

AN INVESTIGATION OF THE BREAST CANCER CLASSIFICATION USING VARIOUS MACHINE LEARNING TECHNIQUES

A Thesis submitted in partial fulfilment of the requirements for the degree of

Master of Technology

In

Biomedical Engineering

By

RAJESH KUMAR TRIPATHY

211BM1220

Under The Supervision of

DR. SUBHANKAR PAUL



**Department of Biotechnology & Medical Engineering
National Institute of Technology
Rourkela-769008, Orissa, India
2013**

Dedicated To
Maa Mangala
&
My beloved Parents



National Institute of Technology, Rourkela

CERTIFICATE

This is to certify that the thesis entitled “**An Investigation of the Breast Cancer Classification Using Various Machine Learning Techniques**” by **RAJESH KUMAR TRIPATHY (211BM1220)** submitted to the National Institute of Technology, Rourkela for the award of Master of Technology in Biomedical Engineering during the session 2011-2013 is a record of bonafide research work carried out by him in the Department of Biotechnology and Medical Engineering under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma.

Place: NIT Rourkela

Date:

Dr. Subhankar Paul
(Associate Professor)
Department of Biotechnology & Medical Engineering
National Institute of Technology
Rourkela-769008

ACKNOWLEDGEMENTS

This work was carried out at the Department of Biotechnology and Medical Engineering, National Institute of Technology, Rourkela, Orissa, under the supervision of Prof. Subhankar Paul.

I would like to express my sincere gratitude to Prof. Subhankar Paul for his guidance and support throughout my work. Without him I will never be able to complete my work with this ease. He was very patient to hear my problems that I am facing during the project work and finding the solutions. I am very much thankful to him for giving his valuable time for me.

I would like to express my sincere thanks to the PhD scholars Mr. Deependra Kumar Ban and Mr. Sailendra Kumar Mahanta and my friend Mr. Uday Kumar Gupta for their precious suggestions and encouragement to perform my project work.

I thank God for giving me the chance to pursue my academic goals and a wonderful, complete life full of amazing people.

Finally, hearty thanks to my family member for their sacrifice that encouraged me to proceed with the work.

Date:

Place: NIT Rourkela

Rajesh Kumar Tripathy

Roll No.-211BM1220

ABSTRACT

It is an extremely cumbersome process to predict a disease based on the visual diagnosis of cell type with precision or accuracy, especially when multiple features are associated. Cancer is one such example where the phenomenon is very complex and also multiple features of cell types are involved. Breast cancer is a disease mostly affects female population and the number of affected people is highest among all cancer types in India.

In the present investigation, various pattern recognition techniques were used for the classification of breast cancer using cell image processing. Under these pattern recognition techniques, cell image segmentation, texture based image feature extraction and subsequent classification of breast cancer cells was successfully performed. When four different machine learning techniques: Kth nearest neighbor (KNN), Artificial Neural Network (ANN), Support Vector Machine (SVM) and Least Square Support Vector Machine (LS-SVM) was used to classify 81 cell images, it was observed from the results that the LS-SVM with both Radial Basis Function (RBF) and linear kernel classifiers demonstrated the highest classification rate of 95.3488% among four other classifiers while SVM with linear kernel resulted a classification rate of 93.02% which was close to LSSVM classifier. Thus, it was demonstrated that the LS-SVM classifier showed accuracy higher than other classifiers reported so far. Moreover, our classifier can classify the disease in a short period of time using only cell images unlike other approaches reported so far.

Keywords: Breast Cancer; cell images; Image Segmentation; feature extraction; Principal Component Analysis; ANN ; KNN; SVM ; LSSVM ; Classification Rate.

CONTENTS

Acknowledgements	iv
Abstract	v
Contents	vi
List of Figures	viii
List of Tables	x
Notations and Abbreviations	xi
1. INTRODUCTION	2
1.1 Objective of The Thesis	5
2. LITERATURE REVIEW	7
3. MATERIALS AND METHODS	11
3.1 Experimental Documented image collection	13
3.1.1 Collection of MDA-MB-231/MCF7/MCF-10A Cell images	13
3.2 Image Processing	15
3.2.1 Image Segmentation	15
3.2.2 Feature Extraction	16
3.2.2.1 Chip histogram based Texture Features	17
3.2.2.2 Tamura Texture Features	19
3.2.2.3 Wavelet based texture features	20
4. MACHINE LEARNING	24
4.1 Dimensionality reduction using PCA	25
4.2 KNN Classifier	26

4.3 ANN Classifier	27
4.3.1 MFNN Classifier	28
4.3.1.1 Training algorithms for MFNN Classifier	29
4.3.1.2 Selection of hidden neurons	31
4.3.1.3 Selection of ANN parameters	31
4.3.1.4 Training evaluation criteria	32
4.4 Introduction to SVMs	32
4.4.1 SVM Classifier	33
4.4.2 LS-SVM Classifier	36
4.5 Performance Evaluation Criteria for Classifier	40
5. RESULTS AND DISCUSSION	43
5.1 KNN Classifier performance	45
5.2 MFNN Classifier performance	46
5.3 SVM Classifier performance	48
5.4 LS-SVM Classifier performance	50
5.5 Comparisons of the Classifiers Performance	53
6. CONCLUSION	56
Future Work	57
References	58
Dissemination	62

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
1	Flow chart shows the components of pattern recognition system to classify breast cells.	12
2	Phase contrast microscopic image of MCF-7 breast cancer cell line.	14
3	Phase contrast microscopic image of MDA-MB-231 breast cancer cell line.	14
4	Phase contrast microscopic image of MCF 10A normal breast cell line.	14
5	Showing three level wavelet decomposition of image for computation of horizontal, vertical and diagonal details.	22
6	Classification of breast cancer cells and noncancerous cells from reduced feature set.	26
7	Block diagram of Multilayer feed forward Neural Network.	28
8	Working of SVM for linear separable data (the symbol triangle used for cancerous and circle used for noncancerous class).	34
9	Architecture of LS-SVM with 'n' as the number of support vectors, $x_1, x_2, x_3, \dots, x_n$ are the input feature vectors and $K_1, K_2, K_3, \dots, K_n$ are the kernel functions).	37
10	Dilated Gradient image of MDA-MB-231 cancer cell.	43
11	Dilated Gradient image of MDA-MB-231 cancer cell.	43
12	Dilated Gradient image of MDA-MB-231 cancer cell after filling interior gaps.	43
13	Segmented cells from the image background with region of interest of MDA-MB-231 cell.	43
14	Segmented cells from the image background with region of interest of MCF-7 cell.	44
15	Variation of Classification rate with respect to 'K' of the KNN	45

classifier.

16	Variations of training MSE with respect to number of iterations.	47
17	(A) ROC plot for RBF kernel SVM classifier for testing data evaluation. (B) ROC plot for linear kernel SVM classifier for testing data evaluation.	50
18	(A) Training data separation with respect to Hyperplane for linear kernel LS-SVM. (B) Training data separation with respect to Hyperplane for RBF kernel LS-SVM.	51
19	(A) ROC plot for linear kernel LS-SVM classifier for testing data evaluation. (B) ROC plot for RBF kernel LSSVM classifier for testing data evaluation.	53

LIST OF TABLES

Table No.	Table Caption	Page No.
1	General procedure to construct confusion matrix for a binary classifier	40
2	Various Parameters setting for MFNN to get optimized training performance	46
3	Confusion matrix after testing data evaluation for MFNN	47
4	Optimized parameters during Training phase of SVM by considering RBF kernel	48
5	Optimized parameters during Training phase of SVM by considering linear kernel.	48
6	Confusion matrix for SVM with RBF kernel	49
7	Confusion matrix for SVM with linear kernel	49
8	Optimized parameters during Training phase of LS-SVM by considering linear kernel	50
9	Optimized parameters during the Training phase of RBF kernel LS-SVM classifier	51
10	Confusion matrix for linear kernel LS-SVM classifier	52
11	Confusion matrix for RBF kernel LS-SVM classifier	52
12	Comparisons of CR, sensitivity, Specificity and area under ROC curve for classifiers	54

NOTATIONS/ABBREVIATIONS

SVM	Support Vector Machines
LS-SVM	Least Square Support vector machine
AR	Association rules
ANN	Artificial Neural Network
MFNN	Multilayer feed forward Neural Network
KNN	K th nearest Neighborhood
PCA	Principal Component analysis
CR	Classification rate
DMEM	Dulbecco modified eagle medium
FBS	Fetal Bovine Serum
MCF	Michigan cancer foundation
WBCD	Wisconsin breast cancer diagnosis
ROC	Receiver operating characteristics

CHAPTER 1:

INTRODUCTION

&

OBJECTIVE

1. Introduction

Breast cancer is second most commonly diagnosed cancer worldwide (Parkin et al. 2001). In order to find the cure it is necessary to quickly diagnose the disease accurately and treat it based on the kind of symptoms appeared. Breast cancer has several classifications, which may help to determine the best treatment. The most important of these classifications are binary classification, either benign or malignant. If the cancer is in benign stage, less invasive and risk of treatments is used than for malignant stage (Rangayyan et al. 1997). The reason being the chances of survival of patient is high; it is not beneficial to increase the speed of recovery at the risk of introducing potentially life threatening side effects caused by aggressive treatment. On the other hand a patient with malignant cancer is not so concerned about the kind of treatment or side effect of the treatment.

The main cause of breast cancer is when a single cell or group of cells escapes from the usual controls, that regulate cellular growth and begins to multiply and spread. This activity may result in a mass, tumor or neoplasm. Many masses are benign that means the abnormal growth is mainly restricted to a circumscribed, single and expanding mass of cells (Gokhale 2009). Some tumors are malignant that means the abnormal growth invades the surrounding tissues and that may metastasize or spread to remote areas of the body (Rangayyan 2004). The benign masses may leads to complications where as Malignant tumors are serious cancer. The majority of breast tumors will have metastasized before reaching a tangible size. So far, a various numbers of imaging techniques are discovered for breast cancer detection in tissue level. These are categorized as Infrared Imaging and Mammography.

Infrared imaging or Thermography is one of the best methods in detecting breast cancer in early stage, and also useful for recording more advanced stages of breast cancers (Tan et al. 2007). This imaging technique shows the dramatic temperature differences that correlate with various types of breast tissue pathology. Thermal imaging is also great value in monitoring the effectiveness of treatment. The use of thermal sensors with a wavelength range of $3\text{-}5\ \mu\text{m}$ may be used to capture the heat emitted from the breast tissue and form as an image. Thermal imaging is observed as a potential tool for the detection of breast cancer (Arora et al. 2008). A tumor is expected to be more vascularized than its neighboring tissues, and hence could be at a slightly higher temperature. The skin surface near the tumor may also demonstrate a relatively high temperature. Temperature differences up to two degrees have been measured between surface regions near breast tumors and neighboring tissues. Thermography can help in the diagnosis of advanced cancer, but has limited success in the detection of early breast cancer.

Mammography has gained recognition as the single most successful technique for the detection of early, clinically occult breast cancer (Jinshan et al. 2009). Mammograms are generally analyzed by radiologists to detect the early stages of breast cancer (Rojas Domínguez and Nandi 2009). A normal mammogram typically describes the converging patterns of vessels and fibro glandular tissues. Any feature that causes a departure from or distortion with reference to the normal pattern is viewed with suspicion and analyzed with extra attention (Zolghadrasli and Maghsoodzadeh 2006). The important signs of abnormalities and cancer obtained from mammography are Calcifications, Localized increase in density, Masses, Asymmetry between the left and right breast images and Architectural distortion (Moradmand et al. 2011).

The methods discussed above are mainly used for the detection of breast cancer in tissue level. For detecting breast cancer cells and differentiating from non-cancer cells in the cell cultured medium, is one of the vital tasks now a day. So far impedance spectroscopy is a single technique developed to analyze and detect the cancer and non-cancer cell lines on the basis of their bioimpedance responses (Hong et al. 2011; Srinivasaraghavan et al. 2011). The diagnostic and prognostic capability of a MEMS-based micro-bioimpedance sensor was investigated (Srinivasaraghavan et al. 2011). The complex cell culture consisting of three different cell types has been analyzed using MEMS-based impedance spectroscopy. The sensor response showed an insignificant decrease in peak bioimpedance if no cancer cells were present on the electrode. This method is an invasive method for detection and classification of breast cancer cells.

Here, I proposed an alternate method to MEMS-based impedance spectroscopy; i.e. the use of the pattern recognition techniques to classify the cancer and non-cancer cells from cultured cell images. These cultured cell images were taken during culturing of various breast cell types with a suitable medium. The objectives of the thesis are given as below:

1.1 Objective of the Thesis

The main objective of this investigation includes the development of a new computerized pattern recognition technique for the classification of cancer and non-cancer cells from cultured breast cell images. This technique includes:

- (i) Collection of breast cancer cell images as well as non-cancerous cell images during cell culturing.
- (ii) Developing the various pattern recognition techniques to classify breast cancer and non-cancerous cell from these cell images.
- (iii) Comparison of the performance of various classifiers used under this pattern recognition techniques in terms of Classification rate (CR), sensitivity, specificity and area under receiver operating characteristics (ROC) curve, to determine the best classifier for the breast cell classification.

