Thesis submitted to the College of Science and Engineering in partial fulfilment of the requirements for the degree of Master of Science (Computer Science) at Flinders University, Adelaide, Australia.

COMP9700-Master Thesis Classification of masses in dense mammograms using Enhanced Local Ternary Pattern.

Author: Maryam Hammad Almaeen

Supervisor :

Dr. Mariusz Bajger

Declaration

I certify that this work does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Maryam Hammad Almaeen
24/11/2019

Contents

Abstract

The World Health Organization (WHO) has declared breast cancer as the second leading cause of cancer death in adult women worldwide after lung cancer. The possibility that breast cancer will result in a woman's death is 2.6% (1 in 38). The high mortality rate of breast cancer is due to imperfect detection techniques available. Technology used for diagnosis or mammography has utmost importance in clinical research, as a mammogram image provides a detailed information. This study proposes a technique that uses regions of interest to classify lesions present in mammogram. The proposed method utilizes an extended local ternary pattern to extract the feature vector from regions of interest. In mammograms, information about texture plays a vital role in the classification of lesions. Therefore, the extended local ternary pattern is adopted in order to give information in depth texture features of the regions of interest (ROIs). To classify the lesions, different machine learning algorithms are used such support vector machine, k-nearest neighbours, and artificial neural networks classifier. The Digital Database for Screening Mammography (DDSM) is used which is publicly available. In total 101 mammograms are considered, out of 51 malignant mammograms and 50 benign mammograms. Then, 1302 benign ROIs, and 1632 ROIs malignant are extracted. To standardize the ROIs, each ROI is kept fixed in size, which is 51x51. In this study, efforts are made to propose a model that can be used for effective classification of malignant and benign regions of interest, so that efficient and early diagnosis of breast cancer can be made possible. The proposed technique using KNN classifier achieved the highest sensitivity of 88.73%, and achieved AUC value of 93.89% with 17 patterns.

Acknowledgment

First of all, I would like to thank my supervisor for his valuable guidance at every step of my whole journey of this thesis. Further, I would like to thank my family for constant support in this whole meaningful journey.

1 Introduction

Breast cancer is one of the most common types of cancer among women. It is responsible for more number of deaths among women, than any other kind of cancer across the world. It is the second most common cause of death in the West, after lung cancer [1][2]. According to research, conducted by a leading breast cancer awareness organization, Breastcancer.org, about 12% of women in the world, may have breast cancer in their lifetime, making it the second most common type of cancer with incidence increasing every year [1]. Breast cancer can occur in any individual, irrespective of their age and gender. It, not only affects females but also affects a large number of males. The number of patients of all genders are increasing every year, and the risk of breast cancer in males is also at its peak. In United States of America alone, roughly 41,760 women are estimated to die from breast cancer this year [2].

It is very difficult to detect breast cancer in its early stages, due to the small number of cancer cells present in the beginning. In most cases patients do not feel any pain, they just notice a lump in their breast. In most cases breast cancer blows in the lymph nodes that results in swelling. There are few symptoms patients may experience such as, breast pain or a feeling of heaviness in the breast, persistent variations in breast such as puffiness, condensing of the breast, redness of the breast's skin, and abnormalities in the nipples such as unprompted release or retraction [3].

Figure 1: Breast Cancer Statistics by U.S. Breast Cancer Organization [1]. Image has been removed due to ccopyright issues.

The statistical research carried by the U.S. Breast Cancer Organization. As shown in Figure 1, many developed countries have a high number of breast cancer patients. Australia and New Zealand, in particular, display an alarming number of breast cancer patients. These statistics clearly indicate the prevalence of the disease. A better technique is needed to detect it at its initial stages to reduce the mortality rate. Breast cancer is a serious public health concern. Once breast cancer reaches at the higher stages, the chances of patient's survival drop dramatically, thus early detection is extremely important for curing the disease. In this regard, a systematic analysis of mammogram is made in order to diagnose breast cancer in its early stages. Computer-aided detection of breast cancer can be carried out using mammograms. A process which helps in the primary detection and diagnosis of this disease. Mammography is a process that uses a spectrum of X-rays to observe the internal structure of the breasts [4]. The images taken through the process of mammography are commonly known as mammograms. An X-ray (radiograph) is a non-invasive medical test that helps physicians to diagnose and treat medical conditions. For breast imaging, mammography utilizes the minor infiltration of ionizing radiation in order to capture images of the body covered by the skin of the breast. In this regard, the technique of mammography to take images of the breast is also modernized. One of the major advancements in the field of mammography is digital mammography. Digital mammography is a pictorial system in which an image is created using a film produced by X-rays. In this process, electronic circuits are used to translate the X-ray intensities into digital colours. The output of this system is a monochromatic image of the internal structure of the breast. This system is similar to the technology that can be found in digital cameras. These pictures of the internal structure of the breast, are extremely helpful in designing a proper diagnostic system using computer-aided techniques [4]. Mostly, mammograms are used to screen patients in order to detect breast cancer in its early stages. These images can also be used to perceive and analyse breast diseases. For screening purposes, mammograms play a vital role in the premature discovery of breast cancer, because these images are effective in predicting future changes in the breast. According to a research study, annual mammograms can help in the initial detection (CAD) of breast-related diseases, especially cancer [4].

Figure 2: Sample Mammogram images with malignant and no-malignant dense masses [33]. Image has been removed due to copyright issues.

Figure 2 shows two mammograms. The arrow in the left image shows the malignant region or region of interest (ROI) that is required for the detection of cancer. A small cancer can easily be seen in the fatty breast (left) as indicated by the arrow. The image in the right shows the presence of a larger cancer in breast. Processing of mammographic images is a widely open research field and the researchers have advanced the use of various computer aided detection techniques for proper detection of breast cancer and other breast related diseases. According to study in [5], features based on intensity of pixels provide meaningful information for detection of breast cancer as masses typically have higher difference in intensities as compared to other mass tissues. The drawback of using features based on intensity of pixels is that there is only a minor or no difference in the intensities of diseased part and not diseased part that appears dense. Some other features that are frequently used in detection of breast cancer are morphological features. These features are highly dependent of the process of segmentation as all these features are extracted from region of interest [6]. The dependency of these morphological operations on the process of segmentation make them incompatible as accurate segmentation is another challenging task in the detection of breast cancer. Some of the key researches to improve the process of segmentation of cancer deseased part and non deseased part, are described in [7], [12], [21] but design of computational efficient system that can be used in real time applications remains a challenging task. In some researches, commonly Grey Level Co-occurrence Matrix (GLCM) is used along with Histogram of Oriented Gradient (HOG) [8]. It is suggested in Ref. [8] that the efficiency of feature vector decreases in the presence of highly dense mass tissues. Along with the use of digital image processing and machine learning techniques, deep learning methods were also used by some researchers to improve the overall classification accuracy of breast cancer classification system. In this regard, convolutional neural networks (CNNs) were used in [9] for classification of breast masses. For a deep learning model such as convolutional neural network to exhibit high accuracy, huge number of training dataset is required. However, this large training mammographic images are not available in this case and hence this makes the use of CNN approach inefficient.

The development in the field of computer generated images have introduced many advanced algorithms which can help increase the performance of the computer-aided detection (CAD) system. This advancement in the field can be utilized to use efficient feature descriptor based on texture of the ROIs. Based on the observations listed above, segmentation of breast masses from thick and dense tissues is required for that a better algorithm is needed in order to produce accurate results. In order to use a machine learning algorithm for classification purposes, a powerful feature vector is required that can differentiate between diseased part and dense tissues that do not show signs of disease. For this study, a robust and efficient technique to distinguish between malignant mass and benign mass in dense mammograms is developed based on the Enhanced Local Ternary Pattern (ELTP) of texture. This technique is resistant to noise, hence making it beneficial in designing a proper diagnostic system.

2 Literature review

In the past few years, many research efforts have been made to design an accurate and computationally efficient diagnostic system that is capable of early detection of breast cancer by using mammograms. These studies include the use of both the image processing techniques and machine learning approaches. Some of these techniques are briefly discussed below. In [10] a method is proposed for the classification of mammographic images into mass and non-mass, based on the region of interest (ROI) extracted from mammograms. For experimental purposes, online available dataset namely the Digital Database for Screening Mammography commonly known as DDSM was used. In order to describe the texture of ROI, two parameters- the taxonomic diversity index and the taxonomic distinctness, were used. Moreover, two techniques that are internal and external masks were used for analysis of texture of ROI. Machine learning based classifier, namely support vector machine was employed in order to classify the input parameters into discrete classes.

