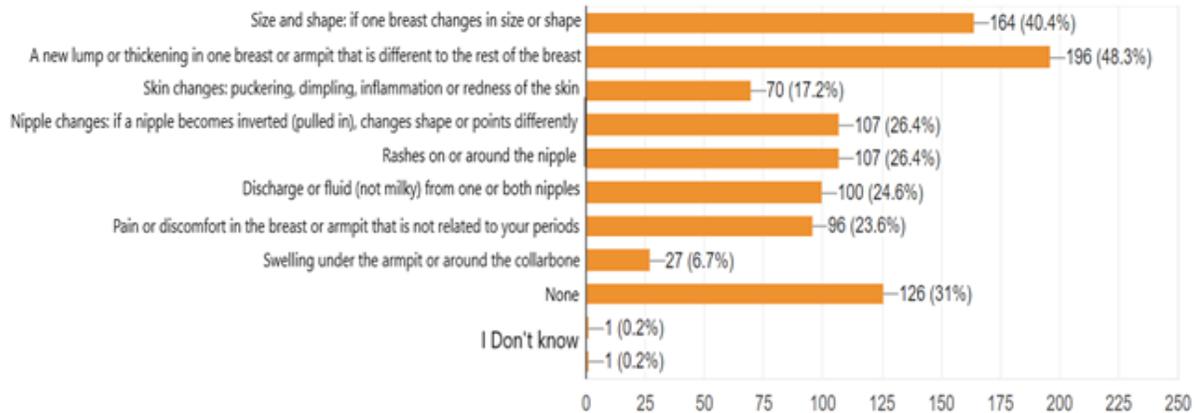


Figure 3.6. Awareness of Breast Cancer Symptoms Chart

3.1.2 Use case Diagram

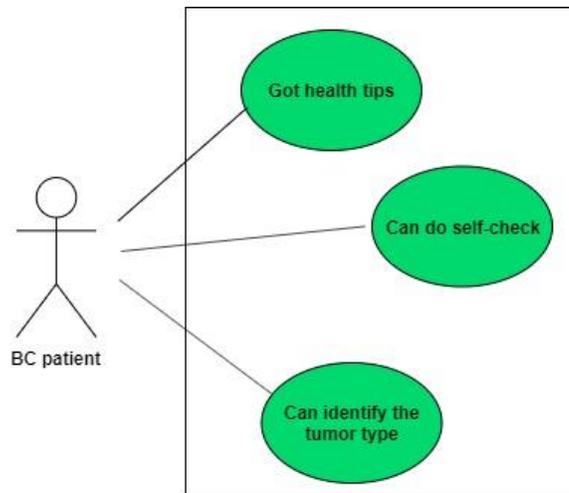


Figure 3.7. Diagram for general user

Follow this above diagram in the below given explanation about mobile application general user features (proposed)-

Got health tips- General User will get health tips which helps them for doing confrontation.

Can do self-check- If a general user know the checking steps which helps them to check their own-self and match with symptoms then they can measure that they are already victimized or not.

Can identify the tumor type (if cancer exist)- After a user self-check and their provided matched symptoms data they can identify that in which tumor type they have.

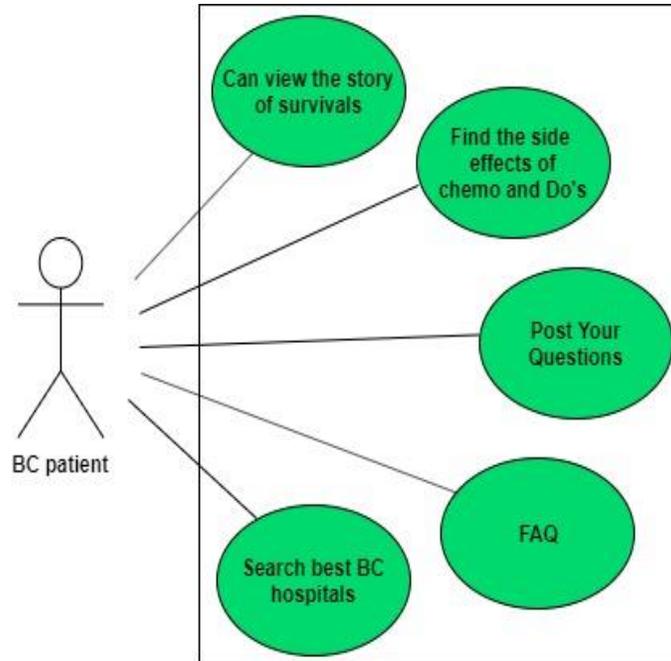


Figure 3.8. Diagram for cancer patient

Follow this above diagram in the below given explanation about mobile application breast cancer patients features (proposed)-

Can view the story of survivals- The story of a survivor will inspire a cancer patient that they can also survive their-self.

Find the side effect of chemo & do's- Side effects of chemotherapy are usually temporary and most can be controlled.

Post your questions- Post your unknown all the question for your betterment.

FAQ- Frequently Ask Questions.

Search best BC hospitals- For your better treatment and also considering the financial aspects it will help to find out your best BC hospitals.\

3.2 Research Subject and Instrumentation

3.2.1 Machine Learning Intelligence Library

With other Scientific Computing Libraries ' useful resource such as Numpy [22], SciKit-learn [23], Pandas [24] are used in this study to implement the Machine Learning algorithm.

3.2.2 Techniques

Actually Machine learning techniques are deeply attach with Data Mining techniques because it's all about data. Those techniques are in collusion with another. In this study, we used K-Nearest Neighbors (KNN) [25] and Support Vector Machine (SVMs) [26] machine learning technique to predict the types of Breast Cancer Tumors.

K-Nearest Neighbors (KNN) [25] is a classifier and regression. It actually classify directly from dataset when data are loaded and by using number of K (k=1) and predict that how closest data are taken. By this classification it reveals the stage of breast cancer.