CHAPTER 2:

LITERATURE REVIEW

2. Literature Review

During the past few years, various contributions have been made in literature regarding the application of pattern recognition techniques for breast cancer diagnosis in tissue level. Rejani and his group proposed a pattern recognition procedure to classify the breast tumor (Rejani and Selvi 2009). They used the image segmentation to segment the breast tissue corresponding to the tumor and used the discrete wavelet transform (DWT) as a feature extraction method to extract various features from the segmented images. Then they also used SVM classifier to classify the breast tissue corresponding to the features and achieved an accuracy of 88.75%. Martin and his group proposed the technique for detection of mass on digitized mammograms (Martins et al. 2009). They used K-means clustering algorithm for image segmentation and gray level co-occurrence matrix to describe and analyze the texture of segmented structures in the image. The classification of these structures was achieved through Support Vector Machines, which separate them into two groups; using shape and texture descriptors: masses and non-masses. The classification accuracy obtained from that method was 85%.

Karabatak and his group proposed an automatic diagnosis based pattern recognition system for detecting breast cancer based on the association rules (AR) and Artificial neural network (ANN) (Karabatak and Ince 2009). In that study, they used AR method for reducing the dimension of breast cancer database and ANN for intelligent classification. The proposed system i.e. the combination of AR and ANN, performance was compared with only ANN model. The dimension of input feature space was reduced from nine to four by using AR. In the testing stage, they use 3-fold cross validation method to the WBCD to evaluate the proposed pattern recognition system

performances. The correct classification rate obtained from that AR + ANN system was 95.6%. Jele and his group proposed a framework for automatic malignancy grading of the fine needle aspiration biopsy tissue (Jele et al. 2008). They used an SVM classifier to assign a malignancy grade based on pre extracted features, with accuracy up to 94.24%. Arodz and his group proposed the Pattern recognition system for automatic detection of suspicious looking anomalies in mammograms (Arodź et al. 2005). They used adaboost algorithm based classifier and achieved an accuracy of 90%. Brook and his group proposed a method for Breast Cancer Diagnosis from Biopsy Images using SVM (Brook et al. 2006). They applied multi-class SVM on generic feature vectors to achieve a high recognition rate.

Fatima and his group used an approach for classifying the breast cancer from Wisconsin breast cancer diagnosis (WBCD) database using adaptive neuro-fuzzy inference system (ANFIS) (Fatima and Amine 2012). By using the ANFIS classifier they were achieved an accuracy of 98.25 % in tissue level. Mousa and his group proposed a pattern recognition method, to classify masses for micro calcification and abnormal severity such as benign or malignant from mammographic image (Mousa et al. 2005). They used wavelet analysis as a feature extraction technique and fuzzy-Neuro as a classifier to achieve a better classification rate. Niwas and his group proposed the pattern recognition task; by considering Color Wavelet Features as feature extraction technique from segmented histopathology image (Issac Niwas et al. 2012). They used SVM classifier which gives an accuracy of 98.3%.

Akay proposed another pattern recognition method by considering SVM classifier to classify benign and malignant masses (Akay 2009). They used WBCD breast cancer database and achieved

an accuracy of 98%. Shi and his group proposed another technique for detection and classification of masses from breast ultrasound images. They used textural features, fractal features as feature extraction methods and SVM, fuzzy support vector machine (FSVM) as the classifiers, to classify the benign and malignant masses (Shi et al. 2010). After classification, they were achieved a classification accuracy of 96.4% for FSVM classifier. Schaefer and his group proposed a pattern recognition technique, by considering basic statistical features, histogram features, cross co-occurrence matrix features and mutual information based features as the feature extraction method and Fuzzy rule based classifiers as classification method to classify the benign and malignant tumor types from the thermogram images (Schaefer et al. 2009). They achieved a classification rate of 79.53% with the help of 14 partitions based fuzzy rule classifier.

CHAPTER 3:

MATERIALS

&

METHODS

3. Materials and Methods

The technique Pattern recognition mainly originates from the need for automated machine recognition of signals, images and objects or any decision based approach, on the basis of the set of features. The goal of Pattern recognition is to predict the correct level corresponding to given feature set based on a better knowledge obtained through training. The pattern recognition can well understood by considering an example: the human being can easily identify the gender based on the face while machine can't, so the aim is to train the machine by considering various features as: facial expression, facial bone structure, hair length and others. After training, the machine can easily identify the required class the new test object belongs. From this above example, it is quite obvious that the heart of pattern recognition system is feature extraction and classification.

The main objective of this research is dealing with the classification of the cancerous and non-cancerous cells from cultured breast cell images, using image processing and machine learning techniques. The pattern recognition system for this classification task is given by the flowchart; as shown in Fig.1. Here, the proposed Pattern recognition system is developed with the help of 81 numbers of cultured breast cell images; out of these 55 numbers of images are belong to cancerous and 26 belongs to non-cancerous class. The software program to implement this system is developed in MATLAB software. The flowchart as shown in Fig.1 mainly consists of three sections: Experimental documented image collection, Image processing and Machine learning. The experimental documented image collection and image processing parts are discussed in this chapter where the machine

learning part is discussed in chapter-3. Under image processing part, I use the image segmentation i.e. to segment out the cells from image background and feature extraction; i.e. to extract of various meaningful features from each segmented image.

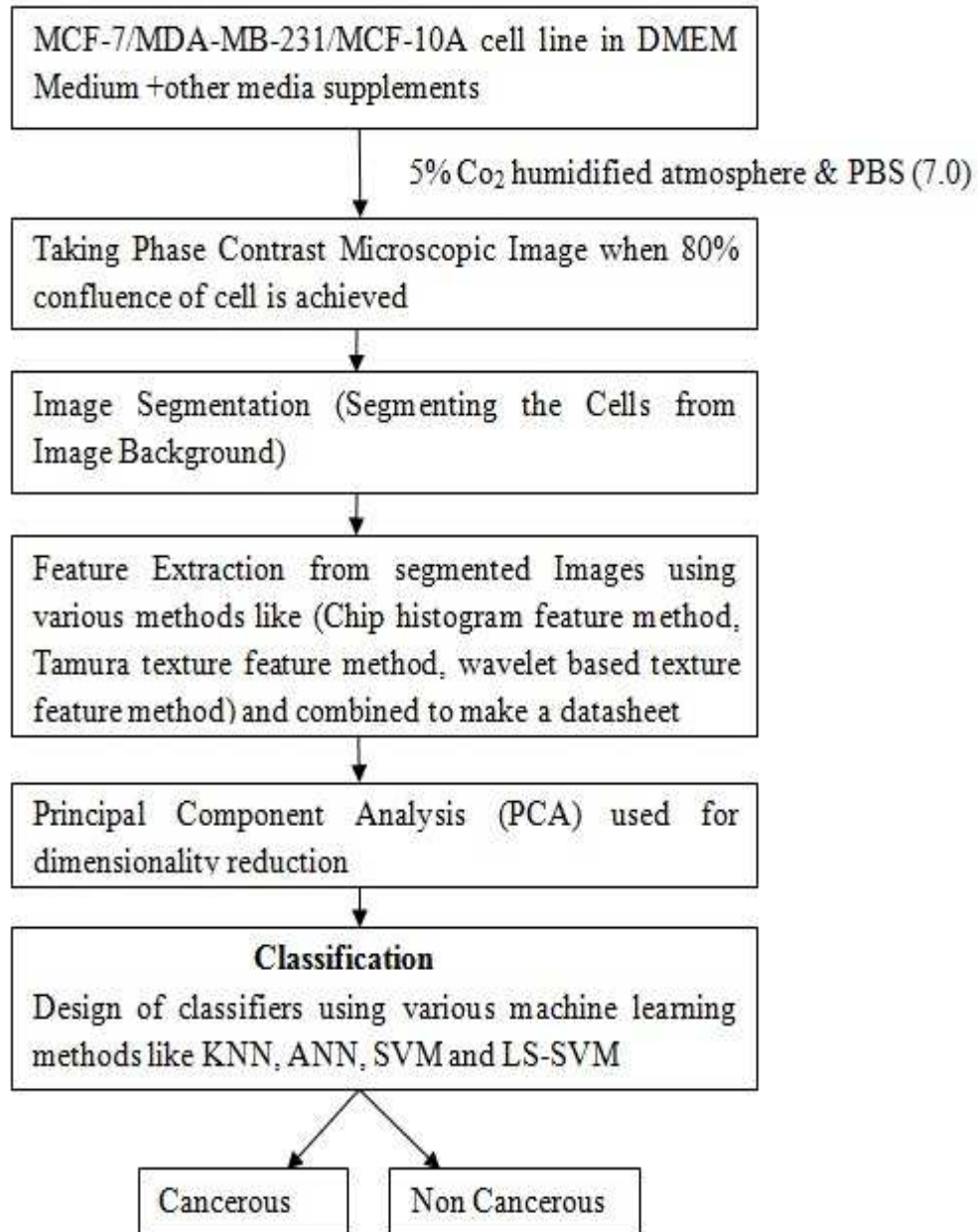


Figure 1. Flow chart shows the components of Pattern recognition system to classify breast cells.

3.1 Experimental Documented image collection

The experimental documented image collection includes the capturing of phase contrast microscopic images during cell culturing. Cell culturing is a process, where the cells are grown under controlled conditions besides the natural environment. Here, the different cancerous and non-cancerous breast cell lines like MDA-MB-231, MCF-7 and MCF-10A were used for cell culturing. The images were taken when 80% confluence of cells is achieved. The MDA-MB-231 breast cancer cell line was obtained first time from a patient at M.D. Anderson Cancer Center. Due to epithelial like morphology the MDA-MB-231 breast cancer cells appear as spindle shaped cells. Similarly, the MCF-7 cells were first obtained from 69 year old woman. These cells have the characteristics like the presence of both estrogen receptor and progesterone receptor. MCF 10A cells are the normal or non-cancerous breast cells. For culturing of these cell lines, I follow the protocols, which are given as: Preparation of aseptic environment, Preparation of cell growth medium and Monitoring the cell growth in terms of confluence.

3.1.1 Collection of MDA-MB-231/MCF7/MCF-10A Cell images

- (i) These Cells were cultured in DMEM cultured media, supplemented with 10% of Fetal Bovine Serum (FBS), 1% penicillin or streptomycin, nonessential amino acids (0.1 mm), Insulin (10ug/mL), Sodium pyruvate (1mM) and 10 nM estrogen.
- (ii) Also other factors like 37°C in humidified and CO₂ (5%) atmosphere are taken during the culturing process.
- (iii) The cultured images were taken with a phase contrast microscope after 80% of confluence of cells was achieved.

After cell culturing the sample cultured cell images are shown in Figure-2, 3, 4. In this study, the MDA-MB-231, MCF-7 cells were cultured in our laboratory, but some of the MCF-10A cultured cell images were collected from various online resources.

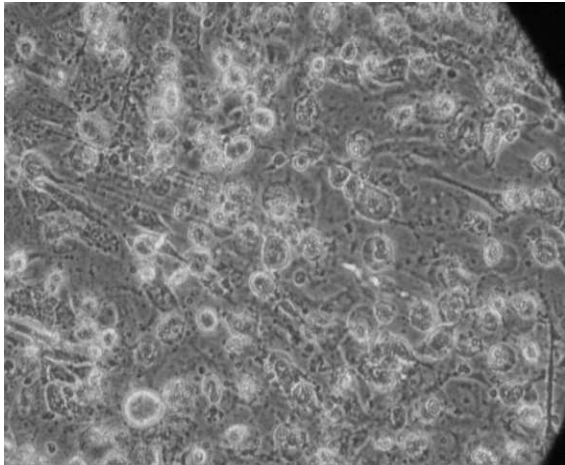


Figure 2. Phase contrast microscopic image of MCF-7 breast cancer cell line.

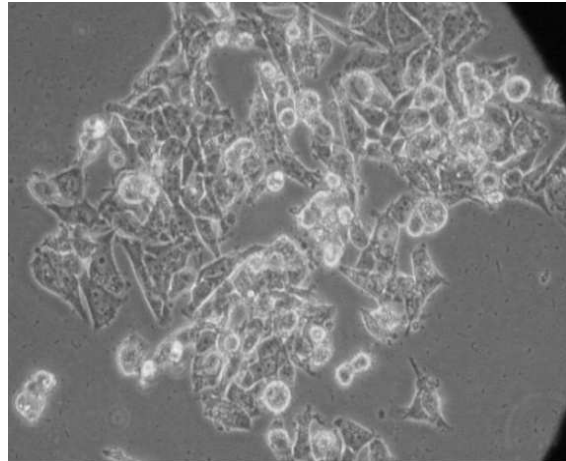


Figure 3. Phase contrast microscopic image of MDA-MB-231 breast cancer cell line.