In [12] a novel technique is proposed for classi- fication of masses in mammograms especially with dense mass tissues. The proposed work makes the use of local binary pattern technique to generate 9 structured super pixel patterns. . The proposed method was found to be efficient in segmenting out masses from dense tissues. The performance of classification system was tested on two freely available mammographic databases namely database for screening mammography (DDSM) and Breast Screen SA (BSSA). A total of 525 ROIs were used, 301 were extracted from DDSM and 224 were extracted from BSSA and Fisher linear discriminant analysis used as classifier. A ROC curve of 0.93 and 0.96 were achieved for BSSA respectively, using only six features. The results indicated that features that were generated using structured super pixel patterns were able to produce efficient and effective texture descriptors of breast masses in dense mass tissues. In [14] research segmentation of mammographic images was improved by combining the LSM algorithm with connected components technique for detection of breast cancer masses. Furthermore, statistical features including grey scale value, mean value of window and standard deviation were extracted from grey level images. These features were then used for extraction of region of interest (ROI). The results of the proposed method were evaluated using statistical evaluation parameters that are sensitivity and specificity. The proposed system displayed encouraging results of 81% sensitivity and 80% specificity. For experimental purposes, the DDSM dataset, available online, was used. The technique developed by Nguyen et al. [15] defines the successful use of Block Variance of Local Coefficients (BLVC) in the field of breast cancer detection. The proposed work successfully classifies the region of interest into two classes namely, masses and non-masses. The classifier used for this purpose was Support Vector Machine or simply, SVM. The experimentation is carried out using publicly available dataset namely Mini-MIAS database. Evaluation of 2700 ROIs detected from the database resulted in an ROC score of 0.93, showing BLVC features to be effective and efficient descriptors for massive lesions in mammograms. In [18] a method is proposed to classify breast cells into two categories which are normal and abnormal. The proposed classification system was based on ROIs which was extracted from mammograms of DDSM database as discussed previously. A powerful and robust feature vector was extracted by using Principal Components Analysis (PCA) technique which send that features as input to SVM and SVM achieved 98.83% sensitivity and 85.48% specificity. In 2013, a Computer Aided Detection (CAD) based methodology was proposed in [19], for the classification of breast cancerous masses. The proposed system in this research effort comprised of three main steps namely segmentation, feature extraction and classification. ROI extraction leads to extraction of powerful and robust feature vector. For the process of feature extraction, Spherical Wavelet Transform was used. In the final stage of classification, SVM were used as machine learning based classifier. A sensitivity of 97% and specificity of 91% was achieved in

[19]. In [22] researchers proposed Local Binary Pattern (LBP) to prepare a set of textural features for the classification of breast tissues. In [23] authors have proposed a system working on super pixel texture analysis for the classification of breast masses and their system have achieved an efficient performance. The authors have used 535 ROIs in total, with 301 extracted from DDSM and 234 extracted from a local database, with all localized in dense backgrounds of breasts. The AUC score obtained using only 4 features for DDSM was 0.957, and for local dataset, it was 0.891. Authors in [40] proposed a random forest, support vector machine (SVM and Artificial neural network (ANN) based approach. The experiment was conducted using digital mammograms obtained from National Hospital Organization, Nagoya Medical Centre, Nagoya, Japan. The database consisted of 322 ROIs for 201 lesions. These were obtained from a total of 186 individuals. They constructed histograms in order to determine the difference in patterns. The variances were discovered to be larger in malignant ROIs. Using the ANN, the AUC value achieved was 0.742. Authors in [24] have introduced a new model of local ternary pattern for the extraction of texture features of an image. As texture features are the strongest features, many methodologies for texture feature extraction are proposed for classification of masses in breast mammogram images. The authors have used the Outex database [32] for the purpose of their evaluation of model, which includes 24 classes of textures that have been collected under three illuminations at nine angles. In addition, they have used the CUReT database, containing 61 classes of real-world textures. They found that the ELTP performs better than LBP on the TC10 dataset obtained from Outex database. On the CURet database, ELTP outperforms LBP and achieves an appreciable classification rate. Local ternary pattern operators are utilized in [25] for extraction of a feature vector and these features were then embedded in support vector machine for the purpose of classification. They used the famous Mammographic Image Analysis Society (MIAS) database, consisting of 322 mammograms of 161 women. They achieved an AUC of 82.33% with their proposed method.

Hence, handcrafted features have shown promising results but when dense background is included, they do not show encouraging results. That's where the research gap is which can be filled by carrying out proper analysis to get best texture features whose possiblity of failure is low in any scenario and can be adopted for the noble cause of life saving.

3 Research hypothesis

In this research, a robust feature extraction and classification technique which is helpful in designing of proper diagnostic system, is represented. These features are based on Enhanced Local Ternary Pattern (ELTP) [24] and are able to distinguish between breast masses and dense background tissues. ELTP not only resistant to noise but also strictly invariant to grey-level transformations. The proposed system is computationally efficient and can be used in clinical applications. ELTP along with the efficient machine learning algorithm is used for classification of malignant and benign masses of breast mammogram images.

Therefore, ELTP can be introduced in order to detect breast cancer at initial stages and increase the performance of the CAD system. It is important to discuss here about the importance of selecting Extended Local Ternary Pattern (ELTP) against other pattern feature extraction techniques. According to research done in [27], ELTP is robust to give good performance even when noise immunity is high. In ELTP noise can be treated as the part in ROI. To evaluate noise immunity, authors of [27] suggested the histogram intersection between the original patterns and their noisy counterparts.

4 Image database

The Digital Database for Screening Mammography (DDSM) [41] is a well-known publicly available resource, which has been used by many research communities for mammographic image analysis. It is a database which consist 2620 scanned film mammography organized in the form of “case” and “volumes”. It contains benign, malignant and normal cases. A case can be described as a collection of images and information related to the mammography exam of a single patient. For the sake of ease of distribution, a volume is present which is simply a collection of cases collected together. All the cases are present with verified pathology information. It is a useful tool for decision support system development and testing because of its scale of database and ground truth validation. Additionally, the DDSM images are stored in non-standard compression files. While the DDSM dataset is made of mammography images with each image having 16-bit in .tif format and mass contour coordinates for each image. There are 51 malignants and 50 benigns. Each image has mass contours, which are in .ovl format. 101 mammograms are considered out of 2620, because in the DDSM dataset, there are only 101 mammograms with mass in dense tissue, while the remaining mammograms are in the panel tissue. The annotation for the dataset is available in the dataset performed manually under the expert team. Thus, by using MATLAB, the core mass regions available in the mammogram images are extracted, after which they are divided into malignant and benign by the help of annotated data.

Figure 3: An example of annotated mammogram image with contour mass region. Figure has been removed due to copyright issues.

Figure 3 shows annotated mammogram images obtained from DDSM database. The contours mark boundaries of the mass regions.

In order to create a uniform feature set, a fixed size rectangular block is extracted from the ROI for all the data set. The rectangular pixels block is made symmetrical about the centre of the ROI. DDSM has 101 mammograms with mass in dense tissue. In total all 101 mammogram images are considered out of which 382 benigns ROI's and 420 malignants ROI's are extracted automatecally. To standardize for both manignant and benign, is kept of fixed size which is 51x51 pixels. For developing and processing the algorithm, MATLAB 2019 is used.

5 Methodology

Figure 4, shows the proposed working algorithm of the thesis in simple steps. The figure explains various stages of data preprocessing and training of classification models and their relation with each other. The proposed methodology completes in four sequential steps which are as follows,

(i) ROI Extraction

(ii) Feature Extraction

(iii) Feature Reduction

(iv) Classification

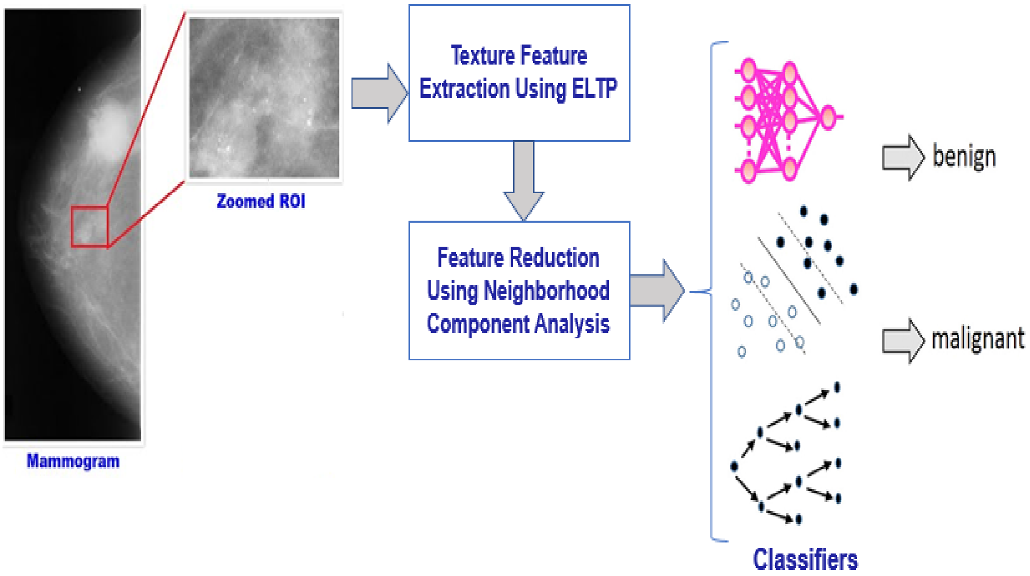

Mammogram
Zoomed ROI
Texture Feature Extraction Using ELTP
Feature Reduction Using Neighborhood Component Analysis
Classifiers
benign
malignant

Figure 4: Pictorial representation of the proposed algorithm stages .