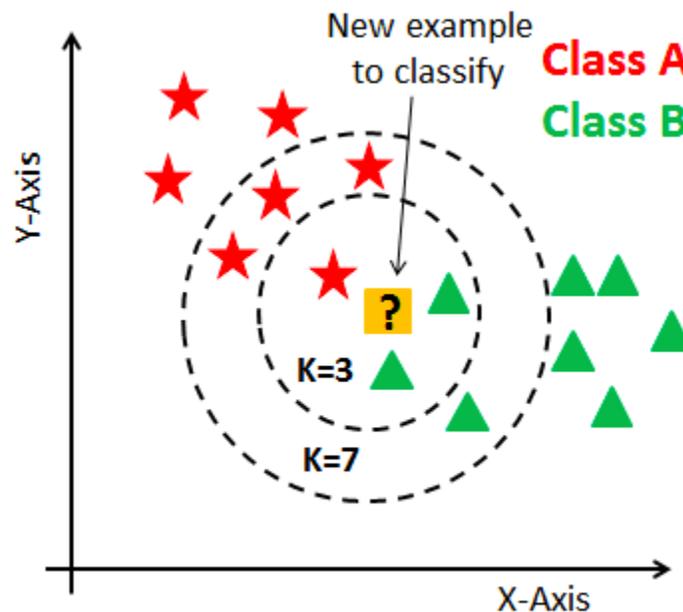


Figure 3.9. K-Nearest Neighbors (KNN) classification for Breast cancer

On the other hand, Support Vector Machine (SVM) [26] is an emerging effective Machine Learning approach to classify cases. SVMs were used to vary of issues and they've been successful in sample focus in Bioinformatics, most of Cancers Prognosis [11] and more.

SVM is an algorithm of maximum margin classification rooted in the theory of statistical study. This is the method for both linear and non-linear data classification. It makes use of a Non-linear mapping technique to radically change the original education records into a greater dimension. It performs classification duties by way of maximizing the margin separating both instructions while minimizing the classification mistakes [27].

Here in the below figure-1 shows the Support Vector Machine (SVMs) topology-

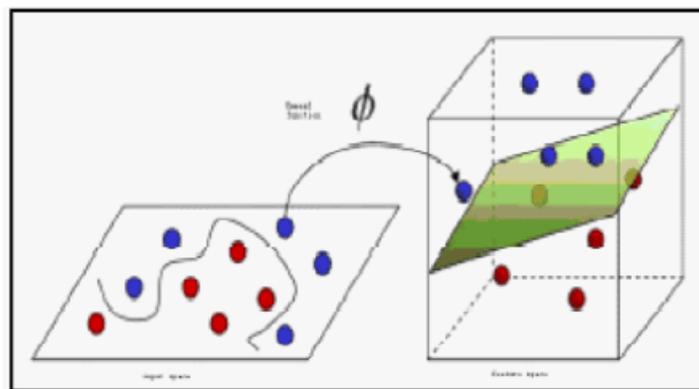


Figure 3.10. Support Vector Machine (SVMs) Topology

3.3 Data Collection Procedure

For the purpose of training out model we checked many sector including hospitals and contacted with personals to collect data. But is a matter of great sorrow that we were unable to gather enough data. We visited Delta Hospital but they refused to share their data set. We also contacted with an organization named Breast Cancer Awareness Society. They also could not able to help us with dataset.

After doing a lot of things here and there we decided to collect dataset from internet. But all of those were out dated. Throughout some research we were able to find a dataset about Mammography report which is used in our project. The dataset was from data.world Mammographic masses repository [28] [29]. And it was shared by JULIO CARDENAS-RODRIGUEZ. It was produced in 2015 which is the newest of all that we checked.

3.4 Statistical Analysis

3.4.1 Data preprocessing

Often data is scattered in a dataset, making it difficult to get accurate results. That's why it's very important to do data pre-processing. Various parameters through which the actual results can be obtained when the prediction is performed have been adjusted from this dataset which for it can improve to get actual accuracy and attain better performance.

3.4.2 Using K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) algorithm:

According to the ideas of previous researchers and their work experience, we have come to realize that compared to all other algorithms, K-Nearest Neighbors (KNN) [25] and Support Vector Machine (SVM) [26] algorithm this two are uses very much for predicting Breast cancer detection and improving accuracy with better performance.

3.4.2.1 K-Nearest Neighbors (KNN)

k-Nearest Neighbors (KNN) is one of the Supervised Machine learning algorithm [30] from others and it easy to implement. After data preprocessing this data are loaded with initialize of K by increase and decreases that k of number which using for classifier and regression. Though this algorithm is significantly slower of predictors but it has been able to bring good results in the field of prediction.

3.4.2.2 Support Vector Machine (SVM)

Statistics show that Support Vector Machine (SVM) was most often used for breast cancer prediction. It can be used for classification, regression and also different tasks for analysis. Actually in SVM there create a model and if classification is linearly separable then SVM fits the decision boundary which is described the closest points in every class by using the biggest margin. Usually known as the Maximum Margin Hyperplane (MMH) [31].

3.5 Implementation Requirements

For our research project we developed our methodology to implement and get the proper result which is absolutely works to predict the Breast Cancer tumor types which is Benign [4] or Malignant [4]. The whole research and project implementation developed by many steps. But for research purpose the major steps are divided into three part for find out the best result. Implementation steps are like:

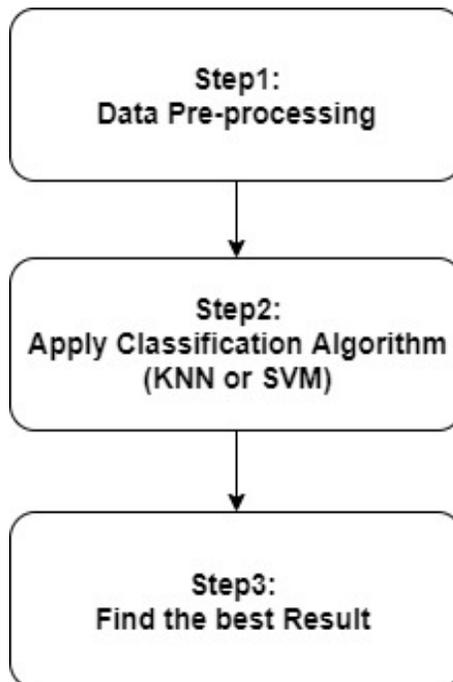


Figure 3.11. Three major part of Methodology

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Introduction

This section represent all the results which we got from using all the methodologies of machine learning techniques. In the above description we describe that we applied different types of algorithm or techniques on breast cancer dataset in our research project which for we got different types of results and it helps to take us the decision that which algorithm is best for prediction. In this result section there also provide web application and smart application user interfaces which related with support vector machine (SVM) to showing users the predicted result.

