Classification of Malignant and Benign tissue with Logistic Regression

A thesis submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Academic Year: 2014-2015

Islamic University of Technology (IUT)

Organization of Islamic Cooperation

Submitted by

Raihan Seraj	**(Student No: 112402)**
Razib Bin Hasan Rezanur	**(Student No: 112431)**
Mohammad Abdul Hasib	**(Student No: 112460)**

UNDER THE SUPERVISION OF

Prof. Dr. Mohammad Rakibul Islam

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

ISLAMIC UNIVERSITY OF TECHNOLOGY

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Gazipur-1704, Dhaka, Bangladesh

Classification of Malignant and Benign Tissue with Logistic Regression

A thesis presented to the Academic Faculty by

Raihan Seraj	(Student No: 112402)
Razib Bin Hasan Rezanur	(Student No: 112431)
Mohammad Abdul Hasib	(Student No: 112460)

Approved By

…………………………………………………

Prof. Dr. Mohammad Rakibul Islam
Thesis Supervisor,
Department of Electrical and Electronic Engineering

…………………………………………………

Prof. Dr. Md. Shahid Ullah
Head of the Department,
Department of Electrical and Electronic Engineering

Islamic University of Technology (IUT)

The Organization of Islamic Cooperation (OIC)

Gazipur-1704, Dhaka, Bangladesh

November – 2015

DECLARETION OF THE AUTHORSHIP

We hereby declare that the thesis titled "Classification of benign and malignant tissue with logistic regression" is an authentic record of our study carried out as the requirement for the award of degree of Bachelor of Science in Electrical and Electronic Engineering under the supervision of the **Prof. Dr. Mohammad Rakibul Islam**, Professor of the Department of Electrical and Electronic Engineering (EEE), Islamic University of Technology, Dhaka, Bangladesh during January 2015 to November 2015.The matter embodied in this thesis has not been submitted in part or full to any other university or institute for the award of any other degree.

Signature of the Authors:

……………………………..
(Raihan Seraj)

Student No.: 112402

……………………………..
(Razib Hasan Bin Rezanur)

Student No.: 112431

……………………………..
(Mohammad Abdul Hasib)

Student No.: 112460

ACKNOWLEDGEMENT

All praise and thanks to Almighty Allah who has showered us with His invaluable blessings throughout our lives, giving us strength and spirit to complete this project.

We would like to express our deepest gratitude to our project advisor **Prof. Dr. Mohammad Rakibul Islam** whose personal supervision, advice and valuable guidance helped us go through all the stages and complete appreciably our final year project. Without his motivation for work and knowledge of the project idea, the completion of the project would have been impossible.

We are also grateful to **Prof Dr. Md. Shahid Ullah**, Head of the Department of Electrical & Electronic Engineering (EEE) for his kind support.

In the end, we would like to show our deepest respect to our parents, family, friends and all those who showed patience and tenacity with us to finish with success.

ABSTRACT

Detection of breast cancer is the major phase in Cancer Diagnosis. So, classifiers with higher accuracy are always superior. A classifier already carrying high accuracy and then leading it to higher accuracy offers very less chance to a patient to be wrongly classified. This book investigates the use of a modified and improved version of the hypothesis used in the logistic regression. Both gradient descent and advanced optimization techniques are used for the minimization of the cost function. A weighting factor β was assigned in the hypothesis which is a sigmoid function. The dependency of this weighting factor to the number of features, the size of the dataset and the type of optimization technique used were observed. The accuracy was improved significantly by appropriately choosing the value of β, which, is a function of both the number of features and the type of optimization techniques used. The obtained results using the weights were promising, resulting in a significant increase in accuracy, sensitivity, and specificity.

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of figures

List of Tables

Chapter 1
Introduction

Breast cancer denotes cancer from a malignant tumor that starts in the cells of the breast tissue. A malignant tumor is a group of cancer cells that can grow into surrounding tissues or spread to distant areas of the body. Breast cancer is uncontrolled multiplication of cells in breast tissue. A group of rapidly dividing cells may form a lump or architectural distortions. The second leading cause of death among women is breast cancer, as it comes directly after lung cancer [1]. Breast cancer is a life taking disease and early detection can certainly reduce the rate of mortality. Machine learning classifiers are very popular for detecting breast cancer. Several research works have been done in this area. Here we have modified a classifier algorithm named “Logistic Regression” to detect the malignancy or benignancy of the tumorous cell more accurately.

1.1 Present scenario

It is a life taking disease & mostly the victims are women. The mortality rate due to breast cancer is very high worldwide. Today, in the United States, approximately one in eight women over their lifetime has a risk of developing breast cancer. An analysis of the most recent data has shown that the survival rate is 88% after 5 years of diagnosis and 80% after 10 years of diagnosis [2].

In 2015, an estimated 231,840 new cases of invasive breast cancer were expected to be diagnosed among women in the U.S. along with 60,290 new cases of non-invasive (in situ) breast cancer [3]. For women in the U.S. breast cancer death rates are higher than those for any other cancer, besides lung cancer [3]. Worldwide breast cancer is the leading type of cancer in women, accounting for 25% of all cases [4]. In 2012 it resulted in 1.68 million cases and 522,000 deaths [4].

It is more common in developed countries and is more than 100 times more common in women than in men [5].Outcomes for breast cancer vary depending on the cancer type, extent of disease, and person's age [6]. Survival rates in the developed world are high[7] with between 80% and 90% of those in England and the United States alive for at least 5 years[8],[9]. Lung cancer is the leading cancer site in males, comprising 17% of the total new cancer cases and 23% of the total cancer deaths.

Breast cancer is now also the leading cause of cancer death among females in economically developing countries. A shift from the previous decade during which the most common cause of cancer death was cervical cancer.

Further, the mortality burden for lung cancer among females in developing countries is as high as the burden for cervical cancer, with each accounting for 11% of the total female cancer deaths. Although overall cancer incidence rates in the developing world are half those seen in the developed world in both sexes, the overall cancer mortality rates are generally similar [10].

In Portugal, each year 4,500 new cases of breast cancer are diagnosed and 1,600 women are estimated to die from this disease [11].

Breast cancer is one of the most common cancers among Egyptian women as it represents 18.3 % of the total general of cancer cases in Egypt and a percentage of 37.3 % of breast cancer is considered treatable disease. Early diagnosis helps to save thousands of disease victims.

About 40,290 women in the U.S. are expected to die in 2015 from breast cancer, though death rates have been decreasing since 1989. Women under 50 have experienced larger decreases. These decreases are thought to be the result of treatment advances, earlier detection through screening, and increased awareness. For women in the U.S., breast cancer death rates are higher than those for any other cancer, besides lung cancer.

Besides skin cancer, breast cancer is the most commonly diagnosed cancer among American women. In 2015, it's estimated that just under 30% of newly diagnosed cancers in women will be breast cancers.

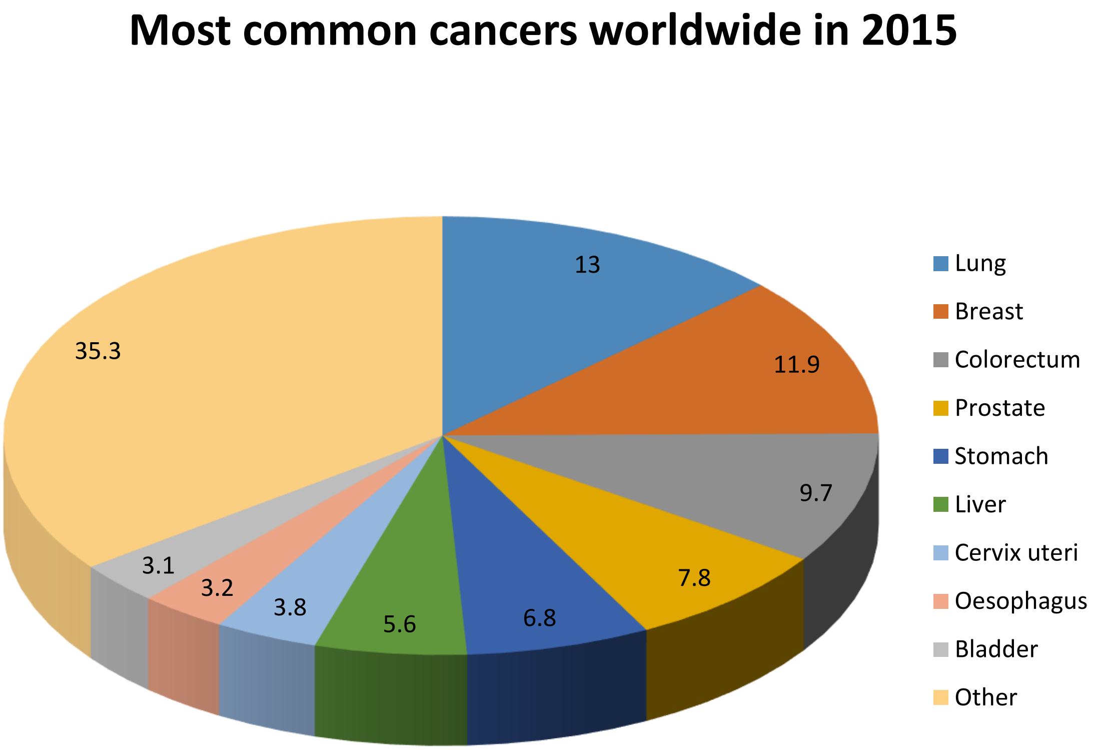

Figure 1.1: Cancer statistics worldwide in 2015

The age of breast cancer affection in Egypt and Arab countries is prior ten years compared to foreign countries as the disease targets women in the age of 30 in Arab countries, while affecting women above 45 years in European countries. Breast cancer comes in the top of cancer list in Egypt by 42 cases per 100 thousand of the population. However 80% of the cases of breast cancer in Egypt are of the benign kind [12]. Worldwide this scenario is getting really hilarious day by day. Most of the cases women are the victim. But now-a-days it is also seen in men. So the breast cancer scenario worldwide is very severe. A woman's risk of breast cancer approximately doubles if she has a first-degree relative (mother, sister, and daughter) who has been diagnosed with breast cancer. About 15% of women who get breast cancer have a family member diagnosed with it.