5.1 ROI Extraction

In computer vision and optical character recognition, the term ROI defines the borders of an object under consideration [37]. The DDSM dataset consists of 101 mammograms with masses located in dense tissue.

Two methods are used to extract the ROI's which are explained below

5.1.1 First method

All 101 mammogram images are processed, out of which 1302 benign ROI's and 1632 malignant ROI's are extracted automatically. To standardize, both malignant and benign each ROI, fixed size 51x51 pixels is maintained. For developing and implementing the algorithm, MATLAB 2019 is used. In the proposed methodology, at first, the smallest rectangle containing mass contour is extracted using mass coordinates provided in DDSM as shown in Figure 5. After this step, random extraction of 51 × 51 size ROI's is performed within the rectangle. The mass contours provided by the radiologist have an irregular shape as shown Figure 5. A block is created which is determined by the minimum and maximum (x,y) coordinates of ROI data. In order to select pixel block for classification, the center of the ROI is determined then a rectangular region is extracted with a symmetrical to the rectangle. Further, these malignant and benign patches are used to extract the features and then perform classification for the detection of breast cancer.

Figure 5: Block representation for ROI extraction. Image has been removed due to copyright issues.

5.1.2 Second method

Mass contour is extracted, from the smallest rectangle containing using mass coordinates as shown in Figure 6. Automatically, 25 ROI's are extracted from inside the mass and 25 ROI's are extreacted from outside the mass from malignant and benign mammograms, respectively. Hence, for malignant 1170 ROIs, for benign 1142 ROIs from inside the mass and 2525 ROIs from outside the mass which represent the normal ROI's, are extracted.

Figure 6: Example of ROI patch extraction, masses present in yellow Contour (rectangular) are extracted for malignant cases and other are taken for benign cases [33]. Image has been removed due to copyright issues.

5.2 Features extraction

For feature extraction, the Extended Local Ternary Pattern (ELTP) method is used. A total of 289 distinct patterns were extracted. All the features extracted from ELTP do not contain discriminatory information. Hence, not all the feature combinations that are generated will be useful.

5.2.1 Working on ELTP

The proposed ELTP attempts to use a clustering method to group the patterns in a meaningful way, the process for converting a region into its ELTP [24] representation. Gray-levels in the range $-t^e$ to $+t^e$ around g_c^e are set to '0', while the values above the range is set to '1' and the values below are set to '-1', as described in equation (1),

$$s^e(g_p, g_c^e, t^c) = \begin{pmatrix} 1, & if & g_p - g_c^c \geq t^c \\ 0, & if & g_p - g_c^c < t^c \\ -1, & if & g_p - g_c^c \leq -t^c \end{pmatrix}, p = 0, 1, 2, ...P - 1 \tag{1}$$

where $g_c^e = mean(G)$, $t^c = mad(G)$, $G = \{g_i | i = 0, 1, 2, ...8\}$, p is the size neighbour set of pixel, $g_p(p = 0, 1, 2...p - 1)$ repesents the gray neighbour value, $mean(G)$ is the mean of the set G, $mad(G)$ is the median absolute deviation of the set G which is the set of the gray-level values in a 3×3 local region. In this method in place of user-defined threshold an auto adaptive threshold t^e which is median absolute deviation (MAD) is adapted, which makes ELTP code invariant to gray-level transformations which does not get affected by noise and indicates the derivation of the local region .

As shown in Figure 7, for ease each ternary pattern is subdivided into two sub parts, $ELTP\_P$ and $ELTP\_N$, and the combination of these to parts give the ELTP descriptor with final computing and histogram. ELTP is rotational invariant as shown in Figure 7, by using equation (1), the ELTP descriptor is defined by

$$ELTP_{P,R} = ELTP\_P_{P,R}{*}(P{+}2){-}(ELTP\_P_{P,R}{*}(ELTP\_P{+}1))/2ELTP\_N_{P,R} \tag{2}$$

$$ELTP\_P_{P,R} = \Sigma_{p=0}^{p-1} e(s^e(g_p, g_c^c, t^c), 1), \tag{3}$$

$$ELTP\_N_{P,R} = \Sigma_{p=0}^{p-1} e(s^e(g_p, g_c^c, t^c), -1), \tag{4}$$

$$e(x,y) = \begin{pmatrix} 1, & x = y \\ 0, & x \neq y \end{pmatrix} \tag{5}$$

Figure 6 shows the working algorithm, from left to right gray-levels are quantized to -1, 0, -1 ternary pattern and corresponding threshold values are indicated in the arrow. The quantized ternanry pattern is divided into two sub patterns such that in upper sub-pattern -1 is replaced with 0, and in the lower sub-pattern -1 is repalced with 1 and 1 replaced with 0. In lower sub pattern two replacements have taken place while in upper sub-pattern only one replacement. As indicated ELPT is rotational invariant (always get both '1s' next to each other no matter how we rotate as shown by arrow in sub division in Figure 3). The combination of these two sub patterns is 19 which is final ELTP as shown in Figure 3.

Figure 7: ELTP Encoding Scheme where patterns are extracted on a similar manner as explained in the above equation (1) [32]. Figure has been removed due to copyright issues.

This mechanism is applied for every ROI on the data set of 51×51 pixels. Data set is divided in 17 blocks each of size 3×3 and features are extracted from each block. A simple calculation is given in appendix J.

5.3 Feature reduction

Features extracted by the ELTP technique are reduced to 17 features, collected out of total number of 289 features from each ROI (17×17 blocks which is collected by dividing each 51×51 ROI into blocks of size 3×3 and each block contribute to 1 feature). This process is done by Neighbourhood Component Analysis (NSA), which finds a feature space such that a stochastic nearest neighbour algorithm gives only those patterns which show better performance. The advantage of using this algorithm is that the number of classes k can be determined as a function of A, which is the change in distance of the data points, up to a scalar constant. Its Leave One Out (LOO) classification can be used for keeping every individual pattern to be analysed individually.

5.3.1 Working principle of neighbourhood component analysis

Neighbourhood component analysis learns a distance metric by calculating a linear change in the data points in such a manner that LOO classification have its maximized performance in transformation space. Moreover, corresponding matrix, A, is calculated, which is the change found in distance of the data points. Further in a transformed space (A*) is determined to know the transformed space. A* is determined in such a manner that objective function is maximized,

$$A^* = argmaxf(A). \tag{6}$$

LOO is a classification technique in which a class label is forecasted using one feature point and associated distance metric. After a linear transformation computation vector of nearest-neighbours can also be used for classification purpose. For multiple data points to contribute in decision making an approach depending on stochastic gradient descent is utilized. The whole transformed dataset is treated as stochastic close neighbours rather than using k nearest neighbours. As shown in Figure 8, 9 and 10.