4.2 Experimental Results

After data preprocessing stage we apply that to the K-Nearest Neighbors (KNN) [25] and also to the Support Vector Machine (SVM) [26] classifier. Whereas from K-Nearest Neighbors (KNN) [25] classifier we got 71.02% classification accuracy. On the other hand, from Support Vector Machine (SVM) [26] classifier we got 81.29% classification accuracy.

4.3 Descriptive Analysis

4.3.1 Training, Testing and the Validation of the model

We created separately training data, testing data and validation data from our dataset for find out the performance of the model. For the training session we used the training set. Actually the model is initially fit on a training dataset [32]. When this time we train the data to test the quality of the model, we used the validation set to help tuning the model's hyper-parameters. At last we used the test data for finding out the performance of final model performance.

4.3.2 Model performance

Confusion matrix [33] is widely used for classification model performance justification. It operates on a set of testing dataset where the true and real values are familiar. In figure-16 it represents a drawing this is our proposed model confusion matrix [33]. In confusion matrix the entries are True positive (TP), False Positive (FP), True Negative (TN), False Negative (FN) rate for each type of dataset.

The overall performance of the classifiers in breast cancer detection can be measured from the confusion matrix analysis and determined below parameters.

Accuracy is the percentage measure of efficiently classified instances for all instances. It obtained by this equation which is given on the below.

$$Accuracy = \frac{(TP + TN)}{(TP+TN+FP+FN)} \quad (33)$$

For those cases which are classified as positive, precision is correctly classified and calculated using the equation,

$$Precision = \frac{TP}{(TP + FP)} \quad (34)$$

Recall is the calculation of the positive instance which can properly classified and can be determined with the equation below,

$$Recall = \frac{TP}{(TP + FN)} \quad (35)$$

F-Measure is the combined precision and recall measure. It demonstrates how accurate the classifier is and how reliable the classifier is. Use of F-measurements below for calculation.

$$F - Measure = \frac{2 * Recall * Precision}{Precision + Recall} \quad (36)$$

Sensitivity is the calculation of positive instances correctly assigned to a maximum number of positive instances.

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (37)$$

Specificity is the calculation of negative instances correctly assigned to a maximum number of negative instances.

$$Specificity = \frac{TN}{(TN + FP)} \quad (38)$$

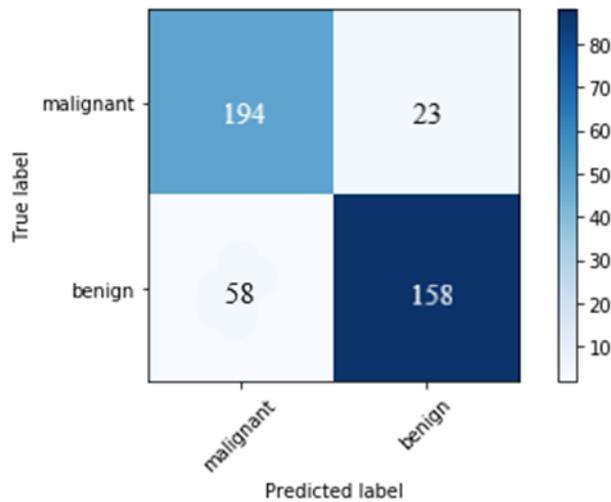


Figure 4.1. Confusion Matrix for Breast cancer using SVM classifier

	precision	recall	f1-score	support
0	0.77	0.89	0.83	217
1	0.87	0.73	0.80	216
micro avg	0.81	0.81	0.81	433
macro avg	0.82	0.81	0.81	433
weighted avg	0.82	0.81	0.81	433

Figure 4.2. Weighted average performance parameters of SVM algorithm for dataset

Here, Precision = 82%

Recall/ Sensitivity/True Positive rate = 81%

F1-score = 81%

Specificity = 87% False Positive rate = 12%

4.3.3 Machine Learning to Web application

Since it is possible to discover new things through research, we want this user to use this new invention directly. As a result, we wanted to bring machine learning and web applications to a platform. Where a user can find out about his / her own predicted result by this platform by providing certain information from his mammogram report and be aware of himself / herself. This is just a simple and user friendly Single Page web Application (SPA). Here describe web application front end design work in the below,

4.3.3.1 Web application Home page

This is the Home page section when a user start browse on there then as a welcome screen he reached in here.