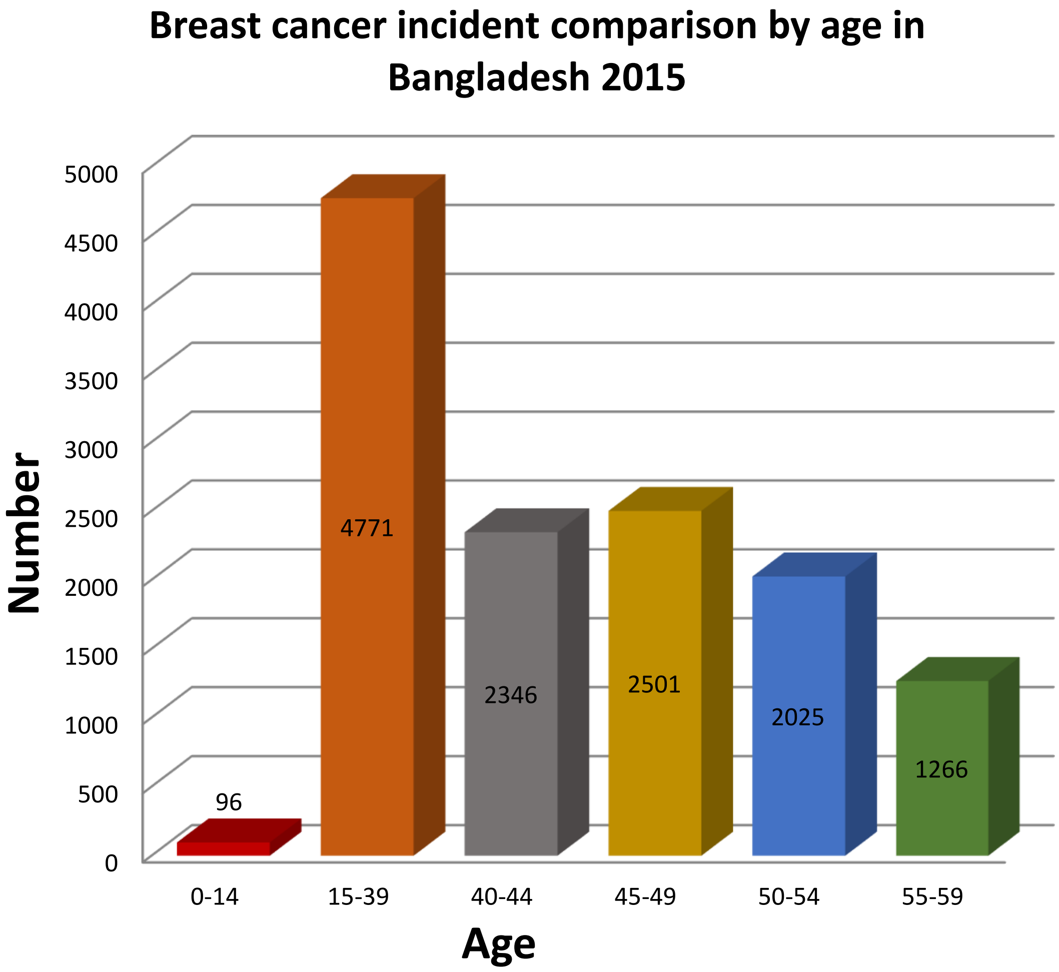

Figure 1.2: Estimated Breast Cancer comparison by age

It is estimated that each year 76,000 women die of breast cancer in South Asia (India, Bangladesh, Nepal, Myanmar, Pakistan, and Tibet). In Bangladesh, there is no national cancer registry. However, age-standardized incidence rates from Karachi, Pakistan (53.8/100,000) and Kolkata, India (25.1/100,000) (both with whom Bangladesh shares many cultural and historical similarities) suggest an annual incidence rate of 35-40/100,000. Therefore, in Bangladesh, we estimate an annual new breast cancer case burden of 30,000 women. It is projected that global breast cancer cases will grow from 1.4 million in 2008 to over 2.1 million cases in 2030. While countries with higher income celebrate significant progress toward curing women with breast cancer, low-income countries like Bangladesh are only beginning to recognize the extent and severity of the disease [13].

1.2 The need for Computer Aided Diagnosis

Computer-aided diagnosis (CAD) has become a part of the routine clinical work for detection of breast cancer on mammograms at many screening sites. This seems to indicate that CAD is beginning to be applied widely in the detection and differential diagnosis of many different types of abnormalities in medical images obtained in various examinations by use of different imaging modalities. In fact, CAD has become one of the major research subjects in medical imaging and diagnostic radiology[14]. Although early attempts at computerized analysis of medical images were made in the 1960s, serious and systematic investigation on CAD began in the 1980s with a fundamental change in the concept for utilization of the computer output, from automated computer diagnosis to computer-aided diagnosis. In this article, the motivation and philosophy for early development of CAD schemes are presented together with the current status and future potential of CAD in the environment of picture archiving and communication systems (PACS).

The number of deaths every year is increasing because of lack of awareness along with scarce medical facilities to detect breast cancer at an early stage. There is huge number of female population living in the rural areas of Bangladesh. If there is a Computer Aided Diagnosis (CAD) system present in the medical facilities in these areas to help the radiologists and doctors to diagnose and detect breast cancer among the huge number of patients there would be huge medical and social impact.

The purpose of CAD is to help Radiologists to make accurate diagnosis, provide a second opinion. CAD can minimize the operator dependent nature of Ultrasound Imaging [15]. It can obtain computational and statistical features that cannot be obtained visually. It also Increases efficiency, saves time and effort.

Radiologists need to undergo years of training and experience to properly read Ultrasound images. This is because ultrasonography is an operator dependent process. But still the diagnosis varies for different radiologists because of the randomness of the cancer lesions. Therefore, computer-aided diagnosis (CAD) has been investigated to help radiologists in making accurate diagnoses.

One advantage of a CAD system is that it can obtain some features, such as computational features and statistical features[16], which cannot be obtained visually and intuitively by medical doctors.

Another advantage is that CAD can minimize the operator-dependent nature inherent in ultrasound imaging [17] and make the diagnosis process reproducible. It should be noted that research into the use of CAD is not done so with an eye toward eliminating doctors or radiologists, rather the goal is to provide doctors and radiologists a second opinion and help them to increase the diagnosis accuracy, reduce biopsy rate, and save them time and effort. In order to increase detection and diagnosis accuracy and save to labor, computer aided detection (CAD) systems have been developed to help radiologists to evaluate medical images and detect lesions at an early stage. In general, CAD is a procedure that employs computers to assist doctors in the interpretation of medical images [18].

A CAD system is an interdisciplinary technology combining elements of digital image processing with radiological image processing. It combines image processing techniques and experts' knowledge for greatly improved accuracy of abnormality detection. In particular, the CAD system for automated detection or classification of masses and micro classification of clusters can be very useful for breast cancer control. CAD systems can provide doctors a "second pair of eyes," whose consistency and repeatability is very good [19], thus greatly reducing the false negative rate and improving the true positive rate.

1.3 Intuition on Computer Aided Diagnosis

A typical CAD application is the detection of tumors in a breast ultrasound image. Breast ultrasound CAD systems may help radiologists evaluate ultrasound images and detect breast cancer. Such systems are used in addition to the human evaluation of the diagnosis. A breast ultrasound CAD system not only improves the ultrasound image quality, increases the image contrast, and automatically determines lesion location, and it also greatly reduces the human workload associated with the diagnosis, and improves the 7 accuracy of detection and diagnosis[20].

Generally, ultrasound CAD systems for breast cancer detection involve four stages [21], as shown in fig.1.3

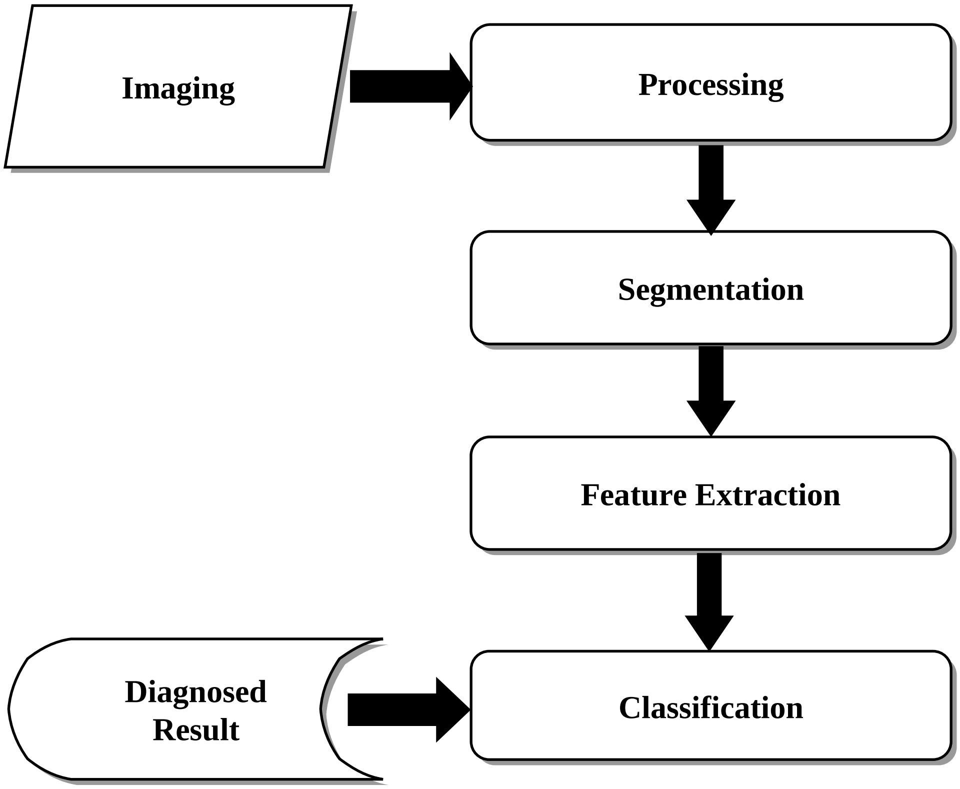

Fig.1.3: A CAD system for breast cancer diagnosis

1. Image preprocessing: The task of image preprocessing is to enhance the image and to reduce speckle without destroying the important features of BUS images for diagnosis.

2. Image segmentation: Image segmentation divides the image into non overlapping regions and it separates the objects (lesions) from the background. The boundaries of the lesions are delineated for feature extraction.

3. Feature extraction and selection: This step is to find a feature set of breast cancer lesions that can accurately distinguish between lesion and non-lesion or benign and malignant. The feature space could be very large and complex, so extracting and selecting the most effective features is very important.

4. Classification: Based on the selected features, the suspicious regions will be classified into different categories, such as benign findings and malignancy. Many machine learning techniques such as linear discriminant analysis (LDA), support vector machine (SVM) and artificial neural network (ANN) have been studied for lesion classification.

1.4 Importance of Early Detection

Early detection means using an approach that lets breast cancer get diagnosed earlier than the disease might have occurred aptly. Early detection of breast cancer can increase the rate of recovery to a great extent. Detected early breast cancer is easier to treat with fewer risks and reduces the mortality by 25% [22].

To give early treatment, it is necessary to detect it in the very early stage. Early diagnosis can save thousands of victims. Early detection of cancer greatly increases the chances for successful treatment.