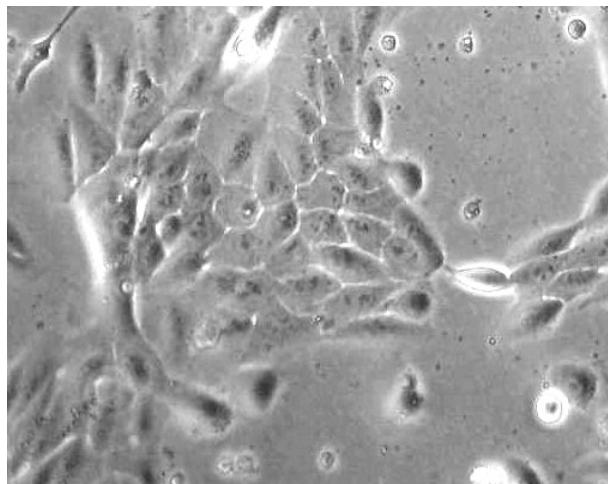


Figure 4. Phase contrast microscopic image of MCF-10A normal breast cell line.

3.2 Image Processing

It is a class of signal processing, where the 2D pictures are processed to collect the meaningful information. In this research, I use two techniques, which come under the task of image processing. These are:

- (a) Image segmentation
- (b) Feature extraction

3.2.1 Image segmentation

Image segmentation refers to finding and labeling areas in an image which gives information about certain things than other part. This can be used to isolate certain areas of interest from an image to distinguish those areas within the image. Applications for image segmentation are found in remote sensing, Medical imaging and other image processing systems. Here I used the texture based segmentation method to segment cells from cultured MDA-MB-231, MCF-7 and MCF-10A cell images. The following steps are used for the segmentation process. These are given as:

Step-1: Detecting cells from image - The cell to be segmented differs greatly in contrast from the background of the image (Anoraganingrum 1999). So the changes in contrast can be detected by calculating the gradient of an image. This gradient of the image is calculated using the Sobel function. After obtaining the gradient of the image, I tuned it with the help of a threshold value to create the binary mask, which contains the segmented cells.

Step-2: Dilation of the gradient image - As the binary gradient image obtained in step-1 gives rise to lines of high contrast in the image. These lines do not quite define the outline of the object

of interest. So the dilation operation is used with the help of horizontal structuring element followed by the vertical structuring element to obtain the dilated image.

Step-3: Filling in Interior Gaps - The dilated gradient image obtained in step-2 shows the outline of the cell quite nicely, but there are still holes presents in the interior of the cell. To fill these holes, I use the hole filling operation of image processing, i.e. to fill the interior gaps.

Step-4: Removing the Connected Objects on Border and Smoothing of the Object –After filling the interior gaps, the border objects are removed for better visualization of the segmented image. In order to make the segmented object look natural, I smooth the object of interests by eroding the image with a diamond structuring element and displaying the segmented object by placing an outline around the segmented cell.

The results obtained from the step by step segmentation technique for MDA-MB-231 cell image is given in chapter-5. Similarly, the same segmentation procedure is carried out for other cell images like MCF-7 and MCF-10A. After segmentation, I switch over to feature extraction stage, i.e. to obtain the meaningful features form each of the segmented cancer and non-cancer cell image class.

3.2.2 Feature Extraction

In the field of Image processing and pattern recognition, the feature extraction is a technique used for collecting various features from image and dimensionality reduction. Here, I use three different texture feature extraction methods (Haralick et al. 1973) (i.e. Chip histogram based texture features, Tamura Texture Features and Wavelet based Texture features) to extract the various features from each segmented cultured cell images.

3.2.2 .1 Chip histogram based texture features

For the image with number of gray levels as 'L', so there exists the gray vector, which varies from 1 to L. The histogram of the image is given as: Histogram= $h(r_k)=n_k$, where ' r_k ' is the k^{th} gray level, n_k is the number of pixels in the image having gray level ' r_k ' and $h(r_k)$ is the histogram of a digital image with gray level r_k (Kaiwei et al. 2011). The image having size of $M \times N$, Gray level probability function is defined as:

$$\text{Gray level probability} = P(r_k) = \frac{h(r_k)}{M \times N} \quad (3.1)$$

The various chip histogram features are extracted from each cancer and non-cancer segmented cell image. These features are explained in details, below:

(a) **Mean**-It measures the mean value of pixels in an image. It is defined as:

$$\text{Mean } (\mu) = \sum_{r_k=1}^L P(r_k) \times r_k \quad (3.2)$$

(b) **Variance**: - The expected value of the square of the deviations of gray level of an image from its mean value. It is defined as:

$$\text{Variance } (\sigma^2) = \sum_{r_k=1}^L P(r_k) \times (r_k - \mu)^2 \quad (3.3)$$

(c) **Kurtosis**- it is a measure of; how the image corresponding to either cancerous or non-cancerous class, is peaked or flat relative to the probability distribution. The image with high kurtosis tends to have different peak near mean value and decline rapidly. Similarly the image having low kurtosis tends to have flat top near mean rather than sharp peak.

The kurtosis of a distribution can be calculated as:

$$\text{Kurtosis (Ku)} = \frac{\sum_{r_k=1}^L P(r_k) \times (r_k - \mu)^4}{\sigma^4} \quad (3.4)$$

Where μ the mean and σ is the standard deviation of the image with gray vector varies from 1 to L.

(d) **Skewness**-It is a measure of the asymmetry of the pixel value of image around the sample mean. If Skewness is negative, the pixels in the image are spread out more to the left of the mean than to the right. Similarly, if Skewness is positive, the pixels in the image are spread out more to the right. The Skewness of the normal distribution (or any perfectly symmetric distribution) is zero. The Skewness is defined mathematically as:

$$\text{Skewness(S)} = \frac{\sum_{r_k=1}^L P(r_k) \times (r_k - \mu)^3}{\sigma^3} \quad (3.5)$$

(e) **Entropy**-It is a statistical measure of randomness that can be used to characterize the texture of the image. The entropy is calculated mathematically as:

$$\text{Entropy} = - \sum_{r_k=1}^L P(r_k) \times \log(P(r_k)) \quad (3.6)$$

(f) **Energy**- Energy of an image gives the concepts about measure of the information. It can be calculated by using a probability distribution function. The energy of an image is given as:

$$\text{Energy} = \sum_{r_k=1}^L P(r_k)^2 \quad (3.7)$$

3.2.2.2 Tamura Texture Features

Coarseness: This feature is calculated by computing six averages for the windows having size $2^m \times 2^m$, $m=0,1,\dots,5$, around the pixel (x,y) and then by computing the absolute differences $E_m(x,y)$ between the pairs of non-overlapping averages in the horizontal and vertical directions. Further, the value of m can be found out, by maximizing the difference $E_k(x,y)$ in either direction and the best window size obtained as $S_{\text{best}}(x,y) = 2^m$. Hence the Coarseness (F_{crs}) is obtained by averaging $S_{\text{best}}(x,y)$ over the entire image (Tamura et al. 1978). Mathematically it is defined as:

$$\text{Coarseness} = \frac{1}{M * N} \sum_{i=1}^M \sum_{j=1}^N S_{\text{best}}(i, j) \quad (3.8)$$

(b) Contrast- Contrast measures, how the grey levels vary in the image and to what extent their distribution is biased to black or white. The second-order and normalized fourth-order central moments of the grey level histogram, i.e., the variance (σ^2) and kurtosis (α_4) are used to define the

contrast (Tamura et al. 1978). The contrast is given by the formulae: $F_{\text{con}} = \frac{\sigma}{\alpha^n}$; where $\alpha = \frac{\mu}{\sigma^4}$,

‘ μ ’ is the mean value and ‘ σ ’ is the standard deviation of the image, given in Eqn. 1 and Eqn. 2.

(c) Directionality- The degree of directionality mainly related to the sharpness of the peaks. It is measured using the frequency distribution of oriented local edges against their directional angles. The edge strength $e(x,y)$ and the directional angle $a(x,y)$ are computed using the Sobel edge detector approximating the pixel-wise x - and y -derivatives of the image.

$$e(x,y) = 0.5(|\Delta_x(x,y)| + |\Delta_y(x,y)|) \quad (3.9)$$

$$a(x,y) = \tan^{-1}(\Delta_y(x,y) / \Delta_x(x,y)) \quad (3.10)$$

A histogram $H_{dir}(a)$ of quantized direction values a is constructed by counting numbers of the edge pixels with the corresponding directional angles and the edge strength greater than a predefined threshold (Tamura et al. 1978).

The degree of directionality given by

$$F_{dir} = 1 - rn_{peaks} \sum_{p=1}^{n_{peaks}} \sum_{a \in w_p} (a - a_p)^2 H_{dir}(a) \quad (3.11)$$

Where n_p is the number of peaks, a_p is the range of the angles attributed to the p^{th} peak (that is, the range between valleys around the peak), r denotes a normalizing factor related to quantizing levels of the angles a , and a is the quantized directional angle (cyclically in modulo 180°).

3.2.2.3 Wavelet based Texture features

Wavelet is a mathematical function that can decompose a signal or an image with a series of averaging and differencing calculations. Wavelets are typically used in image decomposition and compression. The image can be decomposed by passing through a series of low-pass and high-pass filter and then reconstructed by simply reversing the decomposition process (Kingsbury 1998). Wavelets calculate average intensity properties as well as several detailed contrast levels, distributed throughout the image. Wavelets can be calculated according to various levels of resolution and depending on how many levels of averages are calculated. Not only they are sensitive to the spatial distribution of gray level pixels, but also able to differentiate and preserve details at various scales or resolutions. This multi-resolution quality allows for the analysis of gray level pixels regardless of the size of the neighborhood. These properties lead to the idea that wavelets could guide researchers to better texture classification in the field of pattern classification (Unser and Aldroubi 1996).

Here, I use three level 2D wavelet decomposition methodologies to find out vertical, horizontal and diagonal coefficients of each image, which are shown in figure-5. Wavelet transforms are finding intense use in fields as diverse as telecommunications and biology. Because of their suitability for analyzing non-stationary signals, they have become a powerful alternative to Fourier methods in many medical applications. The main advantage of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast frequencies, thus leading to an optimal time-frequency resolution in all the frequency ranges. This wavelet transform is computed by applying a filter bank to the image. The rows and columns of an image are processed separately and down sampled by a factor of 2 in each direction, resulting in one low pass image LL and three other detail images like HL, LH, and HH. The LH channel contains image information on low horizontal frequency and high vertical frequency, the HL channel contains high horizontal frequency and low vertical frequency, and the HH channel contains high horizontal and high vertical frequencies. The wavelet coefficients obtained from an image are given as:

$$W_{\phi}(j_0, m, n) = \frac{1}{\sqrt{M \times N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \phi_{j_0, m, n}(x, y) \quad (3.12)$$

$$W_{\psi}^i(j, m, n) = \frac{1}{\sqrt{M \times N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j, m, n}^i(x, y), i = \{H, V, D\} \quad (3.13)$$

Where $\psi_{j, m, n}^i(x, y) = 2^{j/2} \psi(2^j x - m, 2^j y - n), i = \{H, V, D\}$ and $\phi_{j_0, m, n}(x, y) = 2^{j/2} \phi(2^j x - m, 2^j y - n)$ are the scaled and translated basis functions.

LL3	HL3	HL2	HL1
LH3	HH3		
LH2		HH2	
LH1			HH1

Figure 5. Showing the three level wavelet based decomposition of image for computation of horizontal, vertical and diagonal details.

The wavelet based texture feature such as energy is calculated along all directions i.e. horizontal, vertical and diagonal. The energy for a single wavelet coefficient is given as:

$$E_{C,L} = \frac{1}{L \times L} \sum_{i=0}^L \sum_{j=0}^L (W_{C,L,i,j})^2 \quad (3.14)$$

Where, $W_{C,L,i,j}$ is the wavelet coefficient at (i, j) location at L level in C. ($C \in (LH, HL, HH)$) is the sub band level decomposition (Dua et al. 2012).

In this work, the each segmented image is decomposed into three levels using discrete wavelet transform. Then the energy is calculated for each wavelet coefficient corresponding to the level of decomposition. For a single segmented image, I got a sum total of 81 numbers wavelet energy features. For other segmented images belongs to both cancerous and non-cancerous classes, the same process is repeated to get the desired wavelet based texture features.

CHAPTER 4:

MACHINE LEARNING

4. Machine Learning

Machine learning is a multidisciplinary field of study that mainly concerned with the design of algorithms which allow computers to learn. The term “Machine Learning” comes from the artificial intelligence community but now a day it mainly the focusing area for many branches of engineering and science (Alpaydin 2004). Learning mainly refers to learning from data or feature set. There are different learning methods for statistical data analysis. These are supervised learning, unsupervised learning and reinforcement learning.

The goal of supervised learning is usually generalize the input output relationship and facilitating the prediction of output associated with previously unseen inputs. The primary supervised learning tasks are the classification and regression. Similarly, the goal of the unsupervised learning is typically not related to the future observations. Instead, one seeks to understand structure in data sample itself, or to infer some characteristics of the probability distribution. The main unsupervised learning tasks are the clustering, density estimation and dimensionality reduction. Dimensionality reduction can also be taken as the class of supervised learning but there should be the input-output relationship. In case of reinforcement learning the input patterns are observed sequentially and after each observed pattern the learner must take an action. After each action the author gets a reward from the environment. The main goal of the learner in supervised learning is to determine a policy to maximize the long term reward. This type of learning is mainly useful in economics, robotics (path planning and navigation) and other areas (Hastie et al. 2005). In this study, the PCA algorithm, which comes under the unsupervised learning task, is used for the dimensionality reduction (Kambhatla and Leen 1997).