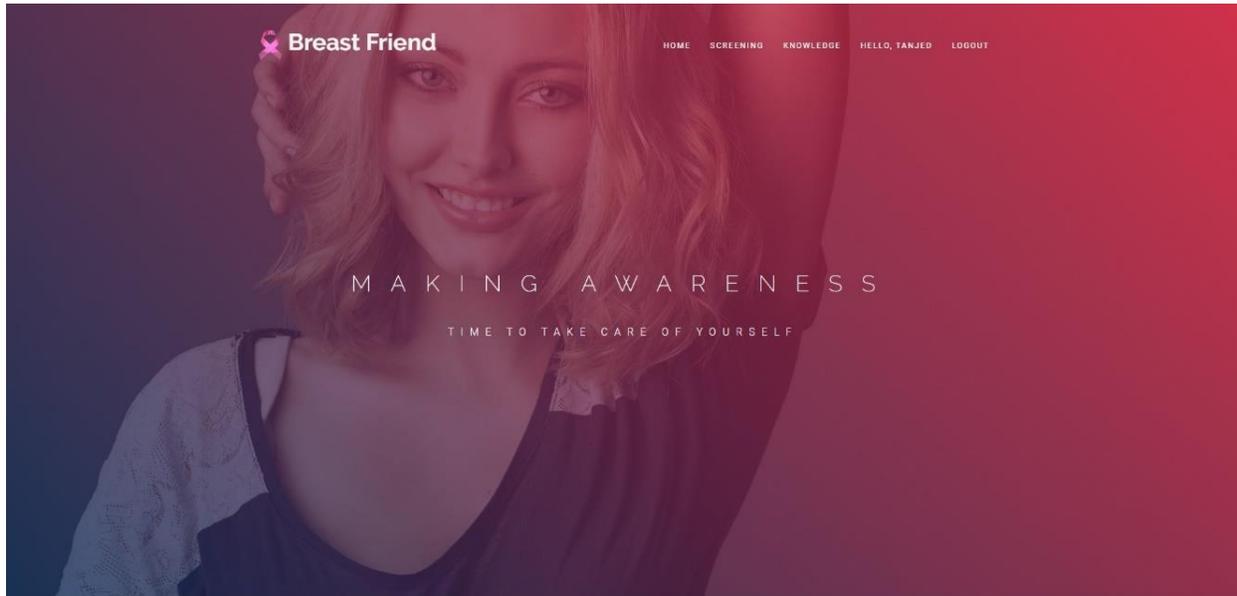


Figure 4.3. Web application Home page

4.3.3.2 About Breast cancer

Through this section, a user can learn about breast cancer. In addition to this, you can be aware of the type of breast cancer (By clicking on the Benign type for get to know about benign also they can get to know about malignant by clicking malignant).

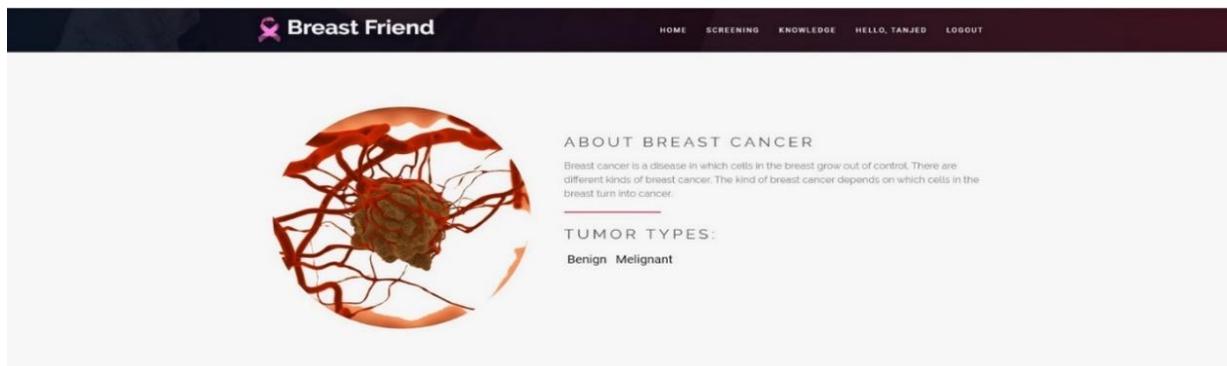


Figure 4.4. About breast cancer

4.3.3.3 Web application Scholars' Quote

There have some very important inspirational speech which for a user can take inspiration that breast cancer or any type of cancer can prevent by their willpower.

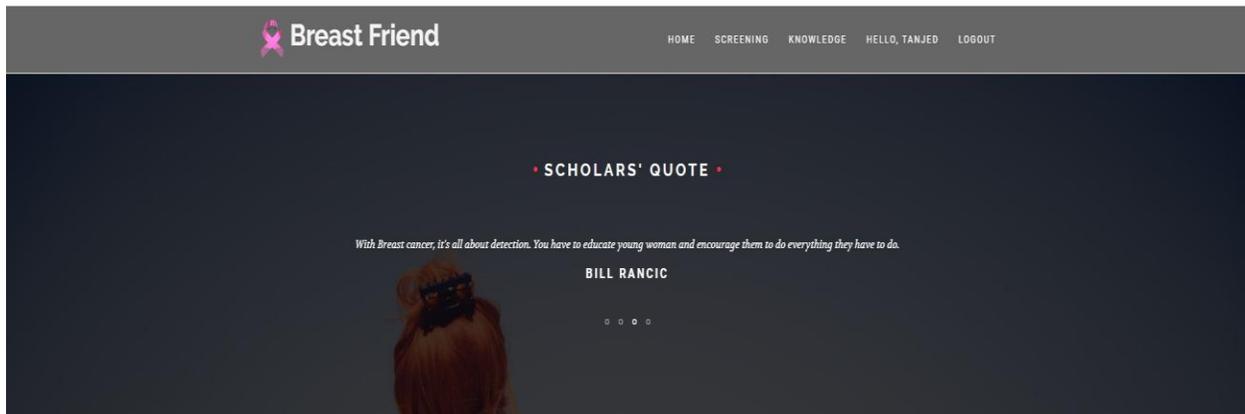


Figure 4.5. Web application Scholars' Quote

4.3.3.4 Web application Knowledge section

In that part a user can get knowledge about Breast cancer main Sign and Symptoms with its Risk factors.