There are two major components of early detection of cancer. They are:

- ➢ Education to promote early diagnosis
- ➢ Screening

Recognizing possible warning signs of cancer and taking prompt action leads to early diagnosis. Increased awareness of possible warning signs of cancer among physicians, nurses and other health care providers as well as among the general public can have a great impact on the disease. Some early signs of cancer include lumps, sores that fail to heal, abnormal bleeding, persistent indigestion, and chronic hoarseness. Early diagnosis is particularly relevant for cancers of the breast, cervix, mouth, larynx, colon and rectum, and skin [23].

Chapter 2
Dataset

When it comes to classification, there is a need of dataset to classify. So, a detailed knowledge of the dataset is certainly a handy tool.

Dataset is a statistical matrix which represents different features. It is a matrix where all the information about different features are given. Each column of the dataset represents the feature of the tumorous tissue and each row represents the number of instances. There are mainly three kinds of datasets which are mostly used in detecting the breast cancer. These are

- Wisconsin Diagnosis breast cancer (WDBC)
- Wisconsin Prognosis breast cancer (WPBC)
- Wisconsin breast cancer (WBC)

These Datasets have some features of their own.

2.1.1 Wisconsin Diagnostic Breast Cancer (WDBC)

The details of the attributes found in WDBC dataset [24]: ID number, Diagnosis (M = malignant, B = benign) and ten real-valued features are computed for each cell nucleus: Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave points, Symmetry and Fractal dimension [25]. These features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image [26].

When the radius of an individual nucleus is measured by averaging the length of the radial line segments defined by the centroid of the snake and the individual snake points. The total distance between consecutive snake points constitutes the nuclear perimeter.

The area is measured by counting the number of pixels on the interior of the snake and adding one-half of the pixels on the perimeter. The perimeter and area are combined to give a measure of the compactness of the cell nuclei using the formula.

Smoothness is quantified by measuring the difference between the length of a radial line and the mean length of the lines surrounding it. This is similar to the curvature energy computation in the snakes.

Concavity captured by measuring the size of the indentation (concavities) in the boundary of the cell nucleus. Chords between non-adjacent snake points are drawn and measure the extent to which the actual boundary of the nucleus lies on the inside of each chord.

Concave Points- this feature is Similar to concavity but counted only the number of boundary point lying on the concave regions of the boundary.

In order to measure symmetry, the major axis, or longest chord through the center, is found. Then the length difference between lines perpendicular to the major axis to the nuclear boundary in both directions is measured.

The fractal dimension of a nuclear boundary is approximated using the "coastline approximation" described by Mandelbrot. The perimeter of the nucleus is measured using increasingly larger "rulers". As the ruler size increases, decreasing the precision of the measurement, the observed perimeter decreases. Plotting log of observed perimeter against log of ruler size and measuring the downward slope gives (the negative of) an approximation to the fractal dimension.

With all the shape features, a higher value corresponds to a less regular contour and thus to a higher probability of malignancy. The texture of the cell nucleus is measured by finding the variance of the gray scale intensities in the component pixel.

2.1.2 Wisconsin Prognostic Breast Cancer (WPBC)

Details of the attributes found in WPBC dataset [23]: ID number, Outcome (R = recur, N = non-recur), Time (R => recurrence time, N => disease-free time), from 3 to 33 ten real-valued features are computed for each cell nucleus: Radius, Texture, Perimeter, Area, Smoothness, Compactness, and Concavity, Concave points, Symmetry and Fractal dimension. The thirty four is Tumor size and the thirty five is the Lymph node status.

It's known from the previous lines that the diagnosis and prognosis has the same features yet the prognosis has two additional features as follows: Tumor Size is the diameter of the exercised tumor in centimeters. Tumor Size is divided into four classes: T-1 is from 0-2 cm. T-2 is from 2-5 cm. T-3 is greater than 5cm. T-4 is a tumor of any size that has broken through (ulcerated) the skin, or is attached to the chest wall.

According to the attributes in WDBC and WPBC datasets, these attributes have 3 values with 3 columns in the data set.

The following equations demonstrate these attributes:

- **The mean** Calculated as:

$$\mathrm{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.1}$$

- **The Standard Error** calculated as:

$$S_e = \nabla \frac{S}{n} \tag{2.2}$$

Where ∇ refers to Standard error parameter, S refers to Standard deviation and n refers to sample size.

- **Worst mean or largest mean:**

  Feature selection is an important step in building a classification model. It is advantageous to limit the number of input attributes in a classifier in order to have good predictive and less computationally intensive models. Chi-square test and Principal Component Analysis are the two feature selection techniques proposed in this paper.

  Chi-square is a statistical test commonly used for testing independence and goodness of fit. Testing independence determines whether two or more observations across two populations are dependent on each other (that is, whether one variable helps to estimate the other). Testing for goodness of fit determines if an observed frequency distribution matches a theoretical frequency distribution.

  Principal Component Analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables.

The mean, standard error, and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 4 is Mean Radius, field 14 is Radius SE, field 24 is Worst Radius.

These data sets involve measurements taken permitting the Fine Needle Aspirate (FNA) test. In case that a patient is diagnosed with breast cancer, the malignant mass must be excised. After this or a different post-operative procedure, a prediction of the expected course of the disease must be determined. However, prognostic prediction does not belong either on the classic learning paradigms of function approximation or classification. This is due to a patient can be classified as a ―recurl case (instance) if the disease is observed, while there is no a threshold point at which the patient can be considered as a ―non-recur‖ case. The data are therefore censored since a time to recur for only a subset of patients is known.

2.1.3 Wisconsin Breast Cancer (WBC)

WBC datasets [23] have the following attributes:

	Attribute	**Domain**
1	Sample code number	Id number
2	Clump Thickness	1-10
3	Uniformity of Cell Size	1-10
4	Uniformity of Cell Shape	1-10
5	Marginal Adhesion	1-10
6	Single Epithelial Cell Size	1-10
7	Bare nuclei	1-10
8	Bland Chromatin	1-10
9	Normal Nucleoli	1-10
10	Mitoses	1-10
11	Class	2 for benign,4 for

Table 2.1: WBC dataset features

In the Clump thickness benign cells tend to be grouped in monolayers, while cancerous cells are often grouped in multilayer. While in the Uniformity of cell size/shape the cancer cells tend to vary in size and shape. That is why these parameters are valuable in determining whether the cells are cancerous or not.

In the case of Marginal adhesion the normal cells tend to stick together, where cancer cells tend to lose this ability. So loss of adhesion is a sign of malignancy.

In the single epithelial cell, the size is related to the uniformity mentioned above. Epithelial cells that are significantly enlarged may be a malignant cell.

The Bare nuclei is a term used for nuclei that is not surrounded by cytoplasm (the rest of the cell). Those are typically seen in benign tumors.

The Bland Chromatin describes a uniform "texture" of the nucleus seen in benign cells. In cancer cells the chromatin tends to be coarser.

The Normal nucleoli are small structures seen in the nucleus. In normal cells the nucleolus is usually very small if visible. In cancer cells the nucleoli become more prominent, and sometimes there are more of them.

Finally, Mitoses is nuclear division plus cytokines and produce two identical daughter cells during prophase. It is the process in which the cell divides and replicates. Pathologists can determine the grade of cancer by counting the number of mitoses.

2.2 Salient Features of Dataset

Each and Every dataset have some salient features like

- Uniformity of cell size
- Uniformity of cell shape
- Smoothness
- Compactness
- Clump thickness
- Marginalm adhesion etc.

2.3 Processing of the Dataset

We had to go through some processing steps before using the dataset in classification. Firstly, missing attributes had to be accounted. Secondly, cross validation had to be performed upon the dataset for improved efficiency and effectiveness.

2.3.1 Account for missing attributes

There were some missing values in the dataset. Again, in the training dataset classification label was given as “M” for Malignant and “B” for Benign. Following steps were taken:

- We have process the missing attributes of the dataset by taking in the mean value that corresponds to the features.
- The Diagnosis column of the dataset was processed so that by taking in 1 for Malignant and 0 for benign instead of M and B so for easier calculation in Binary Classification

2.3.2 Cross validation

The main aim of cross validation is to define a dataset into test and training set so that the function approximator can fit a function using training set only. This method of splitting the data is called hold out method.

In order to improve the hold out method the data set is split into K- subsets, and the holdout method is performed in the subsets. This method is known as K-fold cross validation.

The advantage of K-fold cross validation method is that it matters less how the data gets divided. Every data point gets to be in a test set exactly once, and gets to be in a training set k-1 times.

We performed 4 fold cross validation where each time one of the 4 subsets were held for testing and the rest 3 sets were set apart for training examples.

Chapter 3
Machine Learning Classifier

3.1 Insights on Machine learning classifier

Machine learning classifier is a study of algorithm that can be learned from dataset. Machine learning can be considered as a subfield of statistics. It has strong ties to artificial intelligence and optimization, which deliver methods, theory and application domains to the field.

Machine learning is employed in a range of computing tasks where designing and programming explicit, rule based algorithms is infeasible.

Machine learning algorithms identifies to which of a set of categories a new observations belongs on the basis of a training set of data containing observation whose category membership is known. It explores the construction and study of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions, rather than following strictly static program instructions.

This taxonomy or way of organizing machine learning algorithms is useful because it forces to think about the roles of the input data and the model preparation process and select one that is the most appropriate for your problem in order to get the best result.

However Machine learning is closely related to computational statistics; a discipline that aims at the design of algorithm for implementing statistical methods on computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms is infeasible.

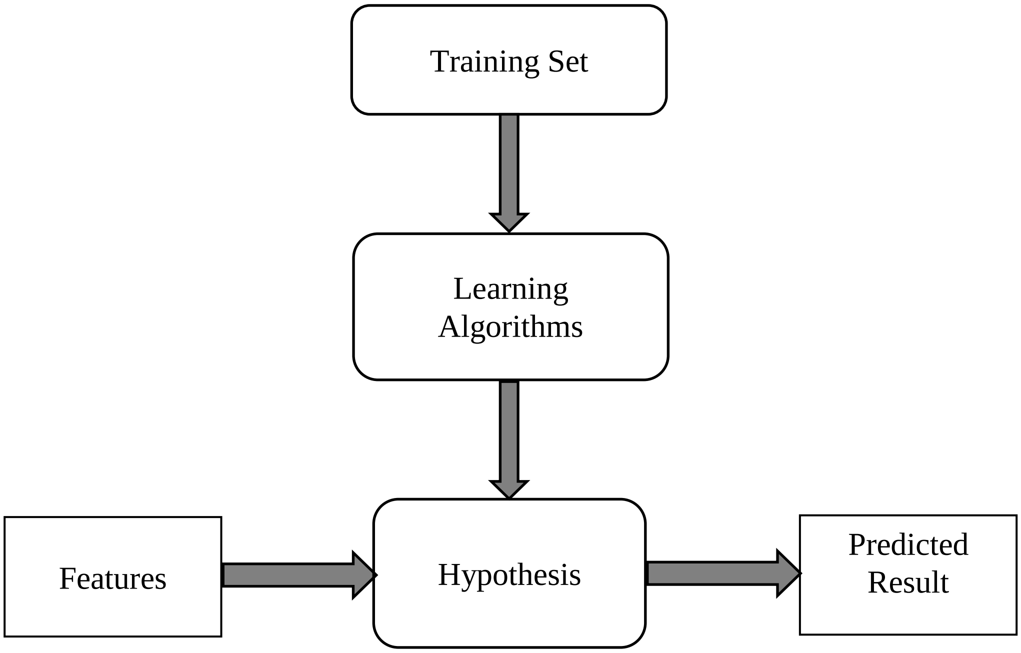

Fig 3.1: Machine Learning Classifier Process

Using dataset in the classifier, the malignancy and benignancy of tumorous cell is identified [27].