4.1 Dimensionality reduction using Principal Component Analysis (PCA)

PCA is a machine learning algorithm which uses an orthogonal transformation to convert a set of correlated variables into a set of uncorrelated variables. These uncorrelated variables are called as principal components. These principal components are less than or equal to the number of original variables (Wold et al. 1987).

PCA mainly consists of two computational steps; first finding the covariance matrix of the data; Then by using this matrix, to compute the Eigen vectors. An Eigenvector is a square matrix, when multiplied by original matrix, yields a vector that differs from the original by a multiplicative scalar λ . The scalar λ is called as Eigen value of the matrix. The Eigenvectors are otherwise called as the principal components. Before using PCA algorithm, it is necessary to first normalize the features by subtracting the mean value of each feature from the dataset and by scaling each dimension, so that they fall in the same range. In this study, after normalizing the data I run the PCA algorithm to compute the principal components. The step by step procedure for this algorithm is given as:

- (i) First, the covariance matrix is calculated: i.e. $S = \frac{1}{m} x^T x$, where x is a ' $m \times q$ ' matrix consists of ' m ' instances in rows and ' q ' columns as features (i.e. the linear combination of chip histogram features, Tamura texture features and wavelet based texture features).
- (ii) Then the matrix ' P ' obtained as: i.e. each row ' p_i ' is an Eigen vector of matrix S .
- (iii) For A to be a square matrix, there exist a non-zero vector ' v ', which is an Eigenvector of A . If there exist a scalar λ then $Av = \lambda v$.

(iv) As there are 'm' Eigenvectors so for reducing the Eigenvector size i.e. 'M' dimension to 'k' dimensions, I choose the 'k' Eigenvectors which related to 'k' largest Eigen values of λ .

After getting the reduced feature set from PCA, the classifiers (KNN, ANN, SVM and LSSVM) which come under the supervised learning task of machine learning are used to classify this set of features into cancerous and noncancerous classes. In machine learning and statistics the classification is a task of identifying, to which class the new observation belongs. An algorithm which implements the classification is mainly called as the classifier. Generally classification is divided into two categories i.e. the binary classification and multi class classification. The Fig. 6 shows the application of various classifiers for the present binary classification task, after getting the reduced feature set from the PCA stage of pattern recognition system.

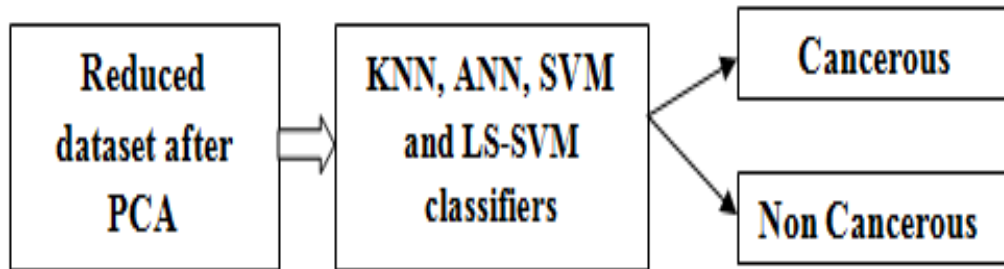


Figure 6. Classification of breast cancer and non-cancer cells from reduced feature set

4.2 K-Nearest Neighborhood (KNN) Classifier

K-nearest neighbor algorithm is a technique for classifying data based on the closest training examples in the feature space. Before using the KNN the protocol should be followed, i.e. given as: First the dataset is divided into a testing and training set. For each row in the testing set, the 'K' nearest training set objects are found, and the classification of test data is determined by majority

vote with ties are broken at random. If there are ties for the K^{th} nearest vector then all the instances are included in the vote (Wold et al. 1987).

The way the KNN classifier works is, first by calculating the distances between the testing data vector and all of the training vectors using a particular distance calculation methodology which is given as follows:

Considering the case of two input variable; the Euclidean distance between two input vectors p and q is computed as the magnitude of difference in vectors i.e. $|p - q|$, Where both the data are having 'm' dimensions i.e. $p = (p_1, p_2 \dots p_m)$ and $q = (q_1, q_2, \dots, q_m)$.

The Euclidean distance between 'p' and 'q' is found to be:

$$D(p, q) = |p - q| = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_m - q_m)^2} \quad (4.1)$$

The KNN classifier takes the test instance 'x' and finds the K-nearest neighbors in the training data and assigns 'x' into the class occurring most among the K neighbors.

4.3 Artificial Neural Networks (ANN) as classifier

Artificial Neural Networks (ANNs) are designed to mimic the decision making ability of the human brain and hence their structure is similar to that of the nervous system. It consists of a large number of highly interconnected processing elements (called neurons), which are working on simultaneously to solve a specific task in machine learning. These tasks are mainly the classification and function estimation. Like human being the ANN is also learning from the instances or examples (Mohanty and Ghosh 2010). To design an ANN based system, a numbers of programming concepts and mathematical models are useful. In this research, I use Multilayer feed

forward neural network (MFNN) as classifier to classify the cancerous and non-cancerous cells from reduced features, which is obtained through the PCA stage of the proposed pattern recognition system.

4.3.1 Multilayer feed forward Neural Network (MFNN) Classifier

The MFNN classifier mainly consists of three layers: which are categorized as input layer, output layer and hidden layer. The hidden layer comes between input and output layers. The Input layer of MFNN mainly consists of different numbers of input variables. Here, I use the input variables, as the reduced features obtained from PCA stage and the output; i.e. corresponds to a binary class. The output is assigned as class-1 for cancerous and class-0 for non-cancerous cell. The architecture of MFNN classifier is shown in Fig. 7.

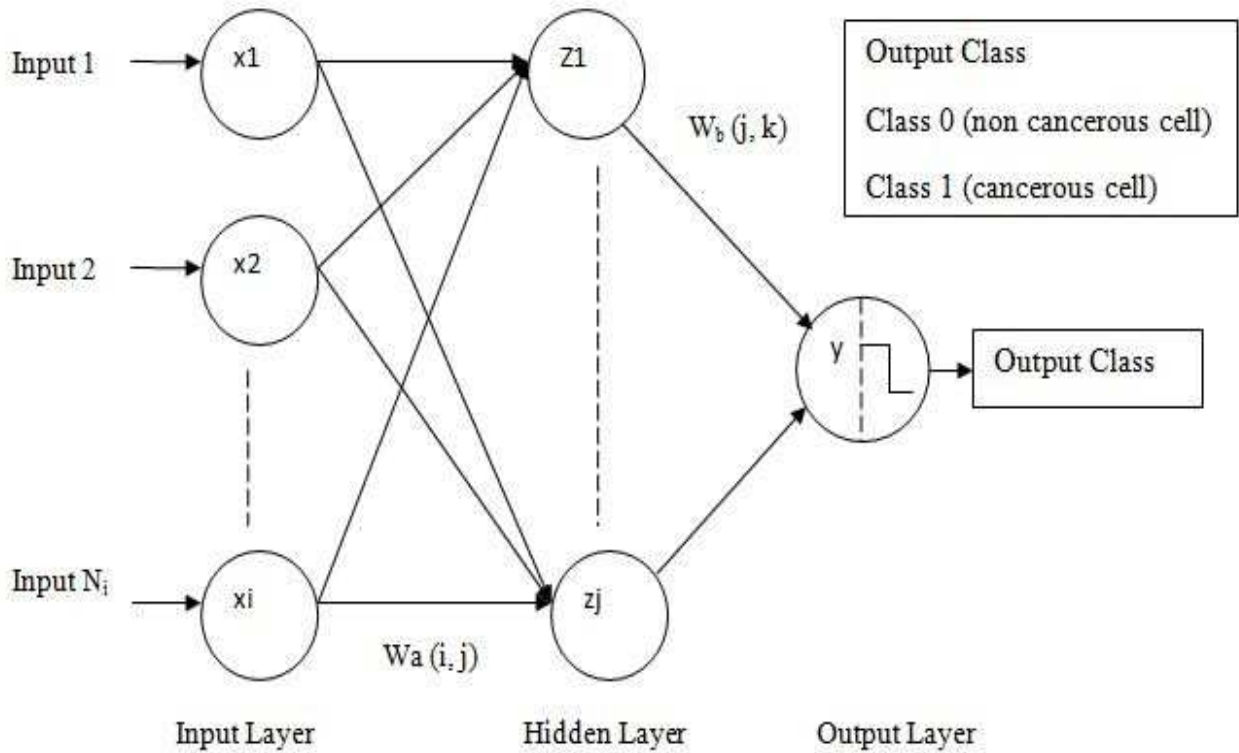


Figure 7. Block diagram of Multilayer feed forward Neural Network

4.3.1.1 Training algorithms for MFNN Classifier

A number of training algorithms are used to train the neural networks. However, Back Propagation Algorithm (BPA) algorithm is the simplest one, which is easy for understanding and implementation. This algorithm mainly consists of four stages: i.e. Initialization of weight values, Feed forward Propagation, Back propagation of errors, and updating the weight values (Sivanandam and Deepa 2006).

Considering a single instance of training vectors for the neural network i.e. $x = (x_1, x_2, \dots, x_n)$ as the input vector, which specifies of the reduced features obtained from PCA stage and 'T' is the output unit corresponds to binary values (i.e. 'T = 0' for non-cancerous cell and 'T = 1' for cancerous cell).

So for 'p' number of instances, the input matrix 'x' consists of 'p' numbers of rows and 'n' numbers of column. The row vector consists of a number of instances while, column vector consists of a number of features. The output column vector consists of the target vector as: $T = (T_1, T_2, \dots, T_p)$, where 'p' is the number of instances taken for neural network analysis, δ_{jk} = error at the output layer (Y_k), δ_{ij} = error at the hidden layer (Z_j), α = Learning rate, η = Momentum factor and $Y = (Y_1, Y_2, \dots, Y_p)$ are the predicted output obtained from the MFNN classifier. The following procedure is applied to train the MFNN network, which is given by:

Initialization of weight values- Here, I Initialize the weights to a small random values in between the input layer to hidden layer i.e. $W_a(i, j)$, as well as from hidden layer to output layer i.e. $W_b(j, k)$.

Feed forward Propagation:- In this case, each of the input receives the signal x_i and transmits all these signals to the hidden layer. For Each hidden unit (Z_j , $j=1, 2 \dots p$), sums its weighted input signals as:

$$Z_{ij} = \sum_{i=1}^n x_i W a_{ij} \quad (4.2)$$

Then, by applying the activation function to the weighted sum the result becomes:

$$Z_j = f(Z_{ij}) \quad (4.3)$$

Further the weighted sum results (Z_j) are transmitted to the output layer. At the output layer, each output unit (Y_k , $k=1, 2 \dots n$) sums the weighted input signal from the hidden layer as:

$$Y_{jk} = \sum_{j=1}^p Z_j W b_{jk} \quad (4.4)$$

Similarly, by applying the activation function at the output layer, the output value is written as:

$$Y_k = f(Y_{jk}) \quad (4.5)$$

The activation function used in the MFNN is a sigmoid function which is defined as:

$$S(q) = 1 / (1 + e^{-q}) \quad (4.6)$$

This function is applied to all neurons except the input neurons.

Back Propagation error- The error obtained due to change in predicting value with respect to the target value (T_k) at the output unit (Y_k) of MFNN classifier during training is defined as:

$$\delta_{jk} = (T_k - Y_k) f'(Y_{jk}) \quad (4.7)$$

Similarly, for each hidden unit (Z_j , $j=1, 2 \dots p$), the error during the training phase is given as:

$$\delta_{ij} = \left(\sum_{k=1}^m \delta_{jk} \times W b_{jk}(m) \right) f'(Z_{ij}) \quad (4.8)$$

The minimum error in hidden layer as well as the output layer corresponds to the change in weight.

These weights are updated as given below:

Updating of weights- The weights between the hidden layer and the output layer are updated as:

$$Wb_{jk}(m+1) = Wb_{jk}(m) + \eta \times \delta_{jk}(m) \times S_b(j) + \alpha \times (Wb_{jk}(m) - Wb_{jk}(m-1)) \quad (4.9)$$

Similarly, the weights between the input layer and the hidden layer are also updated as:

$$Wa_{ij}(m+1) = Wa_{ij}(m) + \eta \times \delta_{ij}(m) \times S_a(i) + \alpha \times (Wa_{ij}(m) - Wa_{ij}(m-1)) \quad (4.10)$$

Where ‘m’ is the number of iterations, $\delta_{ij}(m)$ is the error in the j^{th} output after the m^{th} iteration, $S_a(i)$ is the output of the input layer, $\delta_{jk}(m)$ is the error for the k^{th} output at the m^{th} iteration. $S_b(j)$ is the output of the hidden layer.