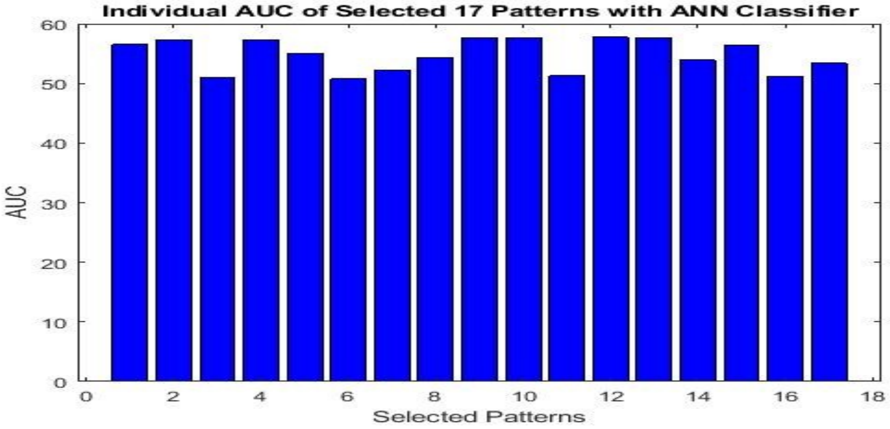


Figure 8: Individual performance for 17 patterns by using ANN classifier

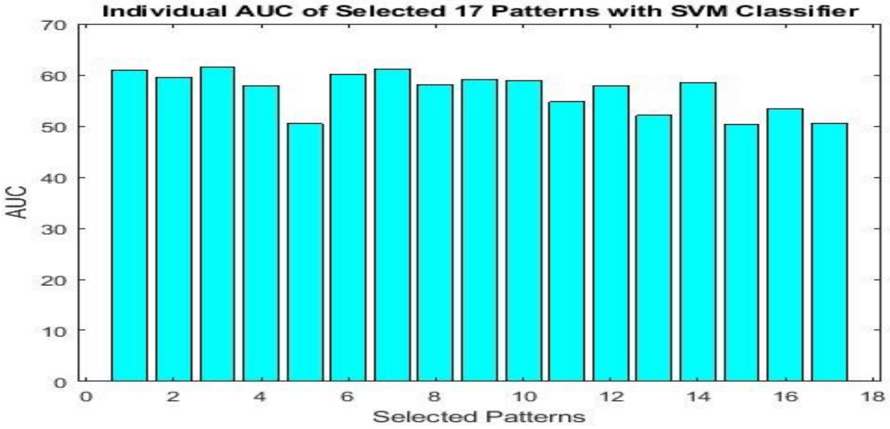


Figure 9: Individual performance for 17 patterns by using SMV classifier

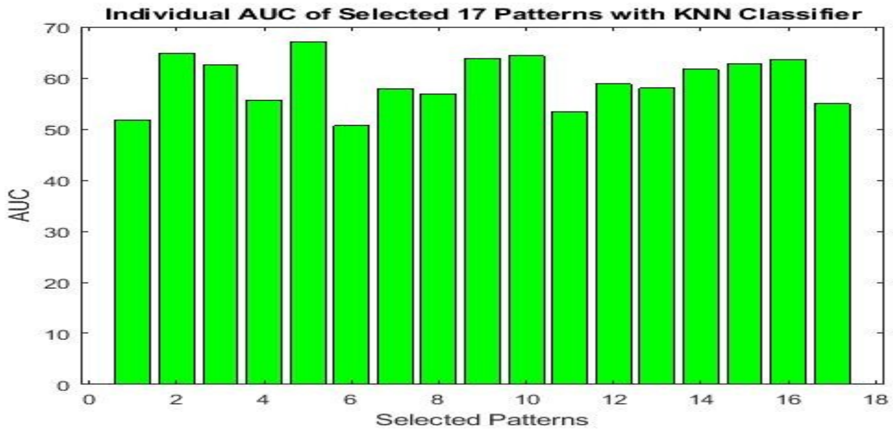


Figure 10: Individual performance for 17 patterns by using KNN classifier

By using fscnca (X,Y) feature selection, 17 such features have been concluded, which on evaluation (individually) perform better than 50% as per the threshold criteria set. This evaluation is performed to use only those features from the data points those have capability to differentiate malignant ROIs from benign ROIs. For better understanding used MATLAB code for NSA is given in appendix.

5.3.2 Feature generation

The 17 patterns that are used as the finalized feature vector as shown below in the Figure 11,

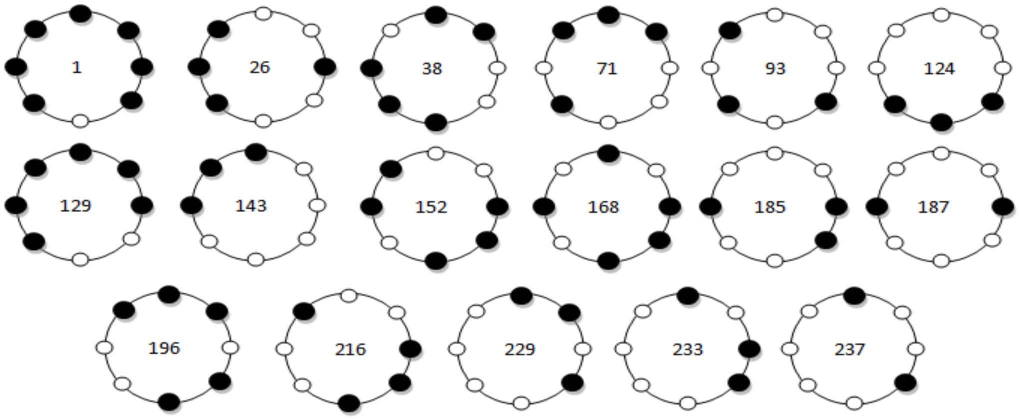


Figure 11: Final Patterns Used For Feature Vector after Feature Reduction

By using ELTP, 17 different patterns are extracted out of 289 features from each ROI where the feature vector extracted was as follows,

$$F_i = |ELTPattern_i|, \tag{7}$$

where i = 0,1,2,3,4……..,255. So total 255 patterns were extracted using ELTP. Then these ELT patterns are used to extract the most efficient patterns. The most efficient patterns were considered and other patterns were ignored for the training the machine learning algorithm. The new feature vector is,

$$F_p = PNCA(ELTPattern_i), \tag{8}$$

where PNCA is applied over ELT patterns and finally selected patterns are indicated with subscrited with p. In proposed system the selected patterns are
F_p = 1, 26, 38, 71, 93, 124, 129, 143, 152, 168, 185, 187, 196, 216, 229, 233, 237.
The feature vector comprising of the features used is represented by F_n
$F_n = f_1, f_{26}, f_{38}, f_{71}, f_{93}, f_{124}, f_{129}, f_{143}, f_{152}, f_{168}, f_{185}, f_{187}, f_{196}, f_{216}, f_{229}, f_{233}, f_{237}.$
As discussed in above section, only those features are taken into account whose individual AUC is greater than 50%. That is also depicted in the figures below

as the minimum AUC achieved by any selected pattern for any classifier is still greater than 50%.

To be a part of the feature vector, a threshold is set then if any pattern which is selected to be in the feature vector needs to achieve that threshold. In this case seventeen features have crossed the set threshold which means they are eligible to be part of the final feature vector which would be used to evaluate the input ROI.

5.4 Classification

For classification purpose, three different classifiers are used to have the comparative analysis about their performance. The classifiers are given 17 different features of each ROI. Those 17 features are the best available patterns that can be used classify the malignant and benign ROI patches. The three classifiers are discussed below in detail.

5.4.1 KNN classifier

KNN is a simple algorithm that is used for the storage and classification of different samples based on similarly measure such as similarity measures. Moreover, KNN is non-parametric technique and can be useful in statistical estimation and pattern recognition. Before using KNN, a few assumptions need to be made. The basic assumption, that KNN makes is that, the data is in the feature space. The data can be identified as scalars or identified as multidimensional vectors. KNN is used in classification where new unlabeled data in the form of data points or data objects for testing is given. Use of KNN classifier helps in better analyzation of mammograms and improve diagnosis of breast cancer. The KNN algorithm can help classify the mammograms into different classes.

Figure 12: KNN classifier working principal for different values of k [34]. Figure has been removed due to copyright issues.

Figure 12 shows the working principal of the KNN classifier, and shows how effectively it chooses the value of K for the classification of two different clusters. Consider the class of the yellow area is to be found. It can either be red or green. The K is the nearest neighbours from where a vote is taken. When K=3, chooses the closest three datapoints on the plane for the purpose of taking a vote.

5.4.2 SVM classifier

One of the most used classifiers in breast masses classification is SVM [39]. SVM [10] is a supervised machine learning algorithm which can be used for classification problems. It uses a linear classifier to classify data into two categories. A classification task usually involves training and testing data, which consists of some data instances which are feature points. The training set contains one "target value" (class labels) and several features. The accuracy of an SVM model is highly dependent on the selection of kernel parameters. SVM as a binary classifier is used in many of the research problems but its performance is highly dependent on the feature data points. Below is an example in Figure 12, about how exactly radial SVM binary classifier works. A separating hyperplane is made in accordance to the distance between the parameters. In proposed methodology, this hyperplane is radial. And data points available within the margin of separating hyperplane are known as Support Vector. So, distance between these data points plays a vital role for deciding the position of hyperplane which finally decides the classifier performance.