3.2 Different classifiers for the detection

There are several classifiers are existing currently. Some classifiers are

- Naïve Bayes
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Tree
- Reinforcement learning
- Neural Network
- *k*-Nearest Neighbors algorithm (KNN,IBK)

We have used Logistic Regression classifier.

3.3 Preferred use of Logistic Regression

We have worked with Logistic Regression classifier in this research work. Because it implies:

- **Probabilistic interpretation:**
  The probabilistic interpretation is easy. But it is not available in Decision Tree and SVM classifier [28].

- **Feature correlation:**
  Under this mathematical model, we do not need to worry about the features being correlated unlike Naïve Bayes [29].

- **Upgrade:**
  The model can be easily updated to take in new data. This can be done using online gradient descent.

- **Adjustment:**
  Logistic regression model helps to easily adjust Classification Thresholds when we are unsure about the confidence intervals.

3.4 Literature review

Up to now, there have been many proposed techniques for classification of breast cancer patterns with high classification accuracies.

In [30], a decision tree method (C 4.5) was used for breast cancer detection with 94.74% classification accuracy.

In [31], a rule induction algorithm based on the approximate classification method was applied to a breast cancer detection problem. The obtained accuracy was 94.99%.

In [32], linear discriminant analysis (LDA) and neural networks (NN) methods were proposed to classify the breast cancer. The accuracy of the proposed LDA + NN was 96.8%.

In [33], a support vector machine classifier was used and the obtained classification accuracy was 97.2%.

In [34], a classification scheme which was based on a feed forward neural network rule extraction algorithm was proposed. The reported accuracy was 98.10%.

A neuro-fuzzy technique was proposed by Nauck and Kruse [35]. The accuracy was 95.06%.

In [36], an AR + NN method was proposed to used in a breast cancer diagnosis problem. The obtained classification accuracy was 97.4%.

In [37], the LVQ, big LVQ, and AIRS methods were applied to breast cancer detection. 96.7%, 96.8%, and 97.2% correction classification rates were reported, respectively.

In [38], a supervised fuzzy clustering technique was proposed for breast cancer detection. The accuracy of 95.57% was obtained.

In [31], the mixture experts (ME) network structure was proposed for breast cancer diagnosis. The authors reported 98.85% correct classification rate for this technique.

An improved Bayesian belief networks (BBNs) using the linear regression technique was applied to breast tumor gene classification in [39]. The authors reported that their proposal could effectively recognize important genes.

An isotonic separation approach for breast cancer detection was proposed in [40]. The proposed method yielded good achievements in breast cancer detection.

In [41], a Bayesian classifiers performance was evaluated for the diagnosis of breast cancer using two different real datasets.

In [42], 1631 different instances were considered by the authors for simulating the breast pathology using BBNs.

A combination of statistical methods and particle swarm optimization (PSO) for mining breast cancer patterns was proposed in [43].

Chapter 4
Performance parameters

This chapter discusses about performance parameters of classifiers. It includes discussion on Receiver Operating Characteristic (ROC) curve, confusion matrix etc. Later complexities regarding algorithms are also explained.

4.1 Receiver Operating Characteristic curve

ROC- Receiver Operating Characteristic curve shows a graphical representation against the true positive rate against false positive rate. A good classifier has a maximum area under the ROC curve[44].

A Receiver Operating Characteristic (ROC) curve succinctly represents the simultaneous variation of TPR and FPR in the same plot. Figure 2.1 shows several ROC curves for different hypothetical classifiers. For an ideal classifier the ROC curve has 100% Area Under the ROC Curve (AUC). However, a real classifier’s AUC falls below that. One logical question arises where the operating point on ROC should be chosen. This depends on the application, e.g, medical diagnosis or radar detection, related with the ROC. In cancer diagnosis in the breast, the operating point needs to be chosen so that TPR or *sensitivity* is 100% even the cost of high FPR or 1-specificity and thus the overall accuracy.

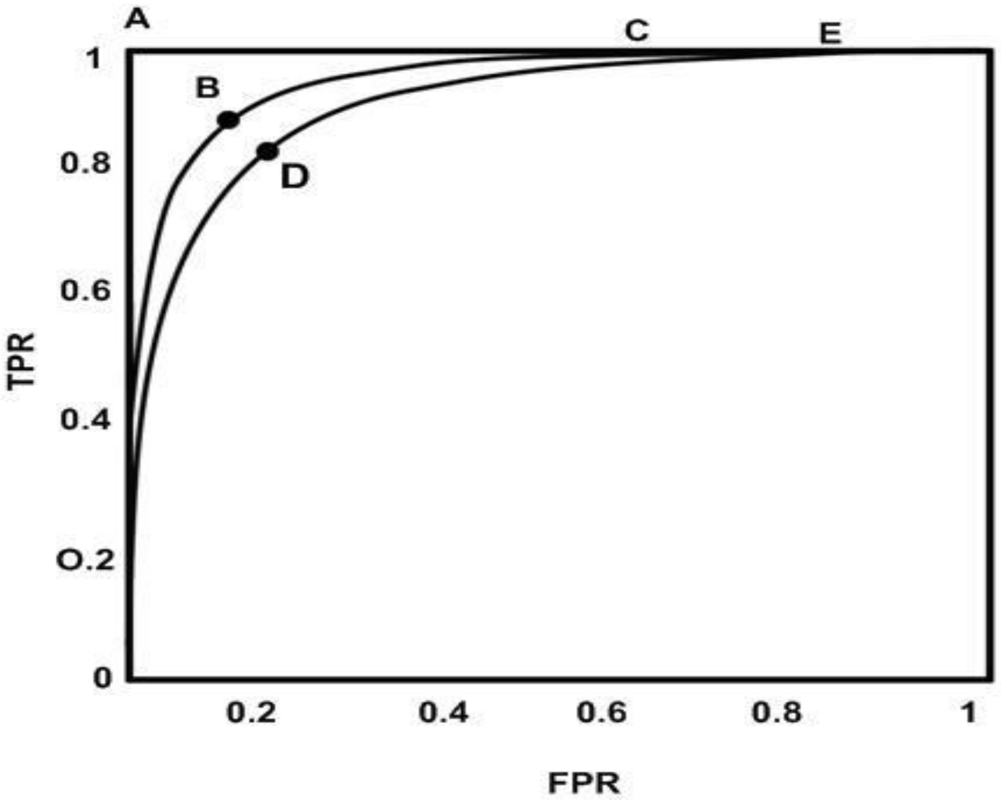

Figure 4.1: Selection of operating points in ROC of a realistic classifier used in medical diagnosis: 'A' is a point in an ideal classifier, whereas 'B', 'C', 'D' and 'E' are the points in realistic classifiers.

However, such high value of FPR implies false cases diagnosed as cancerous and this ultimately would repair the biopsy a good classifier needs to have an operating point that would rule out any false negative case as well as false positive case.

In other words, the *accuracy* of the classifier needs to be close 100%. In the present of such classifier only the cases diagnosed as malignant would be forwarded for the biopsy. Thus, such reduction of the number of biopsy will reduce the monitory involvement and use of the other resources of the patients and the relevant authorities.

For a realistic classifier employed for cancer diagnostic, it needs to operate at the point of ROC curve where sensitivity is 100% and 1-specificity is minimum.

The point ‘A’ in Figure 2.1 denotes the operating point of an ideal classifier that represents 100% TPR with 0% FPR, i.e., 100% accuracy, thus lies on an ideal ROC curve with a perfect AUC of 1. A random classifier, i.e., one that classifies randomly, would form a diagonal line as a ROC curve with an AUC of 0.5.

Points ‘B’,’C’,’D’ and ‘E’ shows the operating points for more realistic classifiers. While points ‘B’ and ‘D’ are the operating points that may correspond to the highest accuracy rate, points ‘C’ and ‘E’ need to be chosen as operating points for the realistic classifiers that correspond to 100% TPR with the minimum FPR.

The closer point ‘C’ or ‘E’ towards ‘A’, the better the classifier. For example, point ‘C’ lies on a better classifier compared to point ‘E’.

In this work, an enhanced classifier based on Sparse Representation-based Classifier (SRC) is represented that provides us near-to-ideal operating points in the ROC curve with an AUC of 0.9754.

4.2 Confusion matrix

Confusion matrix (also known as contingency table or error matrix or matching matrix) can be specified as a performance evaluating criteria for machine learning classifiers. It visualizes the predicted outcomes and actual results from which important parameters i.e. accuracy, sensitivity, specificity can easily be calculated.[45] It actually shows the relationship between predicted & actual classifications. The following demonstration shows a confusion matrix for Breast Cancer datasets.

Confusion matrix is a visualization tool which is commonly used to present the accuracy of the classifiers in classification. It is used to show the relationships between outcomes and predicted classes. The level of effectiveness of the classification model is calculated with the number of correct and incorrect classification in each possible value of the variable being classified in the confusion matrix.

		Actual	
		Predicted Malignant	Predicted Benign
Predicted	Actual Malignant	True Positive TP	False Negative FN
	Actual Benign	False Positive FP	True Negative TN

Figure 4.2: Confusion matrix

4.2.1 Entries in confusion matrix

True Negative (TN): The classifier has classified the instance as Benign & actually it is Benign.

False Positive (FP): The classifier has classified the instance as Malignant & actually it is Benign.

False Negative (FN): The classifier has classified the instance as Benign & actually it is Malignant.

True Positive (TP): The classifier has classified the instance as Malignant & actually it is Malignant.

Only the True Negatives (TN) & true Positives (TP) are correct classifications.