4.3.1.2 Selection of hidden neurons

The selection of optimal number of hidden layer neurons (N) is one of the most interesting aspects of designing the MFNN classifier. Simon Haykin proposed the value of ‘N’, which should lie between 2 and ∞ (Haykin 2009). Similarly, Hecht-Nielsen proposed, for a single hidden layer neural network the number of hidden neurons found to be $2(N_i+1)$, where ‘ N_i ’ is the number of input neurons. A large value of ‘N’ may reduce the training error associated with the MFNN, but increase the computational complexity and time (Stathakis 2009).

4.3.1.3 Selection of ANN parameters

The learning rate (α) and the momentum factor (η) are the most important factors, which affect the learning speed of MFNN. For a small value of α results a very slow rate of learning while If the

value α is large the speed of learning is also increases which makes the MFNN unstable (Haykin and Network 2004). A very simple technique to increase the learning rate without making the MFNN system unstable is simply by adding the momentum factor (Haykin 2009). The values of the momentum factor (η) and learning rate (α) should lie between 0 and 1.

4.3.1.4 Training evaluation criteria

The Mean Square Error (E_{tr}) for the training patterns after the m^{th} iteration is defined as:

$$E_{tr} = \left(\sum_{k=1}^K (t(m) - Y(m))^2 \right) \times \left(\frac{1}{K} \right) \quad (4.11)$$

Where ‘ t ’ is the experimental output corresponding to cancerous and non-cancerous class, ‘ k ’ is the number of training patterns and ‘ Y ’ is the predicted value of the output after m^{th} iteration using neural network. The training is stopped when the least value of training error (E_{tr}) has been obtained and this value does not change much with the number of iterations.

4.4 Introductions to SVMs

SVM and LS-SVM are the category of supervised learning methods, which is mainly used for solving classification and regression tasks in machine learning. The reduced features obtained from PCA stage are directly fed to both SVM and LS-SVM Classifiers. At the output, the classifier predicts, the features corresponding to image belong to either cancerous or non-cancerous classes. The SVM and LS-SVM are acts as non-probabilistic binary linear classifiers and builds a Hyperplane which separates the two classes. The most important concept of SVM and LS-SVM is

separating Hyperplane which can well understand by these three key points (Cristianini and Shawe-Taylor 2000; Suykens and Vandewalle 1999).

- The maximum-margin Hyperplane (This corresponds to the best Hyperplane is the one that separates the two classes and lies at a maximal distance from any samples).
- The soft-margin (This allows for some degree of the misclassification).
- The kernel function (This function maps the input reduced features to a higher dimension feature space, with the goal of finding a separation between the data).

4.4.1 Support vector machine (SVM) Classifier

An SVM classifier classifies the data by determining the optimal Hyperplane which can separates all the data points of one class from that of the other class. This optimal Hyperplane for an SVM classifier means the one with the highest margin between the two classes. The margin means the maximal width of the slab parallel to the Hyperplane that has no interior data points (Campbell and Ying 2011).

For a given training set $\{x_n, y_n\}_{n=1}^K$ with the input data as $x_n \in R^m$ and corresponding the output class level as $y_n \in R$. Here the input data are transferred to the high dimensional feature space with the help of nonlinear function $\phi(\cdot)$.

The decision function associated with Hyperplane to distinguish the binary classes (as shown in fig. 2) is given as:

$$y(x) = \text{sgn}(w^T \phi(x) + b) \quad (4.12)$$

In this eqn-4.12, the 'w' corresponds to weight vector and 'b' corresponds to the bias value. These weight vectors and the bias values for optimal Hyperplane SVM were obtained by maximizing the

distance between the closest training point and Hyperplane. This can be achieved by maximizing

the margin, which is defined as $M = \frac{2}{|w|}$ and same as minimization of:

$$\frac{\|w^2\|}{2} = \frac{1}{2} w^T w \quad (4.13)$$

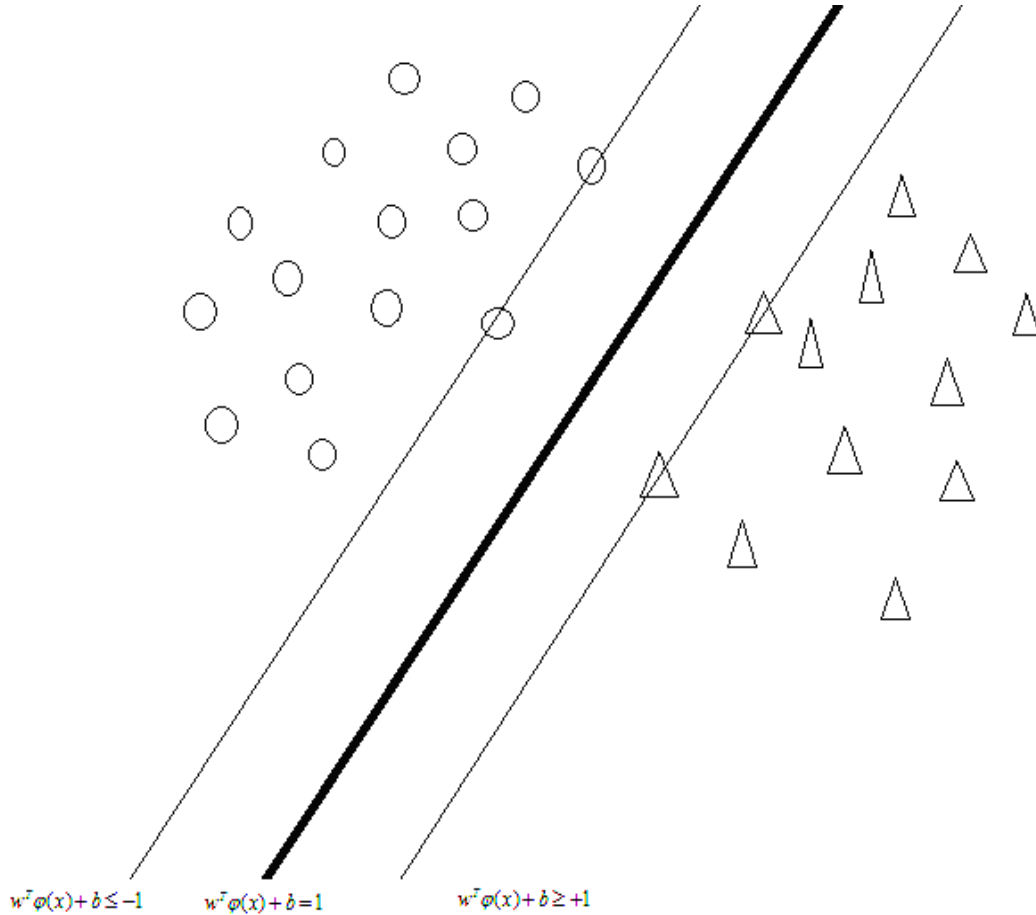


Figure 8. Working of SVM for the nonlinear separable data (triangle symbol used for cancerous and circle for noncancerous class).

The features of input data are separated according to the decision as: the features correspond to the output class ($y_n = +1$) if;

$$w^T \phi(x_n) + b \geq +1 \quad (4.14)$$

Similarly, the features correspond to output class ($y_n = -1$) if;

$$w^T \phi(x_n) + b \leq -1 \quad (4.15)$$

These two sets of inequalities as in eqn-4.14 and eqn-4.15 are combined into one single set, which stated as:

$$y(w^T \phi(x_n) + b) \geq 1, \text{ where } n=1, 2, \dots, K \quad (4.16)$$

The combination of eqn-4.13 and eqn-4.16 corresponds to a primal optimization problem, which is given as:

$$\min_{w, b, \xi} J_{\text{Primal}}(w, \xi) = \frac{1}{2} w^T w + c \sum_{n=1}^K \xi_n, \text{ such that, } y(w^T \phi(x_n) + b) \geq 1 - \xi_n \quad (4.17)$$

Where $\xi_n \geq 0$ and $n=1, 2, \dots, K$.

To solve this primal problem, the lagrangian is introduced, which is given as:

$$L(w, b, \xi, \alpha, v) = J_p(w, \xi) - \sum_{n=1}^K \alpha_n (y_n [w^T \phi(x_n) + b] - 1 + \xi_n) - \sum_{n=1}^K v_n \xi_n \quad (4.18)$$

Where α_n and v_n are the lagrangian multipliers for $n=1, 2, \dots, K$.

The solution of eqn-4.18 is obtained, by differentiating it with respect to w, b and ξ_n , then

assigning to zero value, which is given as:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{n=1}^K \alpha_n y_n \phi(x_n) \quad (4.19)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{n=1}^K \alpha_n y_n = 0 \quad (4.20)$$

$$\frac{\partial L}{\partial \xi_n} = 0 \rightarrow 0 \leq \alpha_n \leq c, n = 1, \dots, K \quad (4.21)$$

The quadratic optimization problem (Dual problem) corresponding to the primal problem is obtained as:

$$\max_{\alpha} J_{Dual}(\alpha) = -\frac{1}{2} \sum_{n,l=1}^K y_n y_l K(x_n, x_l) \alpha_n \alpha_l + \sum_{n=1}^K \alpha_n \text{ Such that } \sum_{n=1}^K \alpha_n y_n = 0 \quad (4.22)$$

In this quadratic optimization task the kernel function used i.e. ‘K’ which maps the input features to a high dimensional feature space. Mathematically the kernel function is defined as:

$$K(x_n, x_i) = \phi^T(x_n) \phi^T(x_i) \quad (4.23)$$

So far various kernel functions are described in SVM literature for this Non-linear mapping of the input features. These are linear kernel, polynomial kernel, Gaussian Kernel, RBF kernel and MLP kernel. Mathematically these functions are expressed as below:

Linear kernel: $K(x_n, x_i) = x_n x_i$,

Polynomial Kernel: $K(x_n, x_i) = (x_n x_i + a)^b$, where $b > 2$ and $a > 0$.

RBF Kernel: $K(x_n, x_i) = e^{-\frac{|x_n - x_i|^2}{2\sigma^2}}$, where σ is the Standard deviation of Gaussian curve.

The new test data evaluated with respect to the kernel trick is given as:

$$y(x) = \sum \alpha_i y_i k(x_n, x_i) + b \quad (4.24)$$

4.4.2 Least square Support vector machine (LS-SVM) as Classifier

Least Squares Support Vector Machines (LS-SVM) is the modifications of SVM and have been introduced in with the goal of obtaining a learning model by solving a set of linear equations instead of more complex (i.e. quadratic programming) problems that need to be solved for the classic SVM problem (Yusof and Mustaffa 2011). The formulation of LS-SVM is same As that of SVM i.e. Initially the training data set is given as $\{x_i, y_i\}_{i=1}^N$ where the input data is $x_i \in R^n$ and

corresponding output as binary class labels which is denoted by $y_i \in \{-1, +1\}$. Further these sets of inputs are mapped to higher dimensional feature space by using a nonlinear mapping function $\phi(x)$. The overall architecture of the LS - SVM model was shown in the Fig. 9.

The equation of the Hyperplane for the LSSVM to separate the binary classes is given as:

$$y_i = w^T \phi(x_i) + b \quad (4.25)$$

The modifications in the cost function and primal optimization problem for LSSVM as that of SVM is given as:

$$\min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (4.26)$$

$$\text{Such that, } y_i [w^T \phi(x_i) + b] = 1 - e_i, i=1, 2, 3, \dots, N \quad (4.27)$$

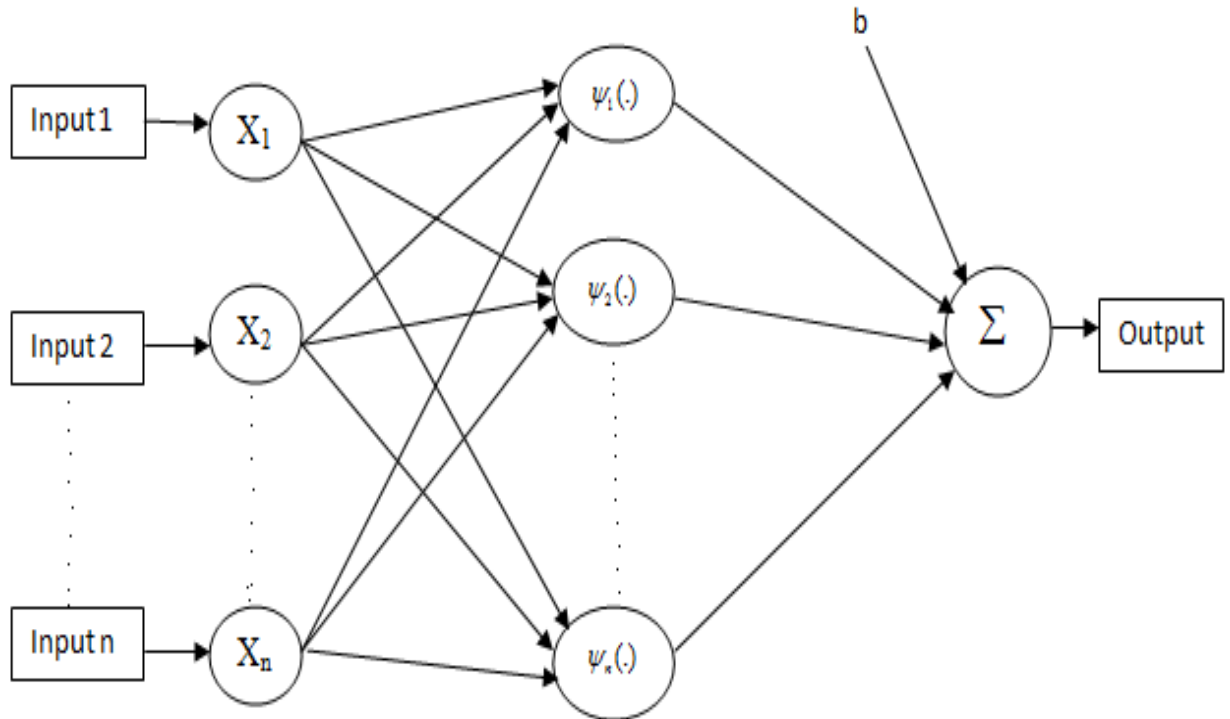


Figure 9. Architecture LS-SVM ('n' is the number of support vectors, $x_1, x_2, x_3, \dots, x_n$ are input feature vectors and $K_1, K_2, K_3, \dots, K_n$ are kernel functions).