Figure 13: SVM Classifier working principal where nonlinear separation boundary is developed for the radial classification of two different CLASSES [35]. Figure has been removed due to copyright issues.

Figure 13 represents the SVM classifier. As shown in Figure 13, It creates a line or hyperplane that separates data into classes. SVM Classifier working principal where nonlinear separation boundary is developed for the radial classification of two different classes.

5.4.3 Artificial neural network

One of the most used classifiers recently is the Neural Network (NN) [36] classifier. The NN consists of units (neurons) arranged in layers. Each unit receives input and applies a (often nonlinear) function to it and then passes the output to the next layer. This function is also known as the activation function. The neurons in each layer are connected to the neurons in the next layer through weighted connections, where the weights are fine-tuned during the training process. The NN has found application in a wide variety of problems, especially in computer vision. ANN uses weights for different neurons which is summed and then passed from an activation function. Activation function here plays a vital role for deciding the pattern to be followed for classification purpose. This activation function gets learning while in training mode and it follows its predefined pattern. Figure 14 shows the architecture of ANN.

Figure 14: ANN Architecture where neurons are assigned with the weight and effective activation function is selected to have efficient classification [36]. Figure has been removed due to copyright issues.

As shown in Figure 14, ANN consists of artificial neurons with assigned weights. The middle layer takes the sum of weighted inputs and applies a logistic/non-linear function to the sum. The result of the function is the output of the middle layer neuron.

6 Results and discussion

6.1 ELTP results by classifying the masses to malignant and benign

By using the first method of ROI's extraction which is described in section 5.1.1. ELTP is used to extract the 256 different patterns from each individual ROI, out of those 256 patterns only 17 different patterns are selected. From 256 ELTP patterns only those patterns are selected which have shown AUC greater than 50%. Table 1 shows results of the proposed technique with 17 patterns. The table illustrates different parameter results. The results shown in Table 1 are of different classifiers, where reduced 17 features/patterns are considered. Neighbourhood component analysis are used for the training and testing purpose. As it can be seen in Table 1, that KNN have shown most promising results. The reason for this is the working principal of the classifier which is discussed in detail in section 5 - methodology. The optimal value of K is chosen by inspecting the data provided, which have played a vital role for the higher value of AUC in the proposed methodology. A larger value of K is more precise than the smaller ones, as it reduces the overall noise, though it is not guaranteed. In cross-validation approach a like manner decide a decent K value by utilizing a free dataset to endorse the K value. Truly, the ideal value of K for most datasets has been in the range of 3 and 10 which produces significantly better outcomes over ANN and SVM, depending on the case scenario. Reduced feature vector holds 17 different ELTP patterns which have shown quite impressive results. The credit for selection of these efficient patterns for classification goes to the NSA. As the working methodology enables the feature reduction technique to analyse each pattern individually. NSA is used because it works on the principal of favouritism with ELTP feature pattern and it can give us only those patterns which have shown better performance. Its Leave One Out (LOO) phenomenon can be used for allowing every individual pattern to be analysed individually. NCA is used to rank the feature set to determine redundant attributes. Half of the maximum ranked patterns are used as a threshold to remove fewer active attributes in the data set. Therefore, only those patterns are considered which have shown AUC more than 50%, individually. There are total 17 such features which have shown more than 50% AUC when tested on individual basis.

Figure 15 depicts a graph with the region of convergence by different classifiers. In this figure, KNN converges fastest among all three different classifiers. This clearly illustrates the high AUC for KNN classifier when compared with other classifiers.

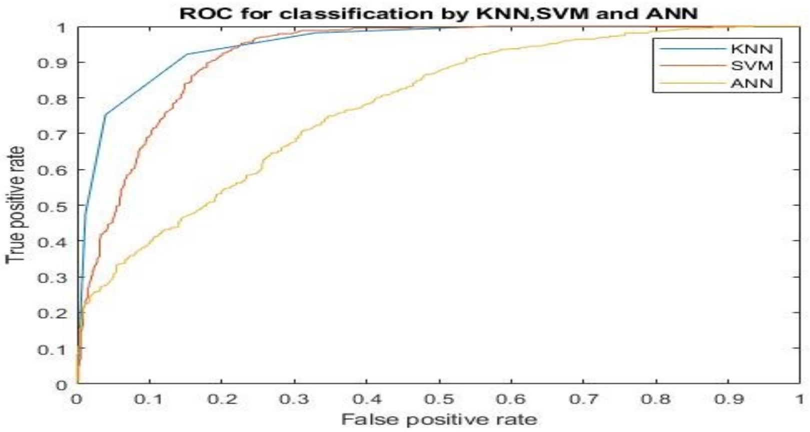


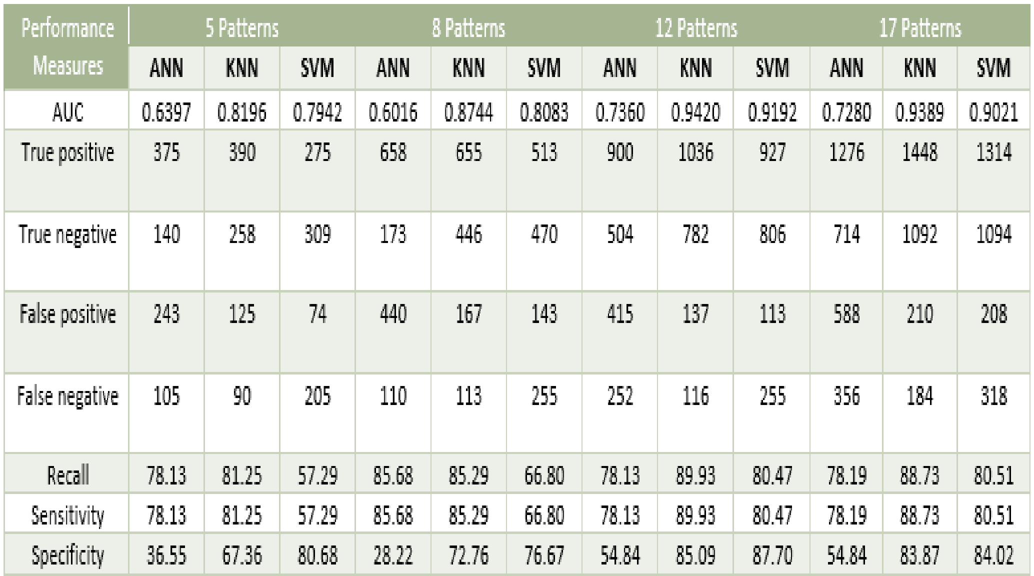
Figure 15: ROC for classification using 17 selected ELTP patterns.

Table 1 is showing results for different number of patterns. As it can be seen that increasing, the number of effective patterns increases the performance. In total we have 17 efficient patterns, analysis is done on 17 of those patterns which has best performance. When different number of patterns are considered starting from 5 patterns (minimum) to 17 patterns (maximum), everytime by introducing new effective patterns the positive change in performance is observed which is why all 17 patterns are recommended to be utilized.

First feature vectors, using 5 patterns are,
$F_1 = f_1, f_{26}, f_{38}, f_{71}, f_{93}.$
second feature vectors by using 8 patterns
$F_2 = f_1, f_{26}, f_{38}, f_{71}, f_{93}, f_{237}, f_{233}, f_{216},$
third feature vector by using 12 patterns
$F_3 = f_1, f_{26}, f_{38}, f_{71}, f_{93}, f_{237}, f_{233}, f_{216}, f_{196}, f_{229}, f_{129},$
fourth by using all the feature vector
F_4= $f_1; f_{26}; f_{38}; f_{71}; f_{93}; f_{124}; f_{129}; f_{143}; f_{152}; f_{168}; f_{185}; f_{187}; f_{196}; f_{216}; f_{229}; f_{233}; f_{237}$

These patterns are selected on the basis of threshold. For top 5 patterns, the threshold was kept high while for the 17 patterns, threshold was kept low. By varying the threshold the number of patterns are increased as per there individual performance. In Table 1, performance can be observed when threshold is reduced as some of the features lying below the threshold are playing some role in the classification AUC. These patterns are selected on the basis of their ranking they hold individually AUC. For 5 patterns top 5 patterns having the highest AUC among all are selected, then for top 8 patterns are selected followed by top 12, and then top 17 patterns. 17 patterns are the highest, being achieved as threshold for selecting the individual patterns, AUC is kept on 50% by going below this won't favour the all feature vector, and overfitting may be introduced. Whilst we need to take care, that overfitting problem should not be occurred. As the objective of using feature reduction is to overcome overfitting problem that comes because of large feature vector. 17 patterns size is not a big feature vector size, so this overfitting problem won't occur, which is clearly depicted in the Table 1 as well. 17 pattern using all together gives us the best result.