- $\mathbf{Condition\ Positive\ =\ Pos}$
- $\mathbf{Condition\ Negative\ =\ Neg}$
- $\mathbf{True\ Positive\ =\ TP}$
- $\mathbf{False\ Positive\ =\ FP}$
- $\mathbf{True\ Positive\ Rate\ (TPR)\ or\ Sensitivity\ =\frac{\mathit{TP}}{\mathit{Pos}}}$ (4.1)
- $\mathbf{False\ Positive\ Rate\ (FPR)or\ 1-Specificity=\frac{\mathit{FP}}{\mathit{Neg}}}$ (4.2)
- $\mathbf{Accuracy\ =\frac{\mathit{TP+TN}}{\mathit{Pos+Neg}}}$
  (4.3)
- $\mathbf{Positive\ Predictive\ Value\ (PPV)\ =\frac{\mathit{TP}}{\mathit{Y}}}$
- $\mathbf{Negative\ Predictive\ Value\ (NPV)\ =\frac{\mathit{TN}}{\mathit{N}}}$ (4.6)
- $\mathbf{Precision\ =\frac{\mathit{TP}}{\mathit{Y}}}$
  (4.7)
- $\mathbf{Recall\ =\frac{\mathit{TP}}{\mathit{Pos}}}$
- $\mathbf{F-measure\ =\frac{2}{\frac{1}{\mathit{Precision}}+\frac{1}{\mathit{Recall}}}}$ (4.8)

4.3 Algorithm Complexity

Algorithmic complexity is concerned about how fast or slow particular algorithm performs. We define complexity as a numerical function T(n) - time versus the input size n.[46] We want to define time taken by an algorithm without depending on the implementation details. A given algorithm will take different amounts of time on the same inputs depending on such factors as: processor speed; instruction set, disk speed, brand of compiler and etc. The way around is to estimate efficiency of each algorithm asymptotically. We will measure time T(n) as the number of elementary "steps" (defined in any way), provided each such step takes constant time.

The goal of computational complexity is to classify algorithms according to their performances. We will represent the time function T(n) using the "big-O" notation to express an algorithm runtime complexity.

For any monotonic functions f(n) and g(n) from the positive integers to the positive integers, we say that f(n) = O(g(n)) when there exist constants c > 0 and $n_0 > 0$ such that

$f(n) \leq c * g(n)$, for all $n \geq n_0$ (4.9)

Intuitively, this means that function f(n) does not grow faster than g(n), or that function g(n) is an upper bound for f(n), for all sufficiently large n→∞

Here is a graphic representation of f(n) = O(g(n)) relation:

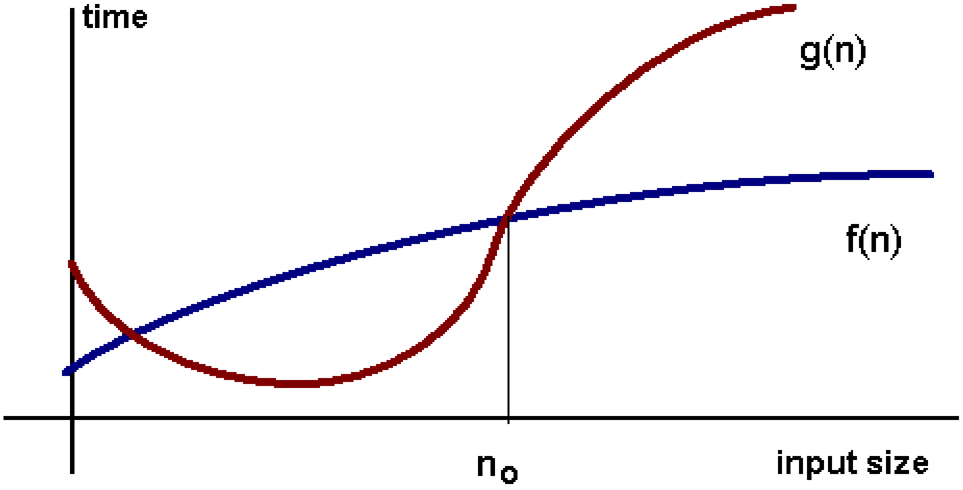

Figure 4.3: Algorithm Complexity explanation

Constant Time: O(1)

An algorithm is said to run in constant time if it requires the same amount of time regardless of the input size. Examples:

- Array: accessing any element
- Fixed-size stack: push and pop methods
- Fixed-size queue: enqueue and dequeue methods

Linear Time: O(n)

An algorithm is said to run in linear time if its time execution is directly proportional to the input size, i.e. time grows linearly as input size increases. Examples:

- Array: linear search, traversing, find minimum
- Array List: contains method
- Queue: contains method

Logarithmic Time: O(log n)

An algorithm is said to run in logarithmic time if its time execution is proportional to the logarithm of the input size. Example:

- **Binary search**

Since in our brute force search algorithm we perform a linear search the algorithm complexity therefore increases with the increase in the range of values being searched. The algorithm therefore has a linear time complexity.

Chapter 5
Proposed model

This chapter of our book discusses about one of the many classification algorithms that are used. This chapter gives an insight on logistic regression and the associated hypothesis, the cost function of logistic regression. It also provides a detailed idea regarding the working principles of Logistic Regression and its features.

5.1 Logistic Regression

The binary logistic regression model is used to predict binary response based on one or more predictor variables (features). Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution. [47]

The term regression comes from the fact that we are fitting a linear model to the feature space. Logistic Regression involves a more probabilistic view of classification.[48]

The outcome of logistic regression should be discreet and not continuous however, the logistic regression can work on multidimensional feature space (features can be categorical or continuous)

5.1.1 Hypothesis and Cost Function

The model for logistic regression involves a vector β in d- dimensional feature space. For a point X in feature space, we project it onto α to convert it into a real number z in the range -∞ $to + \infty$

$$\boldsymbol{Z = \alpha + \beta . x = \alpha + \beta_1 . X_1 + \cdots \beta_d . X_d} \tag{5.1}$$

The value of Z thus obtained is then mapped to the range from 0 to 1 using the logistic function

$$\boldsymbol{p = \frac{1}{1 + e^{-z}}} \tag{5.2}$$

Thus this function transforms a point X in the d dimensional feature space in to a range of 0 to 1.

Thus by applying a probability threshold corresponding to a particular class, we can identify a particular class for a corresponding feature subsets.

In order to fit in the optimum parameters β to the real number Z, we plug in the logistic function into the following cost function.

$$-\frac{1}{m}\left[\sum_{i=1}^{m}\left[y^{(i)}log(hyp) + \left(1 - y^{(i)}\right)log(1 - hyp)\right]\right] \tag{5.3}$$

Here 'I' corresponds to the ***i***th training example and ***hyp*** is our proposed hypothesis which is the sigmoid function.

The value of β was initialized to be zero, then the corresponding cost was calculated, then the value of β is adjusted, so that the cost function is minimized.

The particular cost function chosen for logistic regression is a bowl-shaped cost function having a global minima. The probability threshold for malignant and benign tumors was set to be 0.5.

Thus any probability greater than 0.5 will result in the classified output to be malignant and any probability less than 0.5 will result in the classified output to be benign.

The cost function chosen will penalize the algorithm with high cost, if the predicted output was malignant (1) instead of (0), alternatively if the predicted output was benign (0) instead of malignant (1) the cost function will also penalize with a high cost.

5.1.2 Sigmoid function

Sigmoid function is the most commonly known function used in feed forward neural networks because of its nonlinearity and the computational simplicity of its derivative[49].Sigmoid function is the logistic function of the form

$$\boldsymbol{p = \frac{1}{1 + e^{-z}}} \tag{5.4}$$

This sigmoid function maps the value of Z to a range from 0 to 1. The sigmoid function has the following form.

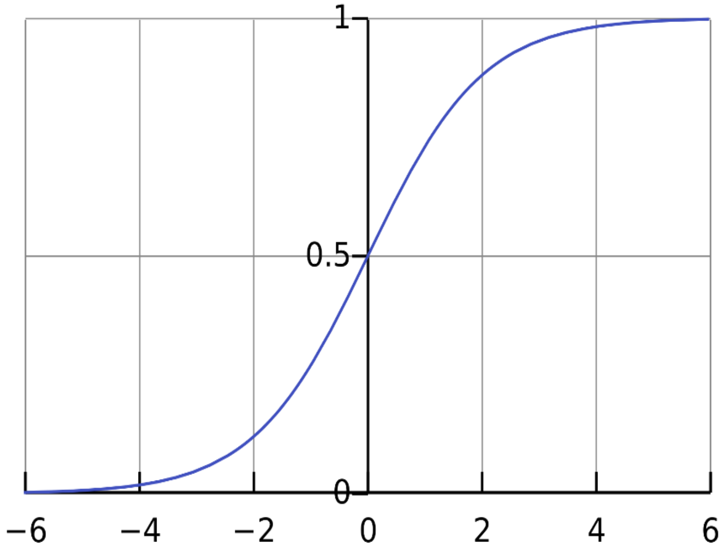

Figure 5.1: Sigmoid Function

The threshold chosen was selected to be 0.5. The hypothesis predicts an output 1 for a probability greater than 0.5 and an outcome of 0 for a probability less than 0.5.

5.1.3 Proposed weighted function

We proposed a weighted hypothesis, where we add a weight factor θ to the existing logistic function. We use brute force search algorithm to find an optimum value for the weight factor what will yield a higher accuracy, sensitivity, specificity. The added weight factor will make an aggressive or submissive approach towards classification. The following expression shows the modified hypothesis with a dynamic weight added to it.

$$\boldsymbol{p = \frac{1}{1 + e^{-\theta z}}} \tag{5.5}$$

Using a weighted sigmoid function, we have observed a significant improvement in the classifier performance compared with the one which used the classical techniques of logistic regression.

5.2 Optimization Techniques

Optimization techniques are used in order to find suitable values of β that will minimize the cost function, so that with this optimized value of β classification can be performed on the test dataset. In our work, we used two optimization techniques:
I) Stochastic Gradient Descent
II) Advanced Optimization Technique

5.2.1 Gradient Descent Algorithm

Gradient descent is a first order optimization algorithm to find the global minima. In order to find the minima, one takes steps proportional to the negative of the gradient (or of the approximate gradient) of the function at the current point.

The gradient descent algorithm has the following format

Repeat {

$$\boldsymbol{\beta_{new}} := \boldsymbol{\beta_{old}} - \mu \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \left(\boldsymbol{hyp} - \boldsymbol{y}^{(\boldsymbol{i})}\right) * \boldsymbol{x}^{(\boldsymbol{i})} \tag{5.6}$$

}

In this optimization technique, we simultaneously update the value of β and iterate the whole process, until a fairly global minima is reached. Thus, we have observed that with an iteration of 100 we fairly reached the global minima. Here μ is called the learning rate which is a measure of the steps that the algorithm will take in order to converge. A low value of μ will mean that the algorithm will require more steps to reach the global minima where as a high value of μ will mean that the cost function will converge at a faster rate[50]. However, a very high value of the learning rate although leads to a faster convergence does not necessarily imply that the function is guaranteed to reach the global minima. Hence in practice an optimum value of the learning rate is chosen which makes a tradeoff between the convergence rate and the global minima reached. In our work, we have chosen the learning rate to be 0.8 and have fairly reached towards the global minima.