For solving this primal optimization problem in LSSVM, the Lagrange function is constructed as:

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i \{w^T \phi(x_i) + b + e_i - y_i\} \quad (4.28)$$

Where ‘ α_i ’ are the Lagrange multipliers: The values of α_i is positive or negative according to the equality constraint, as in eqn-4.27.

The condition of optimality confirms that:

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i y_i \phi(x_i) \quad (4.29)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \quad (4.30)$$

$$\frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i, \quad i=1,2,3,\dots,N \quad (4.31)$$

$$\frac{\partial L}{\partial \alpha_k} = 0 \Rightarrow y_i [w^T \phi(x_i) + b] - 1 + e_i = 0 \quad (4.32)$$

The corresponding Dual function for the primal problem is given as:

$$y(x) = \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \quad (4.33)$$

Where K is the kernel function and is defined as:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (4.34)$$

The value of α and b is obtained by solving the set of linear equations, which is given as:

$$A \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (4.35)$$

Where A is a square matrix and defined as:

$$A = \begin{bmatrix} K + \frac{I}{\gamma} & 1_N \\ 1_N^T & 0 \end{bmatrix} \quad (4.36)$$

K denotes the kernel matrix, I denotes the identity matrix $N \times N$ with value as $1_N = [1 \ 1 \ 1 \dots 1]^T$.

Hence the solution of α and b is given as:

$$\begin{bmatrix} \alpha \\ b \end{bmatrix} = A^{-1} \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (4.37)$$

The test data point x_j to be evaluated as follows:

$$y_j = \sum_i^N \alpha_i y_i K(x_i, x_j) + b \quad (4.38)$$

For LS-SVM, there is various kernel functions used i.e. Radial basis function (RBF), linear polynomial, MLP kernel. These kernel functions are defined as:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_{sv}^2}\right)$$

$$K(x_i, x_j) = (x_i^T x_j + t)^d$$

Where, σ_{sv}^2 is the squared variance of the Gaussian function and 'd' is the degree of polynomial function.

Here, the input data are normalized before being processed in the classifier. The normalization is done as follows (Alpaydin 2004):

$$in_{i,\max} = \max(in_i(p)) \quad p=1 \dots N_p, i=1, \dots, N_i \quad (4.39)$$

Where N_p is the number of patterns in the training set, N_i is the number of neurons in the input

layer and in is the input parameters to the classifiers. After Normalization by these maximum values, the input variables lie in the range 0 to 1, which is given as:

$$in_{i,nor}(p) = \frac{in_i(p)}{in_{i,max}}, \quad p = 1, \dots, N_p, \quad i = 1, \dots, N \quad (4.40)$$

4.5 Performance Evaluation Criteria for Classifier

The performance of classifier depends on various factors like CR, Sensitivity, Specificity and the area under Receiver operating characteristics (ROC) curve (Lu et al. 2004). These values are calculated by considering confusion matrix, which is given as below:

Table 1. General procedure to construct confusion matrix for a binary classifier

TP	FN
FP	TN

Where, TP: Number of True Positives, FP: Number of False positives, TN: Number of True negatives, FN: Number of False negatives.

$$CR = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4.41)$$

Sensitivity-It is the statistical measure of how well a binary classifier correctly identifies the positive cases.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4.42)$$

Specificity-It is also a statistical measure of how well a binary classifier correctly identifies the negative class.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4.43)$$

Specificity is also defined as, $\text{Specificity} = 1 - \text{false alarm rate}$, Therefore, the false alarm rate can be calculated as:

$$\text{False alarm rate} = 1 - \text{Specificity} = \frac{FP}{TN + FP} \quad (4.44)$$

ROC Curve-This curve mainly used in signal detection theory to depict the relationship between the hit rate (specificity) and the false alarm rate in a noisy communication channel. It is also marked as the technique for organizing classifiers and visualizes their performance. The performance of classifier with respect to ROC curve can be determined by considering area under the curve (AUC) (Sing et al. 2005). If the $\text{AUC}(a1) > \text{AUC}(a2)$ then classifier 'a1' has better performance than that of classifier 'a2'.

In the next chapter i.e. chapter 5, I discussed the results in terms of various performance parameters (CR, Sensitivity, Specificity, Area under ROC curve) to evaluate the best classifier for the classification of breast cells into cancerous and non-cancerous class.

CHAPTER 5:

RESULTS & DISCUSSION

5. Results and discussions

Initially, the image segmentation method is used to segment the cells from image backgrounds. The results obtained during step by step segmentation for the MDA-MB-31 sample cancer image is given as in Fig. 10, 11, 12, 13.



Figure 10. Binary gradient image obtained after tuning Sobel matrix with MDA-MB-231 cultured cell image.

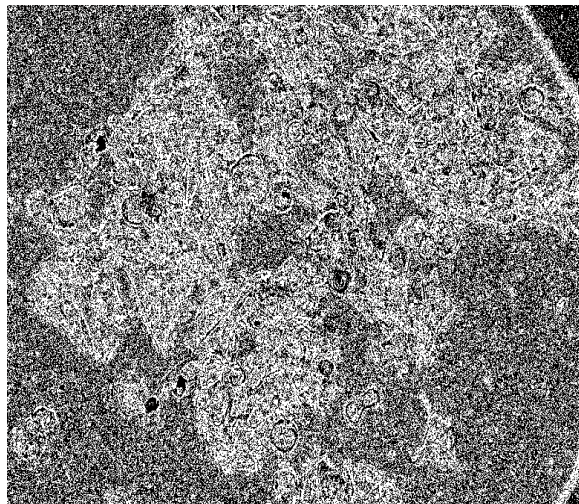


Figure 11. Dilated Gradient image of MDA-MB-231 cancer cell.

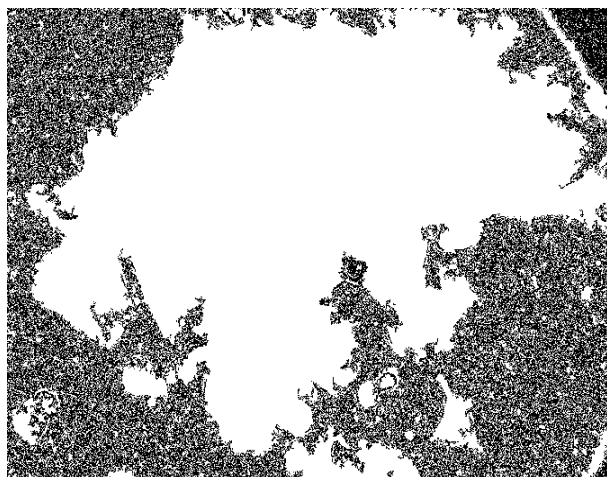


Figure 12. Dilated Gradient image of MDA-MB-231 cancer cell after filling interior gaps.

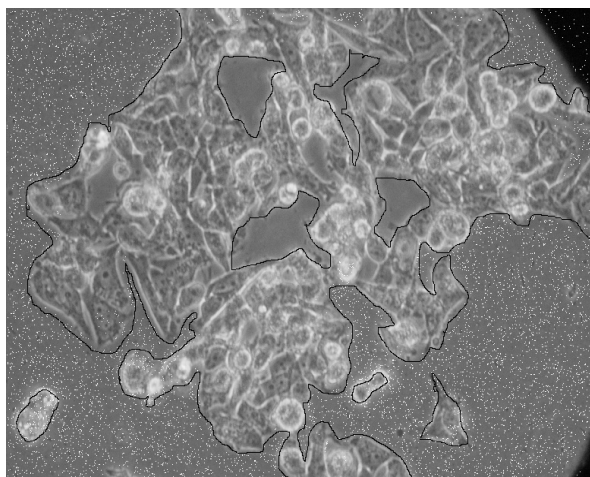


Figure 13. Segmented cells from the image background with region of interest of MDA-MB-231 cell.

Similarly, the segmented MCF-7 Cell image obtained after final step during image segmentation (explained in chapter-3) is shown in Fig 14.

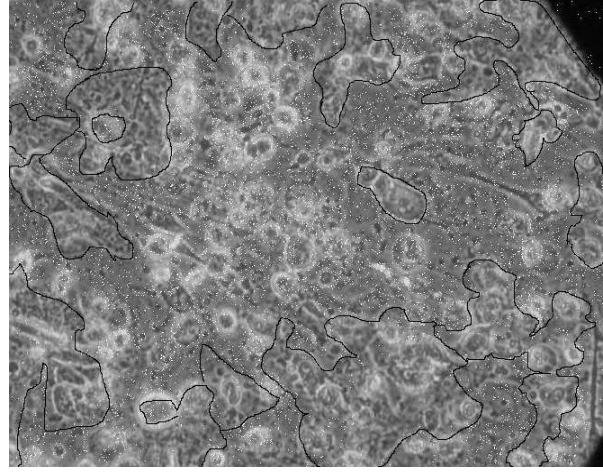


Figure 14. Segmented cells from the image background with region of interest

After segmentation, the required texture features (i.e. chip histogram based texture features, tamura texture features and wavelet based texture features) are extracted from each segmented images. The description of each features are also discussed in chapter-2. From these feature extraction techniques, I got a sum total of 90 features. These 90 numbers of features are extracted for a single segmented image. Similarly for 81 numbers of images the same texture feature extraction procedure is followed, to make the feature dataset.

The dataset, I obtained mainly consist of the number of instances which are in the rows and the feature vectors are, in the column. The size of the dataset is found to be: 81×90 . Further, I applied the Principal component analysis (PCA) algorithm to reduce the size of dataset i.e. from 90 numbers of features to 15 numbers of features. Hence, the size of the reduced dataset after PCA is given as: 81×15 . For the classification purpose, I marked the cancer image instances as logically '1' and non-cancer image instances as logically '0'.

The features in the reduced dataset for different instances are taken as the inputs to the KNN, ANN, SVM and LS-SVM classifier for both training and testing purpose. The outputs used for the training of classifiers are the binary values ('1' for cancerous and '0' for non-cancerous cell). Here, I consider 50% of instances as training and the rest of the instances as testing of the reduced data set. Initially the classifiers are trained, to get optimized training performance. After getting the optimized training performance parameters, the test data are evaluated. Then the predicted output and actual test output are compared to get the classifier performance. The performance of classifier is given as:

5.1 KNN Classifier performance

The parameters used to get optimized training performance are the number of nearest neighbors (K) and the minimum Euclidean distance from the nearest neighbors. After setting the value of K as 2 and the distance as Euclidean, I got the optimized training performance. Further the test data are evaluated by this optimized training performance and the performance of KNN classifier is obtained in terms of classification rate, Which is found to be 83.8506% for K=2. The variation of the classification rate with respect to number of nearest neighbors 'K' is shown in Fig. 15. From this figure, the classification rate for KNN classifier changes when the number of nearest neighbors (K) varies.

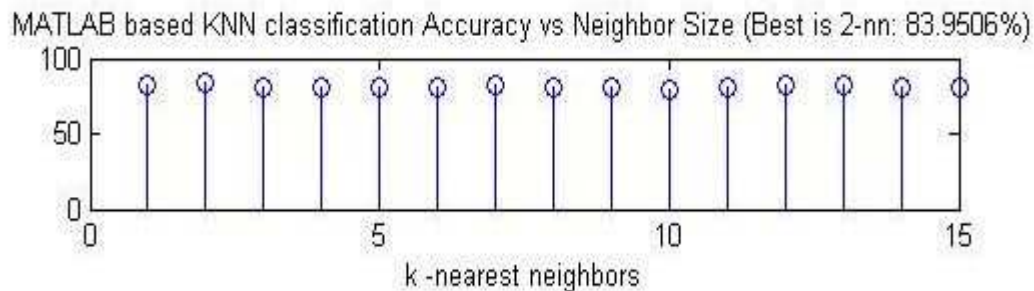


Figure 15. Variation of Classification rate with respect to 'K' of the KNN classifier

5.2 MFNN Classifier performance

For MFNN classifier the different parameters used to evaluate the optimized training performance are the number of hidden neurons, learning rate, Momentum factor, Number of iterations and Training parameter goal. Here, I tuned these parameters to get optimized training performance. By fixing the values of these parameters as in Table 2, I got the optimized training performance.