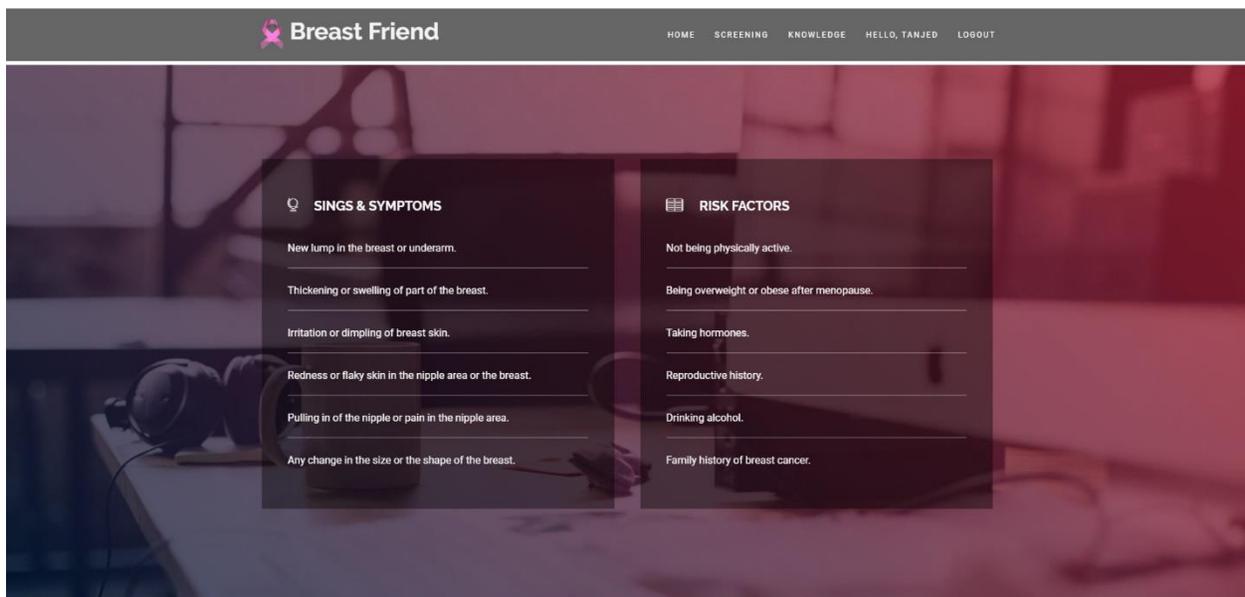


Figure 4.6. Web application Knowledge section

4.3.3.5 Web application Registration page

For a new user or if anyone want to register their own self then they to fill up and submit this registration page. After registration process they can login by their information.

Figure 4.7. Web application Registration page

4.3.3.6 Login

After registration process a user can login to this web application and also if a user want to check themselves then they have to login their own self. After login a user can screening themselves.

Figure 4.8. Web application Login page

But if a user not register themselves and want to test screening then they get this message that they have to register for testing.

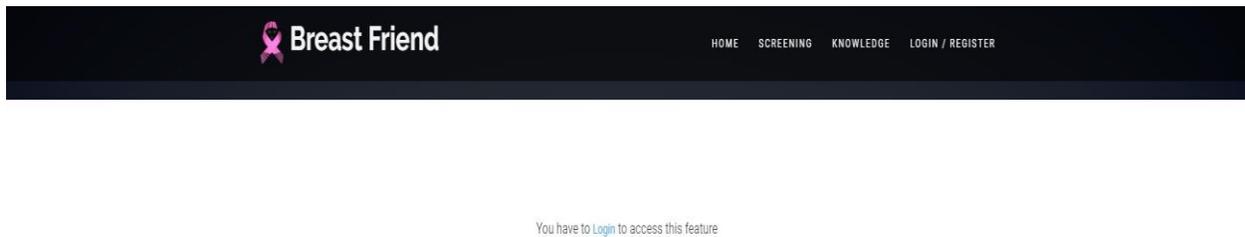


Figure 4.9. Web Unregister user view(for login)

4.3.3.7 Web application Screening test page

In this part user can test themselves given their mammogram report information. The result is directly driven by machine learning method and user can get the predictive result. Result shows that which type of breast cancer or tumor they have.

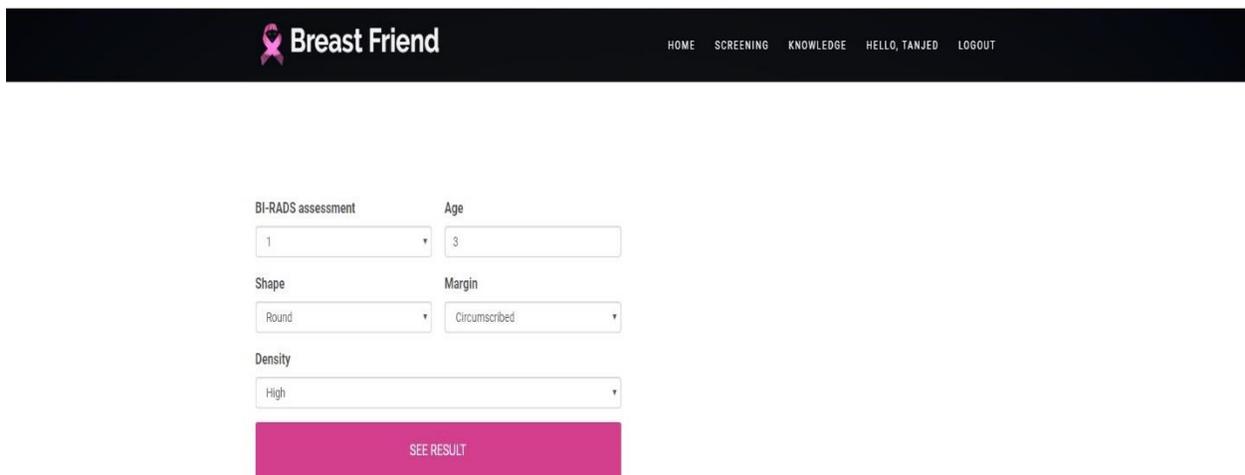


Figure 4.10. Web application Screening test page

4.3.4 Machine Learning to Smart Application

4.3.4.1 Mobile application Sign In page

If a user want to do screening test then he/she can test their own self by using that smart application on their smart phone but at first they have to sign-In on application.

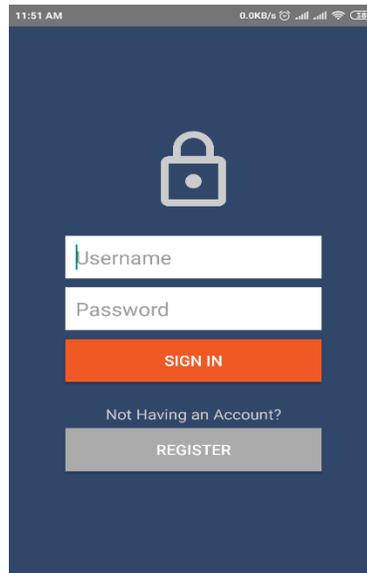


Figure 4.11. Mobile application Sign In page

4.3.4.2 Mobile application Sign Up page

We know that if we want to sign-In to the application then a user have to Sign-up by giving on proper information.