One of the limitations of such an algorithm is that it only works well in finding the global minima of a convex function. If the function is not convex in practice, this algorithm will only help to reach the local minima of the cost function instead of proceeding towards the global minima.

5.2.2 Advance optimization Technique

In the advanced optimization technique, we do not have to manually choose the value of the learning rate μ, which is chosen automatically. In the advanced optimization technique, we have used the **BFGS (Broydon Fletcher Goldfarb Shanno)** algorithm. **BFGS** is an iterative method for solving unconstrained nonlinear optimization problems. The **BFGS** method approximates Newton's method, a class of hill-climbing optimization techniques that seeks a stationary point of a (preferably twice continuously differentiable) function. For such problems, a necessary condition for optimality is that the gradient be zero. Newton's method and the **BFGS** methods are not guaranteed to converge unless the function has a quadratic Taylor expansion near an optimum. These methods use both the first and second derivatives of the function.[51]

Such an algorithm has been proven to perform fairly well even for non-convex function where there may be multiple locally optimal points and it can take a lot of time to identify whether the problem has no solution or if the solution is global.

5.3 Brute Force technique for finding optimum weights

In this section we have used our modified sigmoid function and have placed a dynamic weight of θ as discussed before. We use a brute force or exhaustive search algorithm and varied the value of θ from 0.01 to 1 with a step size of 0.01 and from 1 to 100 with a step size of 0.5. Our search algorithm takes into account the performance parameters of the classifier outputs for each value of θ and finds the optimum value of the weight θ that gives the greatest accuracy, and an improved sensitivity and specificity. The search algorithm compares the performance parameters each time with the previously obtained performance parameters of a particular value of θ. Since the performance parameters somehow changes arbitrarily with the value of θ we therefore compare the parameters for the adjacent values of θ and put the higher performance parameters on a secondary array. The search algorithm then chooses the value of θ from the secondary array which will lead towards obtaining the maximum accuracy, sensitivity and specificity.

5.3.1 Relationship of the weighted hypothesis

We dynamically varied our weight factor θ and obtained the value for θ that gives the maximum classifier performance. The proposed system of adding weights to the sigmoid function was tested with different datasets where we have observed that the addition of the weight factor always leads to an improvement in the classification performance. Moreover our proposed system gave further insights about the magnitude of the weight factor being added to the sigmoid function.

The proposed system was tested with Wisconsin Breast Cancer Dataset consisting of 32 features. Keeping the optimization technique constant to advance optimization technique, when classification was performed with logistic regression on the larger dataset comprising of 32 features compared to 12 in Wisconsin Diagnostic Breast Cancer dataset, an important observation came into role. The magnitude of the weight factor θ increases as the number of features in the dataset increases.

The proposed system was again tested, this time keeping the size of the data fixed and varying the optimization techniques used to reach the global minima of the cost function. The value of θ seems to change depending on whether stochastic gradient descent algorithm or advanced optimization technique was used.

Summarizing the above observation we came to a conclusion that

$$\boldsymbol{\theta = f(optimisation\ technique, number\ of\ features\ in\ the\ dataset)} \tag{5.7}$$

I.e. θ is a function of optimization technique and the number of features in the dataset.

Such an important observation leads us to improve our search algorithm where we can initialize our value of θ to a higher in order to reach the optimum value of θ and save a lot of computation time.

Chapter 6
Result analysis

6.1 Results from Matlab simulation for small features

The proposed system comprising of a weighted sigmoid function was tested on WDBC (Wisconsin Diagnostic Breast Cancer) dataset. A twofold cross validation was performed on the dataset and each time half of the dataset was held as the test set. The next section will give a performance comparison when logistic regression was applied with and without the weighted sigmoid function.

6.1.1 Results obtained from the Classical System

The classical logistic regression without incorporating the weighted sigmoid function, when applied to the dataset consisting of 12 features lead to the following performance parameters. The parameters remained same of each of the 2 cross validated subsets set aside for test purposes.

- **Accuracy:** 95.7082
- **Error Rate:** 0.0429
- **Sensitivity:** 0.9944
- **Specificity:** 0.8333
- **Confusion matrix:** $\begin{matrix} 178 & 1 \\ 9 & 45 \end{matrix}$

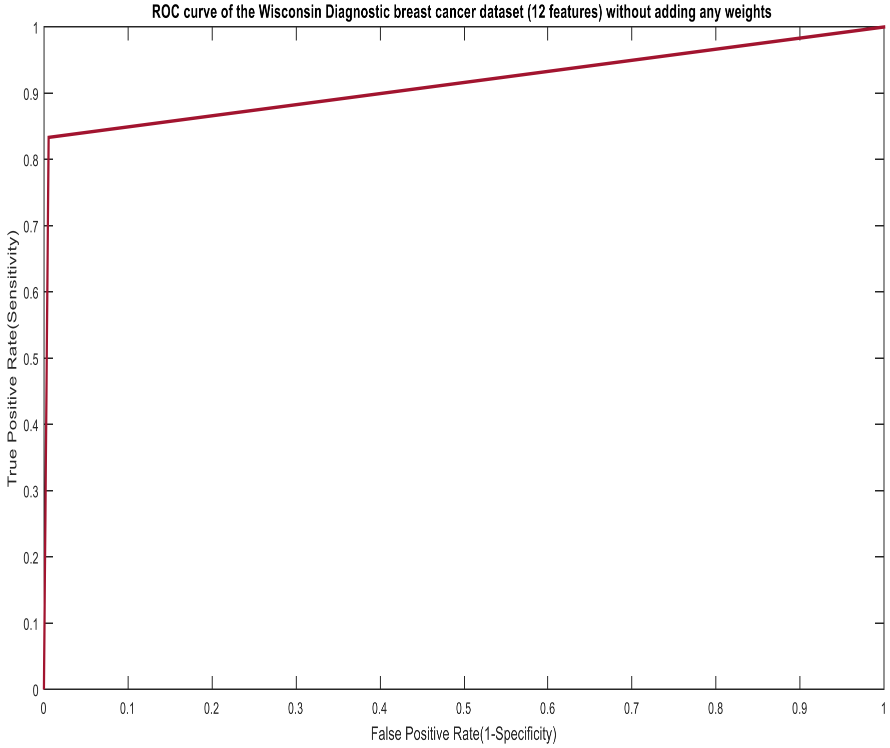

Figure 6.1: Result of classical system for smaller dataset

6.1.2 Results Obtained from the Proposed System

The proposed system has been implemented on the same dataset and using the brute force search algorithm, an optimum weight factor of 0.5 in the sigmoid function lead to the following maximum achievable performance parameters

- **Accuracy:** 97.4249
- **Error Rate:** 0.0258
- **Sensitivity:** 0.9944
- **Specificity:** 0.9074
- **Confusion matrix:** $\begin{matrix} 178 & 1 \\ 5 & 49 \end{matrix}$

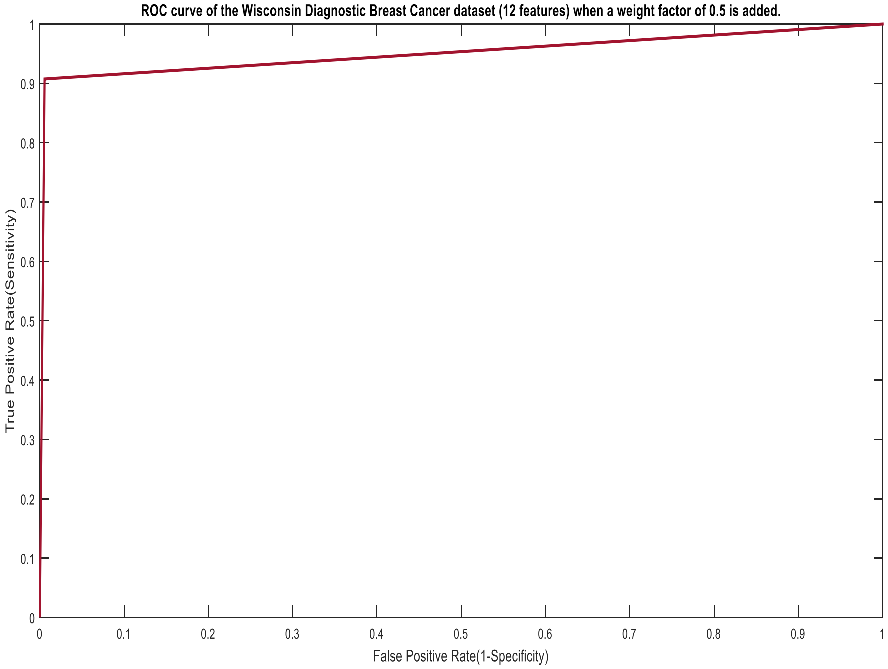

Figure 6.2: Result of proposed system for smaller dataset

Performance Parameters	Existing Method	Proposed Method
Accuracy	95.7082	97.4249
Sensitivity	0.9944	0.9944
Specificity	0.8333	0.9074
Error rate	0.0429	0.0252
Confusion matrix	178 1 9 45	178 1 5 49

Table 6.1: Performance parameters of existing and proposed method for smaller dataset

Comparing the two figures and the performance parameters above, it is fairly easy to visualize the significant change and improvements in the accuracy, sensitivity and specificity when the weighted sigmoid function has been applied. The area under the receiver operating characteristic curve has been increased thus reflecting an improvement in the performance parameters.

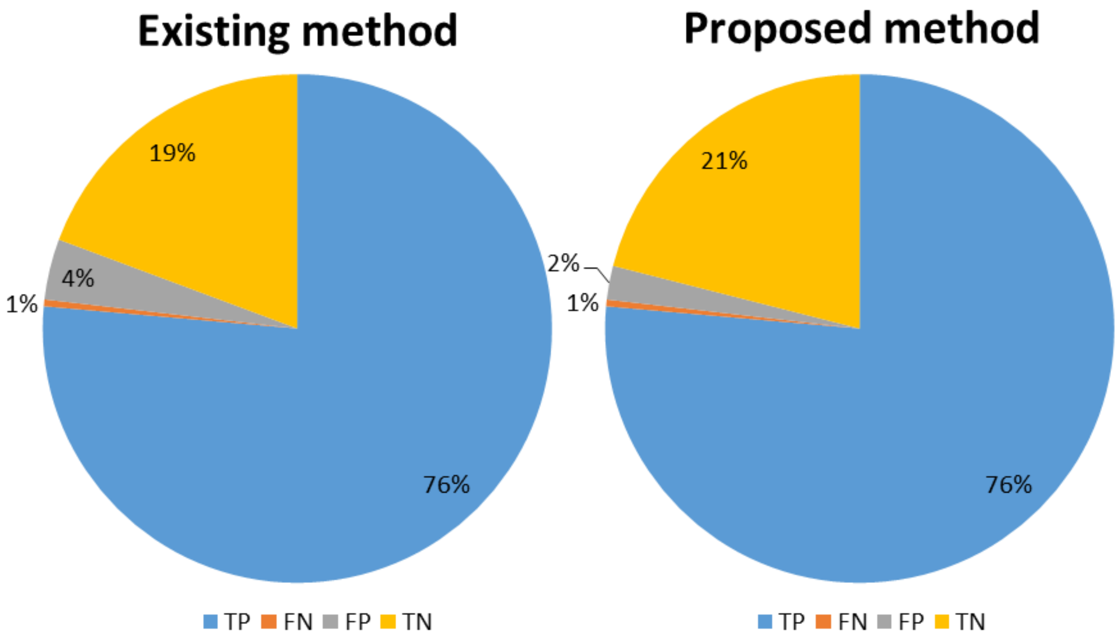

Figure 6.3: Entries of confusion matrix from existing and proposed method for smaller dataset

6.2 Results from Matlab simulation for large features

In order to ensure the effectiveness of the proposed system, the modified sigmoid function was applied to another dataset. The WBC (Wisconsin Breast Cancer) dataset is a larger dataset consisting of 32 features. The optimization technique was kept constant and hence an advance optimization technique with BFGS algorithm was deployed. The classical method without adding the weight factor in the hypothesis and with adding the weight factor was analyzed and the following performance parameters were obtained.