Table 2. Various Parameters setting for MFNN to get optimized training performance

MFNN parameters	values
Number of hidden neurons	9
Learning rate	0.05
Momentum factor	0.7
Number of iterations	1000
Training parameter goal	10^{-9}

These variations of the Mean Square error (MSE) with respect to the number of iterations are shown in Fig 16. After 632 numbers of iteration, the optimal training goals achieved and further the test data are evaluated simply by passing it to that optimized trained MFNN network. The performance of testing data evaluation for the MFNN network is given in terms of CR, Sensitivity, Specificity and area under the ROC curve.

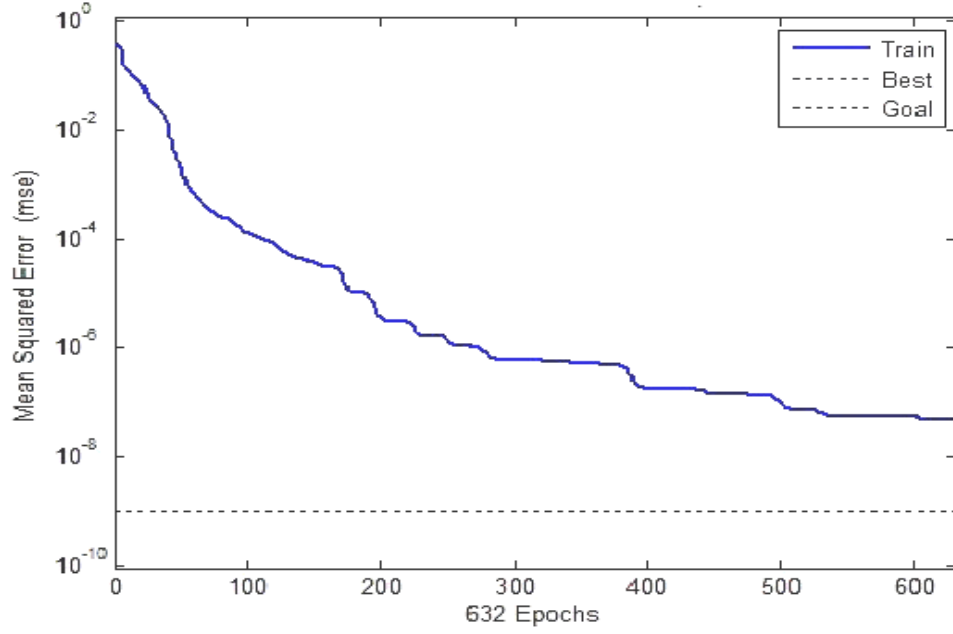


Figure 16. Variations of training MSE with respect to number of iterations

Table 3. Confusion matrix after testing data evaluation for MFNN

Predicted \ True	Cancerous	Noncancerous
	Cancerous	Noncancerous
Cancerous	11	3
Noncancerous	1	28

The confusion matrix obtained during the evaluation of testing data is given in Table 3. From this confusion matrix the CR, Sensitivity and Specificity are calculated, which is given as:

$$CR = \frac{(11+28)}{(11+28+1+3)} \times 100 = 90.6977\%, \text{ Sensitivity} = \frac{11}{11+3} = 0.785, \text{ Specificity} = \frac{28}{28+1} = 0.965$$

5.3 SVM Classifier performance

To obtain the optimized training performance in the case of the SVM classifier with both RBF kernel and linear kernel functions, I tuned the following training parameters as the Number of Support vector and number of iterations. These parameters are shown in the Table 4 and Table 5.

Table 4. Optimized parameters during Training phase of SVM by considering RBF kernel

SVM parameters	Values
Number of Support vector	37
Number of Iterations	1000

Table 5. Optimized parameters during Training phase of SVM by considering linear kernel

SVM parameters	Values
Number of Support vector	17
Number of Iterations	1000

Further the test data are evaluated by passing it to the optimized trained SVM with linear and RBF kernel. The performance of the evaluation of test data is given in terms of confusion matrix and ROC plot. After calculating the confusion matrix (as in Table 6 for RBF kernel SVM and Table 7 for the linear kernel SVM classifier) and ROC curve (as in Fig. 17(A) for RBF kernel SVM and Fig. 17 (B) for the linear kernel SVM classifier), from the comparisons of predicted outputs and actual outputs; I used certain calculations to find CR, Sensitivity, Selectivity and the area under the ROC curve, which is given as:

Table 6. Confusion matrix for SVM with RBF kernel

True Predicted	Cancerous	Noncancerous
Cancerous	2	0
Noncancerous	10	31

$$CR = \frac{(2+31)}{(2+31+0+10)} = 76.74\%, \text{ Sensitivity} = \frac{2}{2+0} = 1, \text{ Specificity} = \frac{31}{31+10} = 0.756$$

Table 7. Confusion matrix for SVM with linear kernel

True Predicted	Cancerous	Noncancerous
Cancerous	12	3
Noncancerous	0	28

$$CR = \frac{(12+28)}{(12+28+0+3)} = 93.02\%, \text{ Sensitivity} = \frac{12}{12+3} = 0.8, \text{ Specificity} = \frac{28}{28+0} = 1$$

From the figure 17 (A) and 17 (B) ; it is quiet obvious that the Area under the ROC curve for Linear kernel is 0.95161, which is higher than that of RBF kernel SVM.

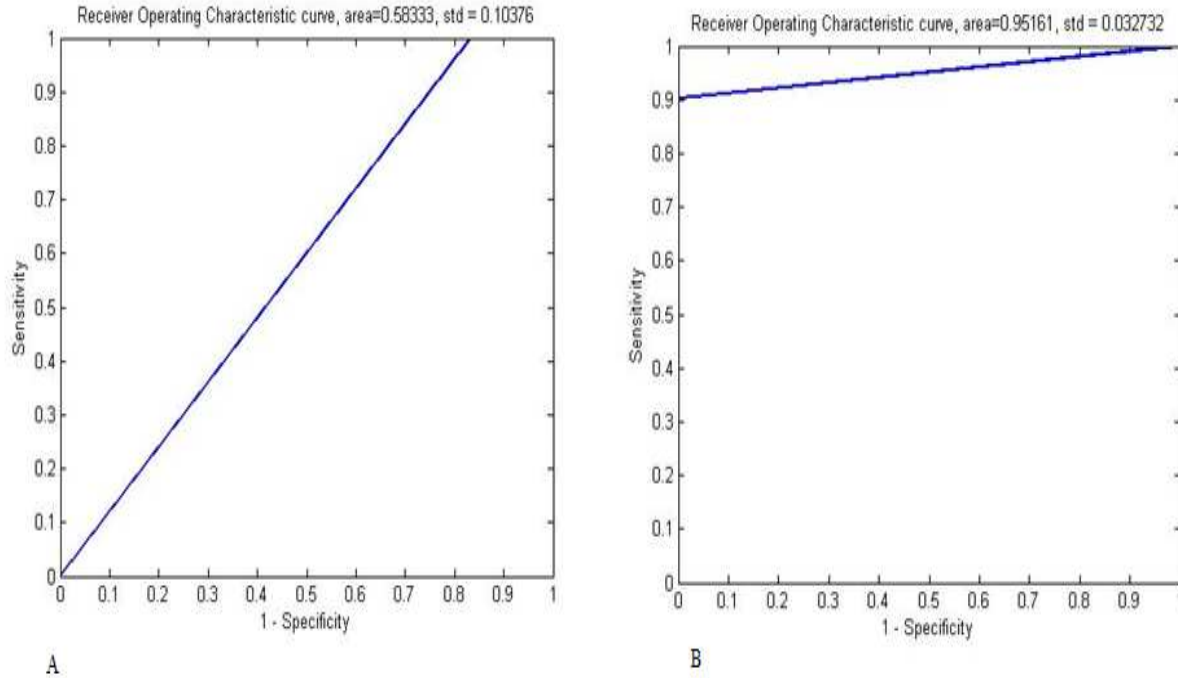


Figure 17. (A) ROC plot for RBF kernel SVM classifier for testing data evaluation. (B) ROC plot for linear kernel SVM classifier for testing data evaluation.

5.4 LS-SVM Classifier performance

The different parameters used to obtain optimized training performance for the Linear and RBF kernel LS-SVM classifier is given in Table 8 and Table 9. After training the distribution of classes with respect to the most appropriate features (important features, which distinguish the classes) for both linear and RBF kernel LS-SVM classifier is shown in Fig. 18(A) and Fig. 18(B).

Table 8. Optimized parameters during Training phase of LS-SVM by considering linear kernel.

LS-SVM parameters	Values
Kernel type	Linear
Gamma(γ)	5.169

Table 9. Optimized parameters during the Training phase of RBF kernel LS-SVM classifier.

LS-SVM parameters	Values
Kernel type	Linear
Gamma(γ)	128.3341
Sig2(σ^2)	59.4891

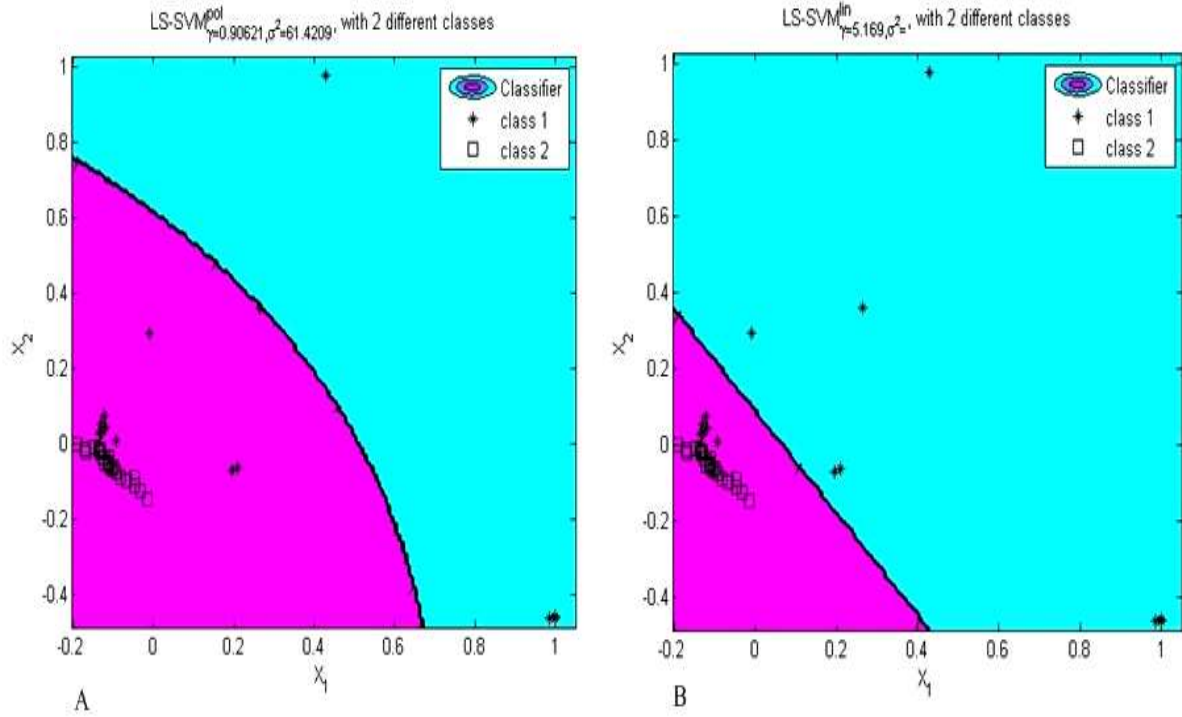


Figure 18. (A) Training data separation with respect to Hyperplane for linear kernel LS-SVM. (B) Training Data separation with respect to Hyperplane for linear kernel LS-SVM.

Further, the test data are evaluated by passing through the optimized trained LS-SVM models with linear and RBF kernel functions. The performance of both LS-SVM classifiers is obtained in terms of classification rate (CR), Sensitivity, Specificity and area under the ROC curve. These parameters are calculated from the confusion matrix: which is given in Table 10 and Table 11. After obtaining the Classification rate, the ROC plot is shown in Fig. 19(A) and Fig. 19(B).

Table 10. Confusion matrix for linear kernel LS-SVM classifier

True Predicted	Cancerous	Noncancerous
Cancerous	12	2
Noncancerous	0	29

$$CR = \frac{(12 + 29)}{(12 + 2 + 0 + 29)} \times 100 = 95.3488\%, \text{ Sensitivity} = \frac{12}{12 + 2} = 0.857, \text{ Specificity} = \frac{29}{29 + 0} = 1$$

Table 11. Confusion matrix for RBF kernel LS-SVM classifier.