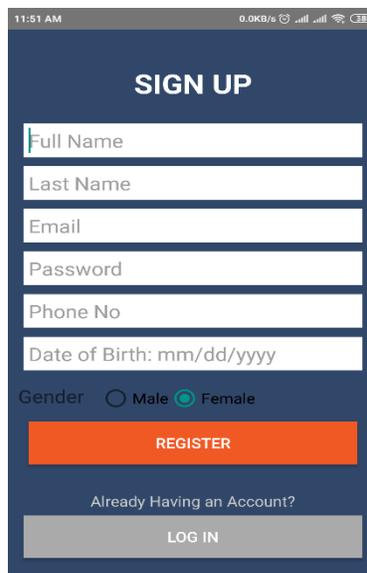


Figure 4.12. Mobile application Sign Up page

4.3.4.3 Screening

After the process of Sign-up and Sign-In then if a user want to do screening test then he/she has to follow this process and have to give information from mammogram report and choose options as like report have. If user do not want to screening test then they can get exit by using logout option as like first picture in the below.

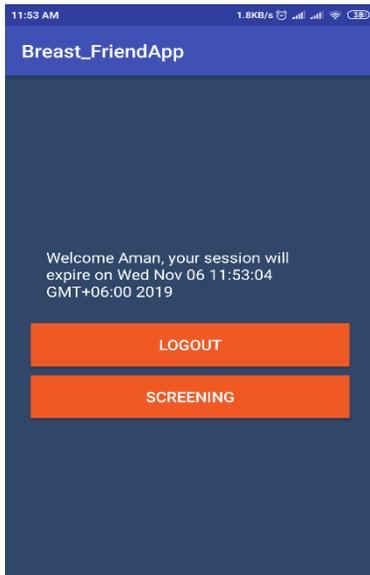


Figure 4.13. Do's or Don'ts

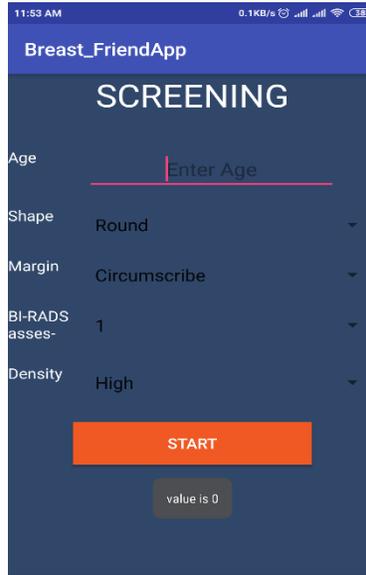


Figure 4.14. Enter Age

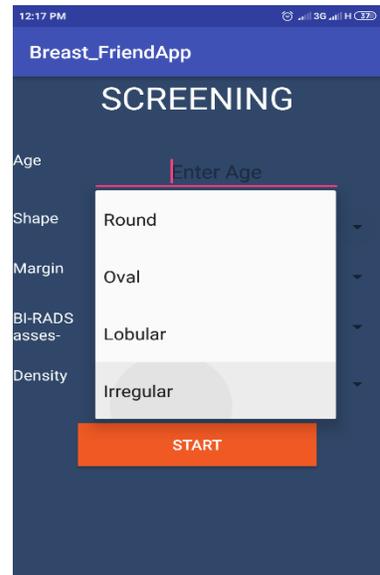


Figure 4.15. Choose the shape

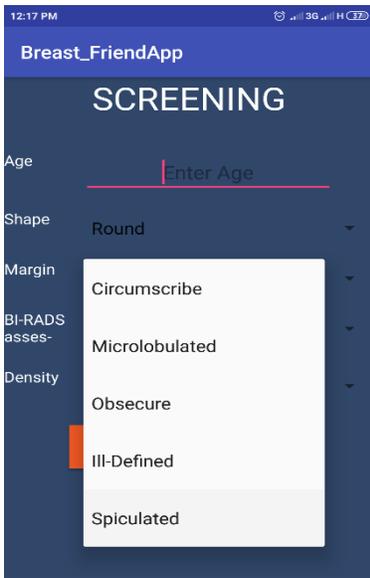


Figure 4.16. Choose Margin

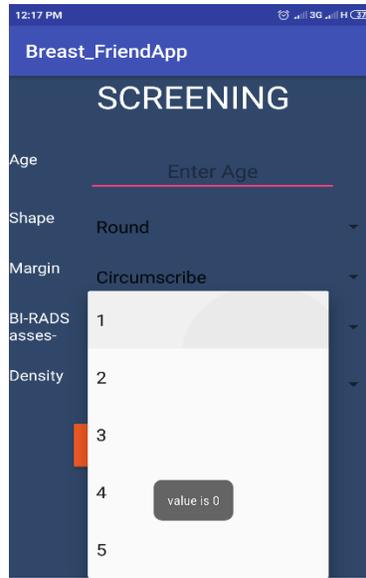


Figure 4.17. BI-RADS Assesment

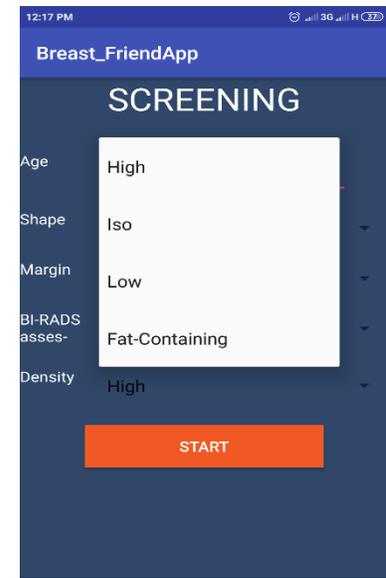


Figure 4.18. Choose Density

Figure 4.19. Mobile application information for doing Screening test (from mammogram report)

4.3.4.4 Result of screening test

If a user given all the information and press to the START as like the first picture in the below then he/she get the prediction result by this application.