6.2.1 Results obtained from the Classical System

The following performance parameters were obtained when no dynamic weights were added in the sigmoid function. The two fold cross validation technique had been applied, and the same performance parameters were obtained for each subsets, set aside for testing purposes.

- **Accuracy:** 95.4225
- **Error Rate:** 0.0458
- **Sensitivity:** 0.9539
- **Specificity:** 0.9552
- **Confusion matrix:** $\begin{matrix} 207 & 10 \\ 3 & 64 \end{matrix}$

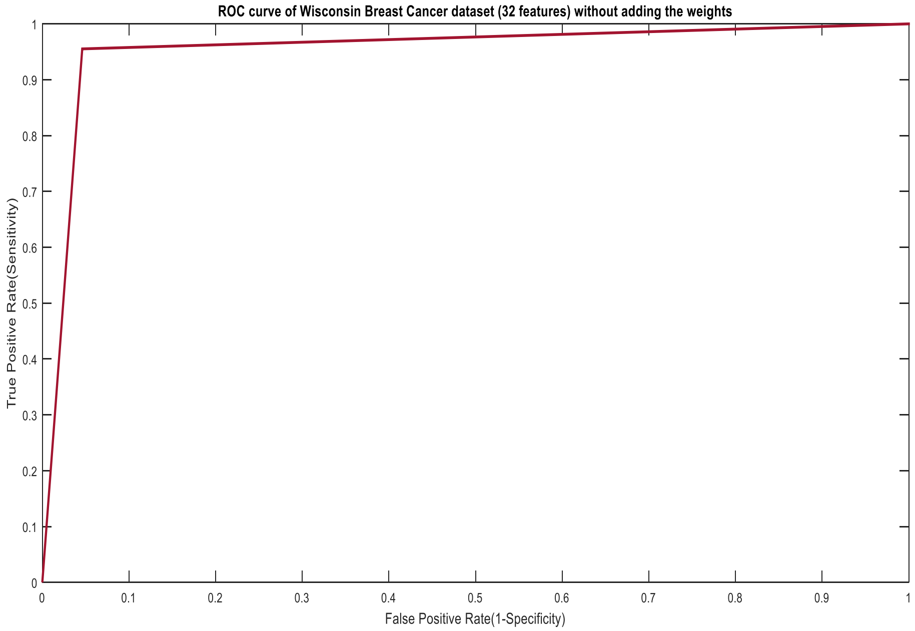

Figure 6.4: Result of existing system for larger dataset

6.2.2 Results Obtained from the Proposed System

The proposed system has been applied to the larger dataset and an addition of a dynamic weights to the sigmoid function using a brute force search algorithm led to an improvement in the ROC curve. The increase in area under the ROC curve also reflects the significant improvement that the proposed hypothesis brings.

The search algorithm found the value of the dynamic weight to be 3 for the larger dataset. The both the subsets, held aside for testing purposes in a 2 fold cross validation gave the following results.

- **Accuracy:** 96.8310
- **Error Rate:** 0.0317
- **Sensitivity:** 0.9631
- **Specificity:** 0.9851
- **Confusion matrix :** 209 8

  1 66

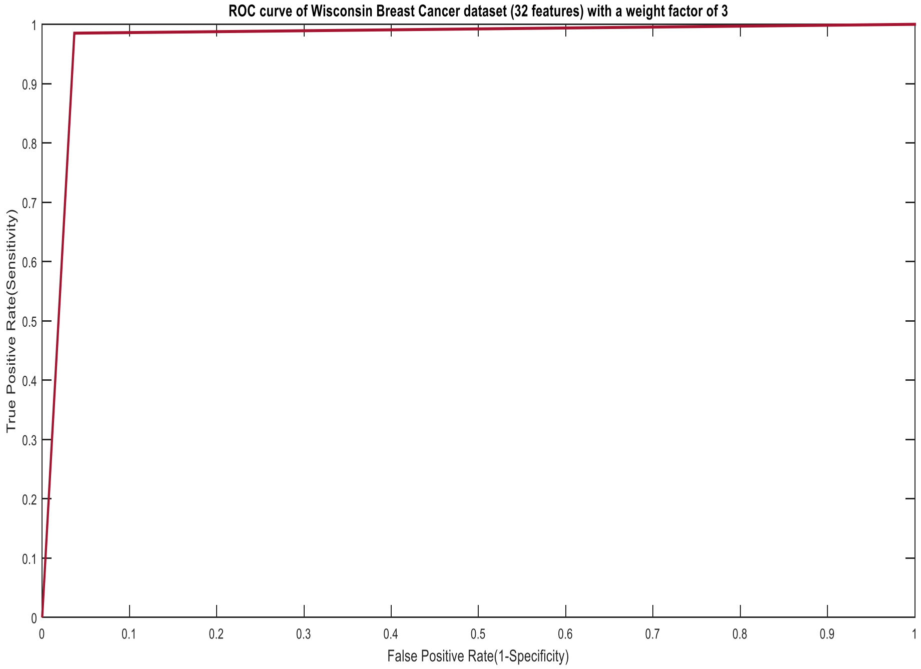

Figure 6.5: Result of proposed system for larger dataset

Comparing the two figures and the performance parameters of the proposed system with that of the classical system, the improvements in the performance of the proposed system is evident. The following section gives further insight of an overall comparison for both the datasets

Performance Parameters	Existing Method	Proposed Method
Accuracy	95.4225	96.8310
Sensitivity	0.9539	0.9631
Specificity	0.9552	0.9851
Error rate	0.0458	0.0317
Confusion matrix	207 10 3 64	209 8 1 66

Table 6.2: Performance parameters of existing and proposed method for larger dataset

Here is the comparison of different entries from confusion matrix.

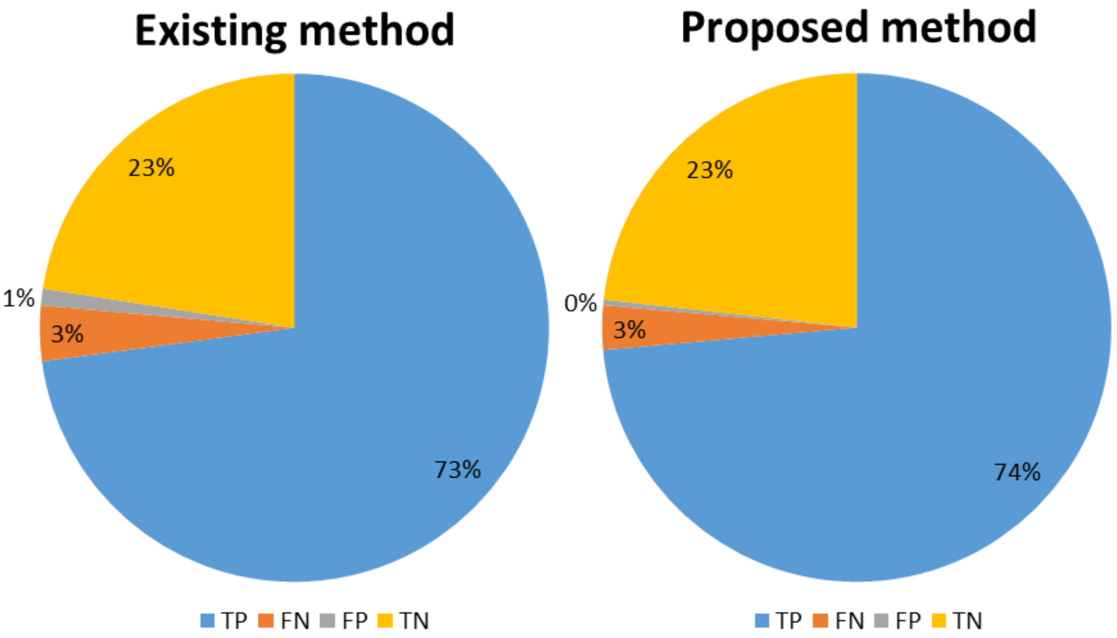

Figure 6.6: Entries of confusion matrix from existing and proposed method for larger dataset

6.3 Overall performance comparison of the proposed system with the Classical System

The proposed modified sigmoid function successfully improved the performance upon the addition of a dynamic weight for both the datasets. Keeping the optimization technique constant the trend of the magnitude of the weighted factor was observed. The magnitude of the weight being added grows with the number of features in the dataset. Similar experiments were performed keeping the size of the data same but changing the optimization technique. Under such circumstances the value of the dynamic weights were also found to be different, thereby concluding the fact that this factor is a function of both the data size and the optimization technique.

The following graph summarizes the comparison between the two methods for both the datasets.