True Positives are the number of malignant cells classified correctly as malignant cells. True positives should be maximum to get the efficient model. As can be seen in Table 1, that increasing number of patterns till 50% AUC on individual basis increases the true positive ratio. The maximum is achieved by 17 patterns which is 1448 when using KNN classifier. This means 1448 malignant ROIs are correctly classified. On other hand, the False Positives are the wrongly classified benign ROIs. This is the scenario where benign ROIs are classified as malignant ROIs. The ratio of false positives with the total number ROI in the class decreases with the increase in patterns. In seventeen patterns there were 210 such benign ROIs which classified wrongly and give as malignant results. True Negatives are those numbers, which shows correctly classified benign ROIs which should be greater to get better specificity. The trend changes here and SVM with 17 patterns have shown the best result this indicates that SVM with 17 patterns will give the best specificity with 1094 true negative cases. False positives are those case which are classified as malignant ROIs but in original they belong to the class of benign. As shown in Table 1, SVM with 17 patterns have shown best performance with begin cells, here as well SVM with 17 patterns have achieved minimum ratio of 208 such benign cases, which are classified wrong. The False positives and false positives should be minimum as they represent the error caused by the classifier during classification.



Performance Measures	5 Patterns			8 Patterns			12 Patterns			17 Patterns		
	ANN	KNN	SVM	ANN	KNN	SVM	ANN	KNN	SVM	ANN	KNN	SVM
AUC	0.6397	0.8196	0.7942	0.6016	0.8744	0.8083	0.7360	0.9420	0.9192	0.7280	0.9389	0.9021
True positive	375	390	275	658	655	513	900	1036	927	1276	1448	1314
True negative	140	258	309	173	446	470	504	782	806	714	1092	1094
False positive	243	125	74	440	167	143	415	137	113	588	210	208
False negative	105	90	205	110	113	255	252	116	255	356	184	318
Recall	78.13	81.25	57.29	85.68	85.29	66.80	78.13	89.93	80.47	78.19	88.73	80.51
Sensitivity	78.13	81.25	57.29	85.68	85.29	66.80	78.13	89.93	80.47	78.19	88.73	80.51
Specificity	36.55	67.36	80.68	28.22	72.76	76.67	54.84	85.09	87.70	54.84	83.87	84.02

Table 1: Results for 17 patterns with different classifiers to classify the mass for malignant and benign.

The individual feature (Reduced Features) performance for the classification is shown in Figure 7, 8 and 9. In bar graphs for individual patterns it can be seen that each pattern have AUC greater than 50%. That's the working principal of feature reduction and threshold at 50% AUC is set to make a pattern eligible to be final feature vector. The case of KNN which clearly depicts the good performance of KNN when compared with other classifiers using the same 1 pattern. Pattern 5 which is '93' shows the best AUC of 67% when used with KNN classifier. The patterns which are not selected or shown are those patterns which were unable to achieve 50% when they were evaluated individually. Further patterns '26', '168' and '152' have also shown high performance of 65%, 64%, and 63% respectively. The performance is best in the case of KNN. It can be analysed how much each model is capable of distinguishing between classes. The higher AUC indicates the better model which is capable of differentiating between patients with and without disease.

These statistics show that the selected patterns are the highly efficient pattern when compared with the fellow patterns extracted using ELTP.

6.2 Comparative analysis among different cases using ELTP selected patterns

An efficient model is required to show the high achievement in classification of giving two distinct classes. In total three different classes, which are normal, benign and malignant have to be considered. Every combination should show high AUC to validate that the selected patterns are highly distinguished when they belong to different classes. By using second method of RoI's extraction, ELTP is applied to compare the following cases.

6.3 Malignant vs normal

The first case is malignant vs normal ROI classification. Figure 16 shows the evaluation of the performance of these two classes on the basis of three different classifiers. In this particular scenario SVM have shown better performance than the other classifiers. Thes second best is ROC curve for KNN, followed by ANN classifier, but the result shows that the selected patterns have high AUC when classify them, which shows the effectiveness of the selected patterns.

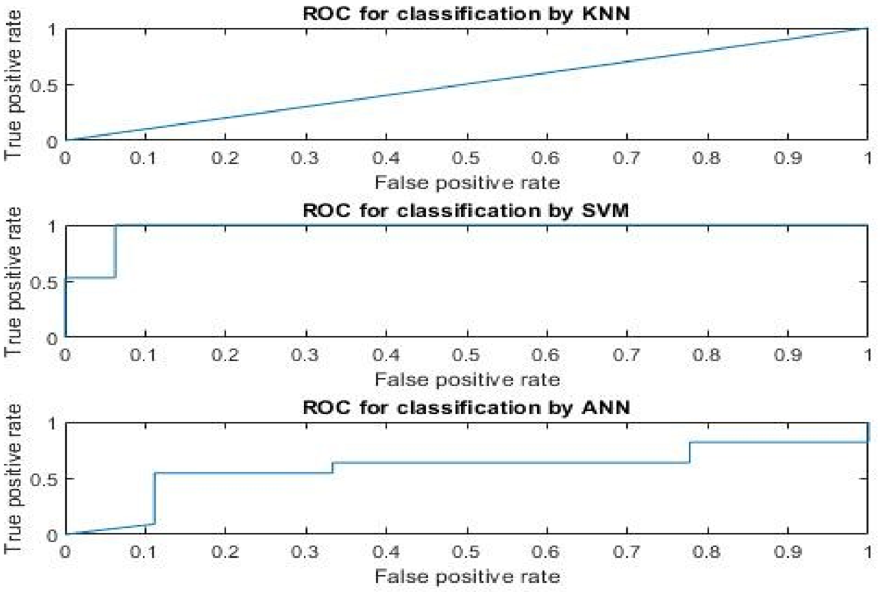


Figure 16: Performance malignant vs normal

6.4 Benign vs normal

Figure 17 shows the ROC curves of this case with usage of different classifiers. For this case best performance is shown by SVM classifier. This shows better results than the previous case, because all three classifiers have converged very rapidly in comparison to malignant vs normal. Hence this shows that the selected patterns are highly effective for the case and plays ideal role on classification of these ROIs.

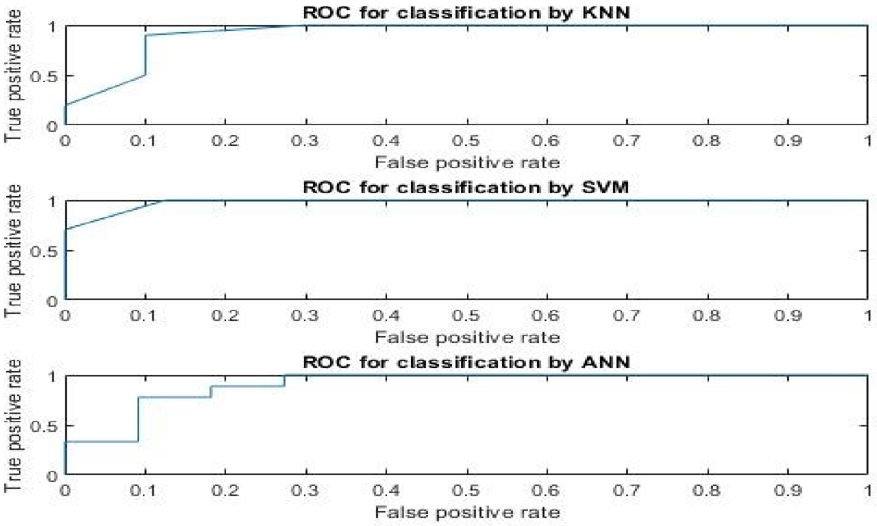


Figure 17: ROC performance for different classifiers, benign vs normal

6.5 Malignant vs benign

The last case is malignant and benign, which is considered to be the most difficult amongst these three cases. Figure 18 illustrates the performance of classification of these classes with these patterns. Here, in this case KNN has shown the best performance followed by SVM and ANN. But the overall ROC curves of all the classifiers clearly depicts good performance by the selected patterns.