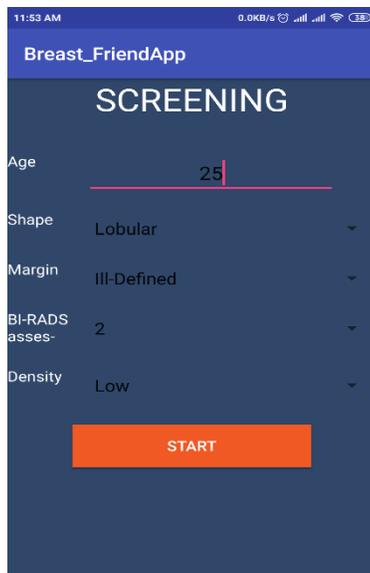


Figure 4.20. Submit Information

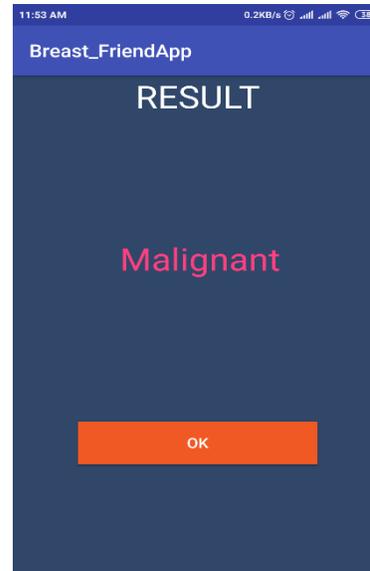


Figure 4.21. Result view

Figure 4.22. Mobile application Result of screening test

4.4 Summary

On the above description we see that actually we uses two classifier. After applying this two classification algorithm by using preprocessed dataset we got 71.02% accuracy for the -Nearest Neighbors (KNN) [25] classifier. From Support Vector Machine (SVM) [26] classifier we got 81.29% classification accuracy.

We've been able to combine web and smart applications with machine learning.

CHAPTER 5

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summary of the Study

The persistent of the study was to show how Machine learning techniques are helps and plays an important role in Health sector and how it can uses to predict the breast cancer tumor types. It was an experimental research study that looked that how data can preprocessed and divide into the three part (training, validation, testing) getting best accuracy and better prediction. We also connect that Machine learning techniques to the web application and also smart application for aiming to inform the user of the new systems and to make them aware. Because it is very important for everyone to be aware of Breast cancer.

5.2 Conclusions

In this research project report we provided different types of machine learning methods explanation and shows up front that data are processed and analyze data for diagnosis and prognosis to detect and getting better result on breast cancer. ML techniques have shown their exquisite potential to improve classification and prediction accuracy. After that preprocess the data we apply different type algorithm which for we can find comparison between those of methods. At the comparison result we got that Support Vector Machine (SVM) is quite well then K-Nearest Neighbors (KNN) algorithm. On the other hand, with this successful rate of Machine Learning algorithm and we applied that on the web application and also smart application.

5.3 Recommendations

On the basis of the result and better accuracy we recommend Support Vector Machine (SVM) [26] algorithm. Because on the comparison of K-Nearest Neighbors (KNN) [25] algorithm and Support Vector Machine (SVM) [26] algorithm it got 71.02% and 81.29% respectively.

5.4 Implication for Further Study

In the early we say that we work on this research project with mammogram report based information of data. But there have some scope to do research one another type area which is image processing dataset. In this research project we work from others datasets but in future we want to work and analysis to our own dataset. On the project site we want to develop this web and smart application in deeply which for a user can test and we can provide them all the predictions like, which stages they belong in the current time, which level of breast cancer they have with more user friendly.

REFERENCES

- [1] Cancer Statistics 2107, available at <<<https://www.cancer.gov/about-cancer/understanding/statistics>>>, last accessed on (Mar 2017).
- [2] M. A. S. M. J. H. M. A. Sm Faysal Bellah, Epidemiology of breast cancer among the female patients in Bangladesh, Oriental Pharmacy and Experimental Medicine, 2016.
- [3] Available at <<<http://news.isncc.org/incidence-of-breast-cancer-in-bangladesh/>>>, last accessed on 2015.
- [4] Malignant and Benign, available at <<<https://www.radiologyinfo.org/en/info.cfm?pg=breast-cancer>>>
- [5] Syed Jamal Safdar Gardezi, Ahmed Elazab, Baiying Lei, Tianfu Wang, Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review, J Med Internet Res. 2019 Jul 26, (21(7):e14464. doi: 10.2196/14464. Review. PMID: PMC6688437).
- [6] Sharanya Rajan , Mithileysh Sathiyarayanan, Breast Cancer Awareness Through Smart Mobile Healthcare Applications from Indian Doctors' Perspective.
- [7] E-learning Opens Door to the Global Community: Novice Users' Experiences of E-learning in a Somali University - Scientific Figure on ResearchGate, available at <https://www.researchgate.net/figure/The-research-model-UTAUT-Venkatesh-et-al-2003-The-UTAUT-model-uses-four-core_fig1_283503814>>, last accessed on 1 Nov, 2019.
- [8] H.L.Story, R.R.Love, R.Salim, A.J.Roberto, J.L.Krieger, and O.M.Ginsburg, Improving Outcomes from Breast Cancer in a Low-Income Country: Lessons from Bangladesh.
- [9] Mohammad Sorowar Hossain, Shameema Ferdous, Henrike E. Karim-Kos, Review Breast cancer in South Asia: A Bangladeshi perspective, The International Journal of Cancer Epidemiology, Detection, and Prevention, 2014,pp. 465-470.
- [10] Yue, Wenbin & Wang, Zidong & Chen, Hongwei & Payne, Annette & Liu, Xiaohui, machine learning with applications in breast cancer diagnosis and prognosis, 2018.
- [11] Cristianini N, Shawe-taylor J (2000) An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, London: Cambridge University Press.
- [12] Hiba Asri, Hajar Mousannif, Hassan Al Moatassime, & Thomas Noël, using machine learning algorithms for breast cancer risk prediction and diagnosis, Procedia Computer Science, 2016 Vol.83. PP. 1064-1069.
- [13] Iliyan Mihaylov, Maria Nisheva and Dimitar Vassilev, Application of Machine Learning Models for Survival Prognosis in Breast Cancer Studies, www.mdpi.com/journal/information, 2019, 10, 93.
- [14] Lei Liu, research on logistic regression algorithm of breast cancer diagnose data by machine learning, International Conference on Robots & Intelligent System (ICRIS), 2018.
- [15] Kihan Park, Jaydev P. Desai, Machine learning approach for breast cancer localization, International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS), 2017.
- [16] V. Chaurasia and S. Pal, Data Mining Techniques : To Predict and Resolve Breast Cancer Survivability, 2014, vol.3, no.1, pp. 10–22.
- [17] Sharkas, Al-Sharkawy & Ragab, Sharkas M, Al-Sharkawy M, Ragab DA., Detection of Microclassifications in mammograms using support vector machine. UKSim 5th European Symposium on Computer Modeling and Simulation, 2011. [CrossRef] [Google Scholar].
- [18] Ragab, Sharkas & Al-sharkawy, Ragab D, Sharkas M, Al-sharkawy M., A comparison between Support Vector Machine and Artificial Neural Network for Breast cancer detection 2 the cad system, 2013. pp.171–176.
- [19] Wichakam I, Vateekul P, Learning K, Convnets CN. Combining deep convolutional networks and SVMs for mass detection on digital mammograms, 8th international conference on knowledge and smart technology (KST), 2016, pp. 239–244. [CrossRef] [Google Scholar].