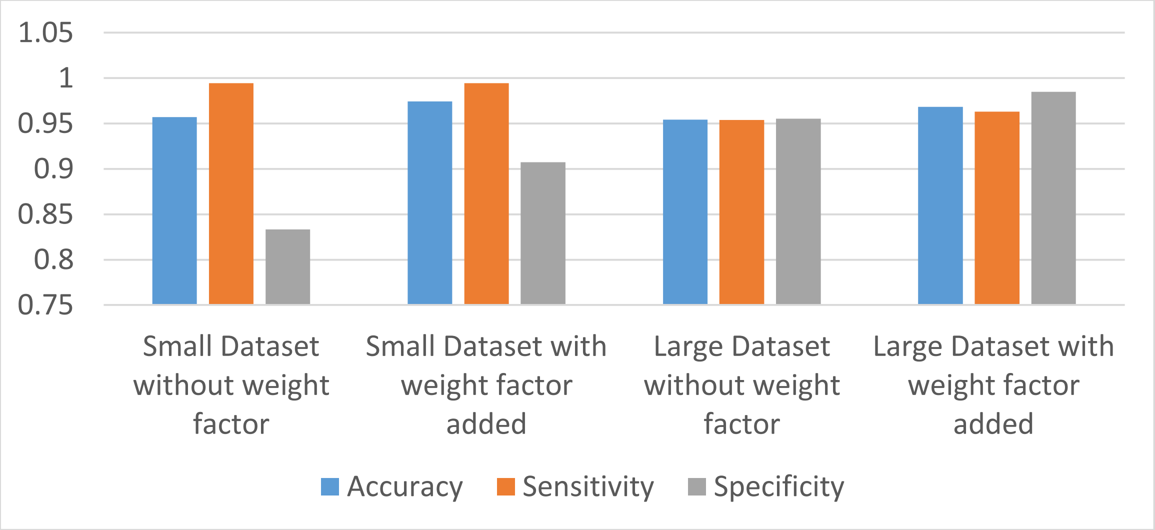

Figure 6.7: Overall performance comparison

Chapter 7
Conclusion & Future work

7.1 Conclusion

We used an improved version of the logistic regression for the binary classification of the breast cancer data set. We have proposed a modified sigmoid function which has a dynamic weight added to it. The improved hypothesis includes a weighting factor β where a unique value of β exists that will maximize the classifier performance. The value of that unique weight factor β depends on the data set and a higher value of the β should be assigned for a dataset with larger number of features.

Furthermore, we have only managed to find the dependency of the weight factor with the size of the data set and the optimization technique. However, the exact relationship is yet to be worked upon. If the exact relationship with the data size can be found then the complexity of the search algorithm can be reduced significantly by initializing the weight with the desired value so that the algorithm does not have to search unnecessarily through the weights. In addition to it, we can also bring certain improvements, by plotting in different features with each other and observe to the extent to which they are correlated. Another important insight towards improving the results obtained by logistic regression, specially our proposed method would include addition of online learning algorithm so as to discover the attributes associated with malignant and benign tumors. We can incorporate methods of automatic feature selection, in case of our data set which has large number of features we can automatically select features that will most likely to contribute towards the classification of benign or malignant tumor.

In order to reduce the features or enhance it if there are less number of features, a method of regularization could also be undertaken so as to improve the accuracy of our proposed system.

7.2 Future work

Logistic regression is a supervised learning in which the classes are labeled. However future work regarding the use of unsupervised learning incorporating this weighted logistic regression can also take a better approach. For instance Reinforcement learning has been used for solving the classification problems. One of the methods involve using neural network and multilayer perceptron that serve as function approximators[52]. Similar technique can be used along with this weighted logistic regression that serve as a function approximators and will aid in a better classification result if used together with reinforcement learning algorithm.

References

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA: a cancer journal for clinicians, 2015. **65**(1): p. 5-29.

2. Masters, G.A., et al., Clinical cancer advances 2015: annual report on progress against cancer from the Ameriican Society of Clinical Oncology. Journal of Clinical Oncology, 2015: p. JCO. 2014.59. 9746.

3. ; Available from: http://www.breastcancer.org.

4. *World Cancer Report 2015.* World Health Organization, 2015: p. Chapter 1.1.

5. Chen, W., et al., *Annual report on status of cancer in World, 2014.* Chinese Journal of Cancer Research, 2015. **27**(1): p. 2.

6. *Breast Cancer Treatment (PDQ®).* NCI, 2014-06-26. Retrieved 29 June 2014.

7. "World Cancer Report".International Agency for Research on Cancer. 2008. Retrieved 2011-02-26.

8. "Cancer Survival in England: Patients Diagnosed 2007–2011 and Followed up to 2012". Office for National Statistics. 29 October 2015. Retrieved 29 June 2015.

9. "SEER Stat Fact Sheets: Breast Cancer". NCI. Retrieved 18 June 2015.

10. Jemal, Ahmedin, et al. "Global cancer statistics." CA: a cancer journal for clinicians 61.2 (2011): 69-90.

11. Veloso, V., “Cancro da mama mata 5 mulheres por dia em Portugal,”. In: (Ed.) CiênciaHoje. Lisboa, Portugal, 2009".

12. Elattar, Inas. “Breast Cancer: Magnitude of the Problem”,Egyptian Society of Surgical Oncology Conference, Taba,Sinai, in Egypt (30 March – 1 April 2005).

13. H. L. Story,1,2 R. R. Love,1 R. Salim,3 A. J. Roberto,4 J. L. Krieger,5 and O. M. Ginsburg1,6, Improving Outcomes from Breast Cancer in a LowIncome Country: Lessons from Bangladesh, International Journal of Breast Cancer Volume 2012 (2012), Article ID 423562, 9 pages.

14. Li, Q. and R.M. Nishikawa, Computer-Aided Detection and Diagnosis in Medical Imaging. 2015: Taylor & Francis.

15. Moon, W.K., et al., Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images. Ultrasound in medicine & biology, 2011. **37**(4): p. 539-548.

16. Yu, S. and L. Guan, A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. Medical Imaging, IEEE Transactions on, 2000. **19**(2): p. 115-126.

17. Cheng, H., et al., Automated breast cancer detection and classification using ultrasound images: A survey. Pattern Recognition, 2010. **43**(1): p. 299-317.

18. Wikipedia. Medical imaging. 2010. http://en.wikipedia.org/wiki/Medical_image_processing#Ultrasound. Jan. 2009.

19. Guo, Y., *Computer-aided detection of breast cancer using ultrasound images.* All Graduate Theses and Dissertations, 2010: p. 635.

20. Tang, J., et al., *Computer-aided detection and diagnosis of breast cancer with mammography: recent advances.* Information Technology in Biomedicine, IEEE Transactions on, 2009. **13**(2): p. 236-251.

21. Shan, J., A fully automatic segmentation method for breast ultrasound images. 2011, Utah State University.

22. Gvamichava, R., et al. "Cancer screening program in Georgia (results of 2011)." Georgian medical news 208-209 (2012): 7-15.

23. Frank, A. and A. Asuncion, *UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]*. Irvine, CA: University of California, School of Information and Computer Science, 2010.

24. Salama, G.I., M. Abdelhalim, and M.A.-e. Zeid, *Breast cancer diagnosis on three different datasets using multi-classifiers.* Breast Cancer (WDBC), 2012. **32**(569): p. 2.

25. Street WN, Wolberg WH, Mangasarian OL. Nuclear feature extraction for breast tumor diagnosis. Proceedings IS&T/ SPIE International Symposium on Electronic Imaging 1993; 1905:861–70.

26. William H. Wolberg, M.D., W. Nick Street, Ph.D., Dennis M. Heisey, Ph.D., Olvi L. Mangasarian, Ph.D. computerized breast cancer diagnosis and prognosis from fine needle aspirates.

27. Cruz, J.A. and D.S. Wishart, *Applications of machine learning in cancer prediction and prognosis.* Cancer informatics, 2006. **2**: p. 59.

28. Yilmaz, I., Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environmental Earth Sciences, 2010. **61**(4): p. 821-836.

29. Xue, J.-H. and D.M. Titterington, Comment on “On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes”. Neural processing letters, 2008. **28**(3): p. 169-187.

30. Quinlan, J.R., *Improved use of continuous attributes in C4. 5.* Journal of artificial intelligence research, 1996: p. 77-90.

31. Hamilton, H.J., N. Shan, and N. Cercone, RIAC: a rule induction algorithm based on approximate classification. 1996: Citeseer.

32. B. Ster, A. Dobnikar, Neural networks in medical diagnosis: comparison with other methods, in: Proceedings of the International Conference on Engineering Applications of Neural Networks, 1996, pp. 427–430.

33. Bennett, K.P. and J.A. Blue. A support vector machine approach to decision trees. in Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on. 1998. IEEE.

34. Setiono, R., Generating concise and accurate classification rules for breast cancer diagnosis. Artificial Intelligence in medicine, 2000. **18**(3): p. 205-219.

35. Nauck, D. and R. Kruse, *Obtaining interpretable fuzzy classification rules from medical data.* Artificial intelligence in medicine, 1999. **16**(2): p. 149-169.

36. Karabatak, M. and M.C. Ince, *An expert system for detection of breast cancer based on association rules and neural network.* Expert Systems with Applications, 2009. **36**(2): p. 3465-3469.

37. Goodman, D.E., L. Boggess, and A. Watkins, *Artificial immune system classification of multiple-class problems.* Proceedings of the artificial neural networks in engineering ANNIE, 2002. **2**: p. 179-183.

38. Abonyi, J. and F. Szeifert, *Supervised fuzzy clustering for the identification of fuzzy classifiers.* Pattern Recognition Letters, 2003. **24**(14): p. 2195-2207.

39. Zhao, X., et al. Automatic 3D facial expression recognition based on a Bayesian Belief Net and a Statistical Facial Feature Model. in Pattern Recognition (ICPR), 2010 20th International Conference on. 2010. IEEE.

40. Ryu, Y.U., R. Chandrasekaran, and V.S. Jacob, *Breast cancer prediction using the isotonic separation technique.* European Journal of Operational Research, 2007. **181**(2): p. 842-854.

41. Cruz-Ramírez, N., et al., *Diagnosis of breast cancer using Bayesian networks: A case study.* Computers in Biology and Medicine, 2007. **37**(11): p. 1553-1564.

42. Maskery, S.M., et al., *A Bayesian derived network of breast pathology co-occurrence.* Journal of Biomedical Informatics, 2008. **41**(2): p. 242-250.

43. Chang, y.w. And w. Chang, a new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. 2009.

44. Hanley, J.A. and B.J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology, 1983. **148**(3): p. 839-843.

45. Foody, G.M., *Thematic map comparison.* Photogrammetric Engineering & Remote Sensing, 2004. **70**(5): p. 627-633.

46. University, C.M. *Algorithm Complexity* Available from: https://www.cs.cmu.edu/~adamchik/15-121/lectures/Algorithmic%20Complexity/complexity.html.

47. Jordan, M.I., Why the logistic function? A tutorial discussion on probabilities and neural networks. 1995, Citeseer.

48. Hilbe, J.M., *Logistic regression models*. 2009: CRC Press.

49. Han, J. and C. Moraga, The influence of the sigmoid function parameters on the speed of backpropagation learning, in From Natural to Artificial Neural Computation. 1995, Springer. p. 195-201.

50. contributors, W. *Stochastic gradient descent*. 2015 29 September 2015 16:50 UTC; Available from: https://en.wikipedia.org/w/index.php?title=Stochastic_gradient_descent&oldid=683330608.

51. Dai, Y.-H., *Convergence properties of the bfgs algoritm.* SIAM Journal on Optimization, 2002. **13**(3): p. 693-701.

52. Wiering, M., et al. Reinforcement learning algorithms for solving classification problems. in Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2011 IEEE Symposium on. 2011. IEEE.