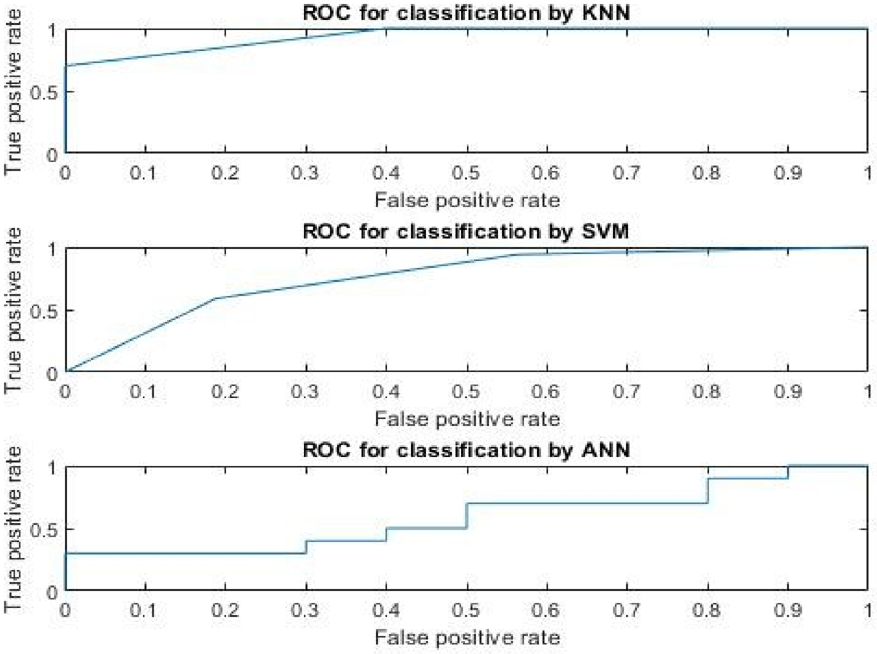


Figure 18: Performance Malignant VS Benign

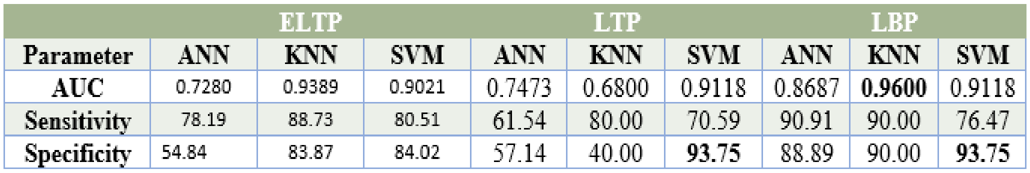
As it can be seen from above discussion that the selected patterns have really distinguished the three classes in a very effective manner. These patterns introduced high interclass variation in different classes of ROIs which clearly depicts that the selected patterns are highly efficient in carrying out the classification task of mammogram ROIs.

6.6 Comparison between ELTP, LTP and LBP by using 17 patterns on the same dataset

For the comparison purpose results are obtained from LBP and LTP pattern also, followed by same NCA, and all three classifiers. This result is obtained to see how the performance of ELTP pattern shows promising results. LBP and LTP is applied on same ROI dataset as for ELTP and the patterns are passed through NCA for selection of most efficient patterns. Same dataset was kept and same feature reduction technique was adopted. Moreover, number of selected patterns were kept constant. These patterns were fixed to 17 different patterns, so that a better comparative analysis can be carried out. The 17 patterns selected for the LTP and LBP were those patterns which gave best individual results when compared with their fellow patterns.

Table 2 shows that in different pattern extraction techniques, using different parameters with different classifiers have shown promising results. But the point here to be noted is that which parameter is the most important among them when it comes to the question of a life. So, the most important parameter here is sensitivity which basically classifies malignant ROI as malignant ROI which is most needed and crucial for any patient suffering from this deathly disease. Sensitivity tells about the actual proportion of positives that are correctly identified as positive by the classifier.

Figure 19-20 shows the ROC curves for LTP and LBP. As shown in above figures, the performance shown by LTP patterns is not as good as LBP patterns. Perticularly in KNN case, like we had in ELTP, that KNN performance outperforms the other classifiers. Therefore, KNN is a better classifier when it comes to use such patterns. Whilst for LTP and LBP patterns most efficient classifiers are SVM and KNN classifiers. The ROC patterns have shown that even LBP and LTP patterns have high converging rate but for some specific classifiers only. Like ROC curve, for LTP using ANN classifier do not converges, while for LBP the ROC does not converge rapidely for ANN and SVM. Whereas in case of ELTP for all classifier's convergence occurred earlier than that of these other patterns. As discussed, the most important thing here is to diagnose the malignant ROI which the ELTP have done best, which is 100%. The best AUC is achieved by the selected patterns of LBP in combination with KNN which is 0.9600. This shows in depth analysis, how different patterns can be used for achieving different targets. Moreover, the high convergence predicts about the efficient ELTP selected patterns performance.



	ELTP			LTP			LBP		
Parameter	**ANN**	**KNN**	**SVM**	**ANN**	**KNN**	**SVM**	**ANN**	**KNN**	**SVM**
AUC	0.7280	0.9389	0.9021	0.7473	0.6800	0.9118	0.8687	**0.9600**	0.9118
Sensitivity	78.19	88.73	80.51	61.54	80.00	70.59	90.91	90.00	76.47
Specificity	54.84	83.87	84.02	57.14	40.00	**93.75**	88.89	90.00	**93.75**