True Predicted	Cancerous	Noncancerous
Cancerous	12	2
Noncancerous	0	29

$$CR = \frac{(12 + 29)}{(12 + 2 + 0 + 29)} \times 100 = 95.3488\%, \text{ Sensitivity} = \frac{12}{12 + 2} = 0.857, \text{ Specificity} = \frac{29}{29 + 0} = 1$$

From the Fig. 19 (A) and (B), it is clear that the area under the ROC of RBF kernel LSSVM classifier is 0.96774 which is best among all other classifiers.

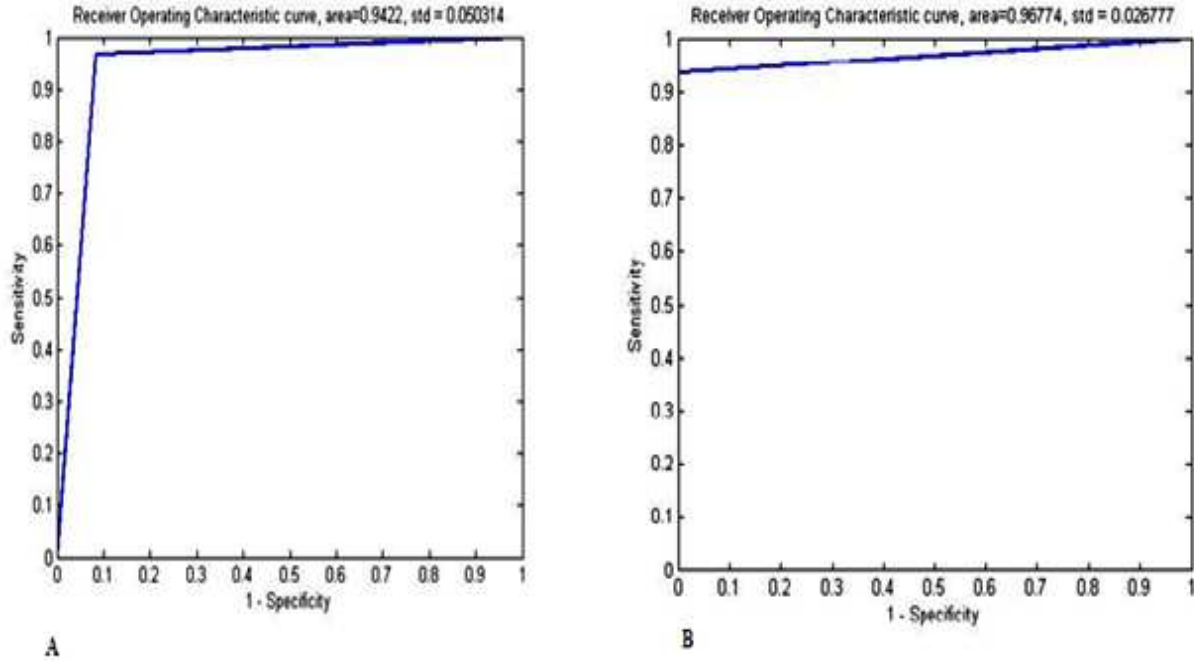


Figure 19. (A) ROC plot for linear kernel LS-SVM classifier for testing data evaluation. (B) ROC plot for RBF kernel LSSVM classifier for testing data evaluation.

5.5 Comparisons of the Classifiers Performance

After calculating the Classification rate, sensitivity, specificity and area under the ROC curve for the various classifiers, I compared these values to obtain the best classifier for the desired classification task, as discussed in chapter-3. The Table 12 gives the comparison results with respect to performance parameters (Classification rate, sensitivity, specificity and area under the ROC curve) for the classifiers. From these comparisons, I got the outcome as the RBF kernel based LS-SVM classifier gives better performance with a CR of 95.3488%, Specificity of unity and the area under the ROC curve of 0.96774, which is much better than the other classifiers.

Table 12. Comparisons of CR, sensitivity, Specificity and area under ROC curve for classifiers

Name of the Classifier	CR (%)	Sensitivity	Specificity	Area under the ROC curve
KNN classifier	83.8506%	N/A	N/A	N/A
MFNN classifier	90.6977%	0.785	0.965	0.9245
SVM with RBF kernel classifier	76.74%	1	0.756	0.58333
SVM with Linear kernel classifier	93.02%	0.8	1	0.95161
LS-SVM with Linear kernel classifier	95.3488%	0.857	1	0.9422
LS-SVM with RBF kernel classifier	95.3488%	0.857	1	0.96774

CHAPTER 6:

CONCLUSION & FUTURE WORK

Conclusion

The selected LS-SVM model successfully exhibited a better predictive capability with a handsome classification rate (CR) for testing data as compared to other classifiers. It was found that LS-SVM model produced highest CR i.e. 95.3488% which is so far highest. Although, the SVM Linear kernel classifier produced 93.02% CR which is close to LS-SVM-derived CR, other classifiers like KNN, ANN were far less accurate compared to both SVM and LS-SVM. Accurately detecting/diagnosing the disease like breast cancer in the early stage is extremely essential for fast recovery or to avoid the death probability. If we achieve more than 95% accuracy by the use of machine learning techniques, the fact will confirm to diagnose the disease in an easy and fast manner. This may lead to reduce the time of detecting and analyzing the cancer cells and prevent death in cancer like lethal disease.

Future Work

The future development of this application might be the followings:

- (i) Implementing the same pattern recognition technique for other diseases like Lungs cancer and other form of cancers.
- (ii) Implementation of Gabor wavelet based feature extraction method and Curvelet based feature extraction method to extract more features from each cancer and non-cancer class of images.
- (iii) Implementation of Fuzzy support vector machine (FSVM) classifier, Relevance vector machine (RVM) classifier and Ensemble classifiers to classify the cancerous and noncancerous cells.

References

- Akay MF (2009) Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications* 36(2, Part 2):3240-3247 .
- Alpaydin E (2004) Introduction to machine learning. The MIT Press.
- Anoraganingrum D Cell segmentation with median filter and mathematical morphology operation. In: *Image Analysis and Processing, 1999 Proceedings International Conference on*, 1999 1999. p 1043-1046.
- Arodź T, Kurdziel M, Sevre EOD, Yuen DA (2005) Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms. *Computer Methods and Programs in Biomedicine* 79(2):135-149.
- Arora N, Martins D, Ruggerio D, Tousimis E, Swistel AJ, Osborne MP, Simmons RM (2008) Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. *American journal of surgery* 196(4):523-6.
- Brook A, El-Yaniv R, Isler E, Kimmel R, Meir R, Peleg D (2006) Breast Cancer Diagnosis From Biopsy Images Using Generic Features and SVMs.
- Campbell C, Ying Y (2011) Learning with support vector machines. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 5(1):1-95.
- Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.
- Dua S, Acharya UR, Chowriappa P, Sree SV (2012) Wavelet-Based Energy Features for Glaucomatous Image Classification. *Information Technology in Biomedicine, IEEE Transactions on* 16(1):80-87.
- Fatima B, Amine CM (2012) A Neuro-Fuzzy inference model for breast cancer recognition. *International Journal of Computer Science & Information Technology* 4(5).
- Gokhale S (2009) Ultrasound characterization of breast masses. *The Indian journal of radiology & imaging* 19(3):242-7.
- Haralick RM, Shanmugam K, Dinstein IH (1973) Textural Features for Image Classification. *Systems, Man and Cybernetics, IEEE Transactions on* SMC-3(6):610-621.
- Hastie T, Tibshirani R, Friedman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* 27(2):83-85.
- Haykin S, Network N (2004) A comprehensive foundation. *Neural Networks* 2.

- Haykin SS (2009) Neural networks and learning machines, vol 3. Prentice Hall New York.
- Hong J, Kandasamy K, Marimuthu M, Choi CS, Kim S (2011) Electrical cell-substrate impedance sensing as a non-invasive tool for cancer cell study. *The Analyst* 136(2):237-45.
- Issac Niwas S, Palanisamy P, Chibbar R, Zhang WJ (2012) An expert support system for breast cancer diagnosis using color wavelet features. *Journal of medical systems* 36(5):3091-102.
- Jele U, Fevens T, Krzy A (2008) Classification of Breast Cancer Malignancy Using Cytological Images of Fine Needle Aspiration Biopsies. *Int J Appl Math Comput Sci* 18(1):75-83.
- Jinshan T, Rangayyan RM, Jun X, El Naqa I, Yongyi Y (2009) Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. *Information Technology in Biomedicine, IEEE Transactions on* 13(2):236-251.
- Kaiwei C, Xiaoqing L, Jianguo S, Xiao W Blind Image Tampering Identification Based on Histogram Features. In: *Multimedia Information Networking and Security (MINES), 2011 Third International Conference on*, 4-6 Nov. 2011. p 300-303.
- Kambhatla N, Leen TK (1997) Dimension reduction by local principal component analysis. *Neural Computation* 9(7):1493-1516.
- Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. *Expert Syst Appl* 36(2):3465-3469 doi:10.1016/j.eswa.2008.02.064.
- Kingsbury N (1998) Wavelet transforms in image processing. *Signal Analysis and Prediction*:27
- Lu Z, Szafron D, Greiner R, Lu P, Wishart D, Poulin B, Anvik J, Macdonell C, Eisner R (2004) Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics* 20(4):547-556.
- Martins LDO, Junior GB, Silva AC, Paiva ACD, Gattass M (2009) Detection of Masses in Digital Mammograms using K-means and Support Vector Machine. *Electronic Letters on Computer Vision and Image Analysis* 8(2):39-50.
- Mohanty S, Ghosh S (2010) Artificial neural networks modelling of breakdown voltage of solid insulating materials in the presence of void. *Science, Measurement & Technology, IET* 4(5):278-288.
- Moradmamand H, Setayeshi S, Targhi HK (2011) Comparing Methods for segmentation of Microcalcification Clusters in Digitized Mammograms *International Journal of Computer Science* 8(6):104-108.
- Mousa R, Munib Q, Moussa A (2005) Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. *Expert Systems with Applications* 28(4):713-723.

- Parkin DM, Bray F, Ferlay J, Pisani P (2001) Estimating the world cancer burden: Globocan 2000. *International journal of cancer* 94(2):153-156.
- Rangayyan RM (2004) *Biomedical image analysis*. CRC press.
- Rangayyan RM, El-Faramawy N, Desautels JEL, Alim OA (1997) Measures of acutance and shape for classification of breast tumors. *Medical Imaging, IEEE Transactions on* 16(6):799-810.
- Rejani YIA, Selvi ST (2009) Early detection of breast cancer using SVM classifier technique. *International Journal on Computer Science and Engineering* 1(3):127-130.
- Rojas Domínguez A, Nandi AK (2009) Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognition* 42(6):1138-1148.
- Schaefer G, Závisek M, Nakashima T (2009) Thermography based breast cancer analysis using statistical features and fuzzy classification. *Pattern Recognition* 42(6):1133-1137.
- Shi X, Cheng HD, Hu L, Ju W, Tian J (2010) Detection and classification of masses in breast ultrasound images. *Digital Signal Processing* 20(3):824-836.
- Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCr: visualizing classifier performance in R. *Bioinformatics* 21(20):3940-3941.
- Sivanandam S, Deepa S (2006) *Introduction to neural networks using Matlab 6.0*. Tata McGraw-Hill Education.
- Srinivasaraghavan V, Strobl J, Agah M (2011) Detection of breast cancer cells in tri-culture using Impedance Spectroscopy. Paper presented at the 15th International Conference on Miniaturized Systems for Chemistry and Life Sciences, Seattle, Washington, USA, October 2-6.
- Stathakis D (2009) How many hidden layers and nodes? *International Journal of Remote Sensing* 30(8):2133-2147.
- Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. *Neural processing letters* 9(3):293-300.
- Tamura H, Mori S, Yamawaki T (1978) Textural Features Corresponding to Visual Perception. *Systems, Man and Cybernetics, IEEE Transactions on* 8(6):460-473.
- Tan TZ, Quek C, Ng GS, Ng EYK (2007) A novel cognitive interpretation of breast cancer thermography with complementary learning fuzzy neural memory structure. *Expert Systems with Applications* 33(3):652-666.
- Unser M, Aldroubi A (1996) A review of wavelets in biomedical applications. *Proceedings of the IEEE* 84(4):626-638 doi:10.1109/5.488704

- Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometrics and intelligent laboratory systems* 2(1):37-52.
- Yusof Y, Mustaffa Z (2011) Dengue Outbreak Prediction: A Least Squares Support Vector Machines Approach.
- Zolghadrasli AR, Maghsoodzadeh Z Modification of Mammograms for early diagnosis of breast cancer using wavelet and Neural networks. In: *Computer Systems and Applications, 2006 IEEE International Conference on*, March 8, 2006 2006. p 288-295.

Dissemination

Journals:

1. Rajesh Kumar Tripathy, Sailendra Mahanta, Subhankar Paul, “Feature identification and cellular classification of breast cancer using Artificial Intelligence Techniques”, *Journal of PLOS One Computational Biology*, 2013 (In review).
2. Rajesh Ku. Tripathy, Subhankar Paul, “Use of Least Square Support Vector Machine as alternative to support vector machine for Breast cancer Diagnosis”, *International Journal of Computer Science and Technology*, Vol. 4, Issue 1, 2013 (In Press).