- [20] Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR, using three Machine learning techniques for predicting breast cancer recurrence, 2013, J Health Med Inform 4, PP. 124.
- [21] Delen D, Walker G, Kadam A, Predicting breast cancer survivability: a comparison of three data mining methods. Artificial Intelligence in Medicine, 2005, PP. 113-127.
- [22] Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux, The NumPy array: a structure for efficient numerical computation. Computing in Science & Engineering Vol. 13 no. 2, 2011, PP. 22–30.
- [23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2011, PP. 2825–2830.
- [24] Wes McKinney and PyData Development Team, pandas: powerful Python data analysis toolkit, Version: 0.25.1, available on << <https://pandas.pydata.org/pandas-docs/stable/pandas.pdf>>>
- [25] Fix, E., & Hodges, J. L., Jr. , Nonparametric discrimination: consistency properties. Unpublished manuscript, University of California, Berkley, California, United States for the United States Air Force, 1951, February.
- [26] SUNIL RAY, available on search- Understanding Support Vector Machine algorithm from examples (along with code), 2017.
- [27] Joachims T, Making large-scale support vector machine learning practical. Advances in Kernel Methods: Support Vector Learning, 1998 MIT Press, Cambridge, MA, PP. 169-184.
- [28] Prof. Dr. Rüdiger Schulz, Wendtland Institute of Radiology, Gynaecological Radiology, University Erlangen-Nuremberg, available at << <https://data.world/julio/mammographicmasses?fbclid=IwAR0by5SHk7C4FLJIMjfBpEBF6JzHA5HVjVMHzyMkxAtUHbyQp1CPx9q13U>>>
- [29] Matthias Elter Fraunhofer, Institute for Integrated Circuits (IIS) Image Processing and Medical Engineering Department (BMT).
- [30] Onel Harrison, available on search- Machine Learning Basics with the K-Nearest Neighbors Algorithm, Sep 11, 2018
- [31] Adebola Lamidi, available on search- breast cancer classification using support vector machine (SVM), 2018.
- [32] James, Gareth, an introduction to statistical learning: with Applications in R. Springer, 2013, PP. 176. ISBN 978-1461471370.
- [33] RW Brause. Medical analysis and diagnosis by neural networks. Lecture Computer Sci 2001, PP.1-13.
- [34] Vaidehi K, Subashini TS. Breast tissue characterization using combined K-NN classifier. Indian J Sci Technol 2015, PP. 23–6.
- [35] Williams K, Idowu PA, Balogun JA, Oluwaranti A., Breast cancer risk prediction using Data Mining Classification techniques, Transactions Networks Communications, 2015, PP. 1–11.
- [36] Xindog Wu, Vipin Kumar. Top 10 algorithms in data mining. Knowledge Information Systems, 2008, PP. 1-37.
- [37] Aruna S, Rajagopalan SP, Nandakishore LV., knowledge based analysis of various statistical tools in detecting breast cancer, 2011, PP. 37–45.
- [38] TM Cover, Geometrical and statistical properties of systems of linear with applications in pattern recognition, IEEE Transactions Electronic Computers EC-14, 1965, PP. 326-34.
- [39] H. D. N. M. M. B. A. A. J. B. Mosammat Mira Pervin, Study on Clinical Presentation of Breast Carcinoma of 50 Cases, Chattagram Maa-O-Shishu Hospital Medical College Journal, 2014, PP. 09.

Apply Machine learning to predict the level of Breast Cancer

ORIGINALITY REPORT

25%

SIMILARITY INDEX

21%

INTERNET SOURCES

18%

PUBLICATIONS

19%

STUDENT PAPERS

PRIMARY SOURCES

1	www.omicsonline.org Internet Source	2%
2	www.mdpi.com Internet Source	2%
3	peerj.com Internet Source	2%
4	innovareacademics.in Internet Source	1%
5	Submitted to University of Petroleum and Energy Studies Student Paper	1%
6	core.ac.uk Internet Source	1%
7	onlinelibrary.wiley.com Internet Source	1%
8	Ibrahim, Parrish, Brown, McDonald. "Decision Tree Pattern Recognition Model for Radio Frequency Interference Suppression in NQR Experiments", Sensors, 2019	1%