Table 2: Comparison results of ELTP, LBP and LTP

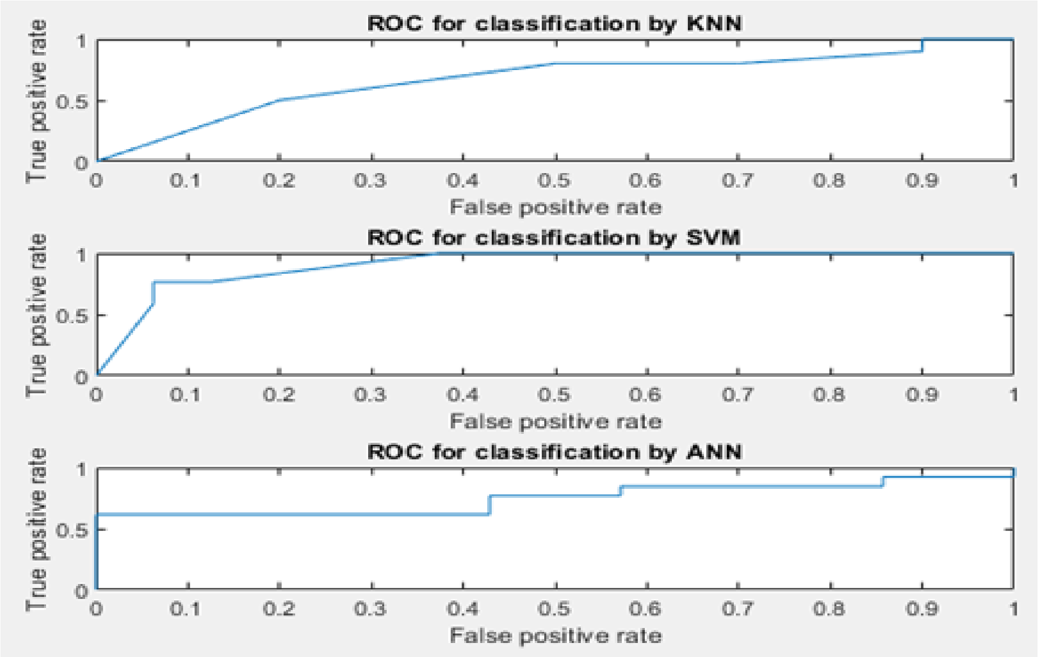


Figure 19: ROC curves for LTP extracted efficient patterns

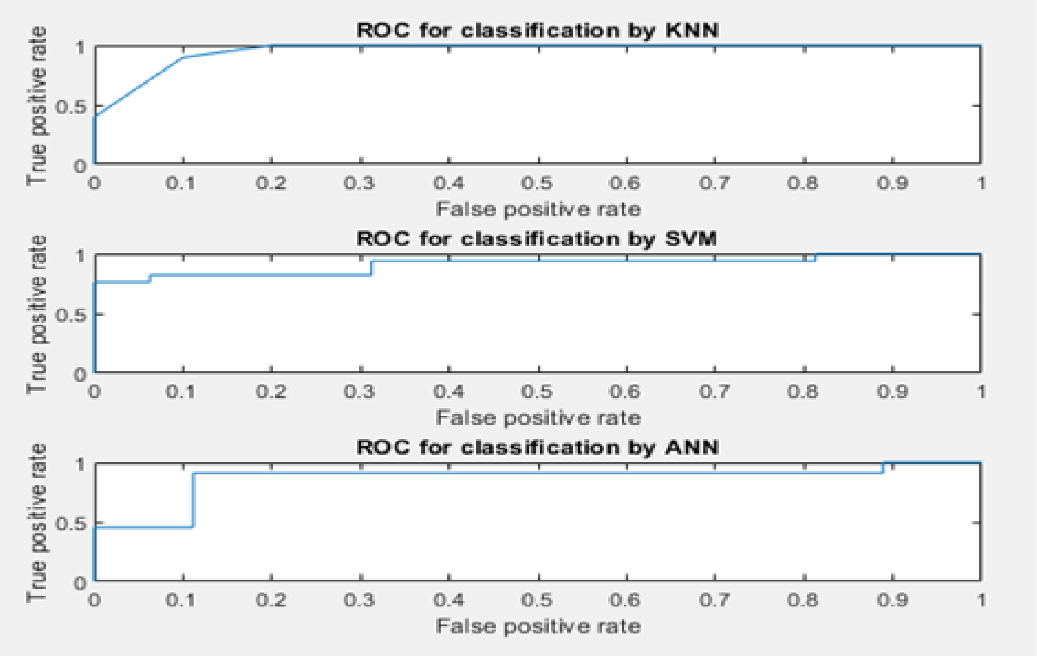
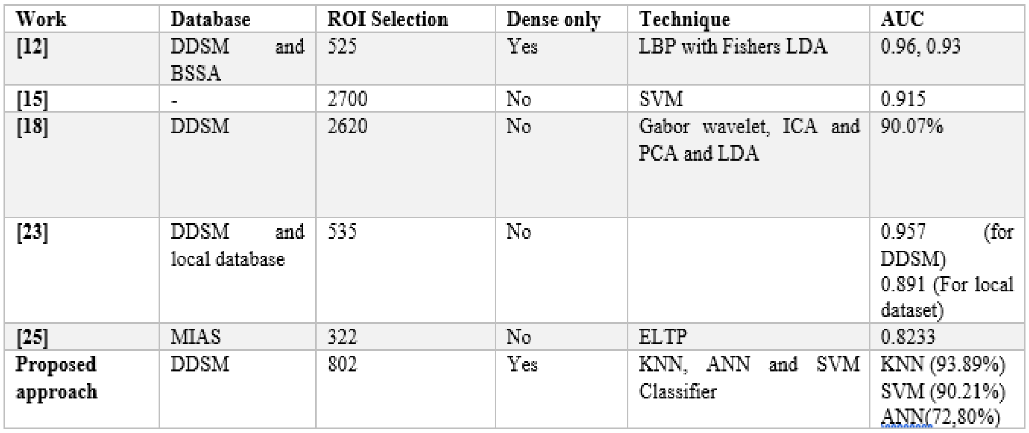


Figure 20: ROC curves for LBP extracted efficient patterns

6.7 Comparison between ELTP and other methods in literature review

It is very important to compare the proposed technique with other existing techniques in literature. Comparison is carried out with already present techniques, it is observed that there are two mostly used databases, which are DDSM and MIAS databases. Further, researchers have used different techniques in which they have used different number of ROIs and different classifiers. The analysis become more complicated when only dense ROIs are considered that decreases the interclass variation, which make it difficult for classifier to classify two classes. As can be seen in Table 3 that which technique and what credentials have used to achieve the AUC. The best AUC is achieved with the technique proposed in [12] for the DDSM database. This is followed by techniques proposed in [23], and then by [18], followed by the proposed technique but here it is important to keep in notice that the proposed technique have shown a greater sensitivity, which is not observed by any other technique as discussed in literature review section. When it comes to sensitivity, which is the most sensitive parameter for medical image processing, the proposed algorithm has shown most promising results. As sensitivity is associated with the detection of malignant cell, the proposed algorithm holds no gap in detecting the malignant region of the breast. That suggest it is quite impressive when it comes to detection of malignant cells.



Work	Database	ROI Selection	Dense only	Technique	AUC
[12]	DDSM and BSSA	525	Yes	LBP with Fishers LDA	0.96, 0.93
[15]	-	2700	No	SVM	0.915
[18]	DDSM	2620	No	Gabor wavelet, ICA and PCA and LDA	90.07%
[23]	DDSM and local database	535	No		0.957 (for DDSM) 0.891 (For local dataset)
[25]	MIAS	322	No	ELTP	0.8233
Proposed approach	DDSM	802	Yes	KNN, ANN and SVM Classifier	KNN (93.89%) SVM (90.21%) ANN(72,80%)

Table 3: Performance comparision of the proposed approach with other mass classification techniques found in the literature.

6.8 Discussion

The ELTP patterns are extracted from the ROIs after which seventeen unique patterns are selected, which are identified based on their individual performance (that is how NCA works). The reduced feature vector holds 17 different ELTP patterns which have shown quite impressive results. As the working methodology enables the feature reduction technique to analyse each pattern individually. NSA is used because it selects only those ELTP feature pattern which can give us only those patterns which have better performance, greater than the set threshold. Its 'Leave One Out' phenomenon can be used for keeping every individual pattern to be analysed individually. Three classifiers are used for this model identification which are ANN, SVM and KNN. Detailed comparative analysis is carried out in result section to know how efficient ELTP reduced feature vector or efficient patterns have performed. Different sizes of reduced feature vectors arenalysed and as a final decision all reduced efficient patterns are included. As discussed in the result section, that every efficient pattern plays its own vital role in the performance. For the comparison purpose results are also carried out on LBP and LTP pattern followed by same NCA and all three classifiers. This result is carried out to see how the performance of ELTP pattern shows promising results. This comparison showed encouraging results for ELTP patterns. In this study, efforts are made to propose a model that can be used for effective classification of malignant and benign region of interests, so that efficient and early diagnosis of breast cancer can be made possible, where the proposed system gives 88% of sensitivity which is ideal. Many techniques have shown better AUC than that of proposed technique but point to be considered here is that the proposed technique have shown high sensitivity which suggests that it is better for the classification of malignant ROIs, which is a major contribution in research of this classification problem.

7 Conclusion

In this thesis, a CAD is proposed for mammogram image classification. ELTP is used for the proposed CAD model. Furthermore, NSA is used to reduce the feature vector extracted from ELTP. 17 unique patterns with are identified based on their individual performance. These 17 patterns are used to train the classifiers. Three classifiers are used for this model which are ANN, SVM and KNN. Moreover, for comparison purpose results are also carried out using LBP and LTP patterns. Among all carried out results, the most convincing key point is the high performance of KNN classifier with 17 efficient patterns for classification of malignant ROIs. It has achieved 100% of sensitivity which means it didn't miss even a single malignant ROI to detect as malignant. The results have shown detailed comparative analysis of different feature vectors as well as of different classifiers. This thesis provides a unique model with the highest possible malignant ROI classification, accurately. A depth analysis carried out in result section, it can be stated that KNN classifier performs better than the other classifiers for the classification of ROIs on basis of patterns which LBP patterns have also depicted in their ROC curves, discussed in result section.

By using the ELTP method 93% AUC, 88% sensitivity and 83% of specificity is achived which is convincing when compared to other available methods. Although LBP has achieved better results but it has a drawback that is, it is not noise resistance which has been removed by using ELPT.

8 Appendix

A. ELTP Code for extraction

```
function ELTP= ELTP2(img,pattern)
%we get the dimensions of the image block
[r,c]=size(img);
%loop the image block and ignoring the borders...
for row = 2 : r - 1
for col = 2 : c - 1
%we consider each image pixel as center and get ELTP for the it so we get the neighboorhood pixels
pixels = double(img(row-1:row+1,col-1:col+1));
% we need to find the threshold at and gc
%t=mad(G), G is the list of pixels in the image block
%gc=mean(G) , G is the list of pixels in the image block
gc=mean(pixels(:));
t=mad(pixels(:));
%loop through the pixels block to quantitized the pixel from 0-255
%to -1,0,+1
se= zeros(3, 3);
for i=1:3
for j=1:3
gp=pixels(i,j);
if abs(gp-gc) > = t
se(i,j)=1;
elseif abs(gp-gc) <t
se(i,j)=0;
else
se(i,j)=-1;
```

```
end
end
end
    ne(1)=se(1,1); ne(2)=se(1,2); ne(3)=se(1,3); ne(4)=se(2,3); ne(5)=se(3,3); ne(6)=se(3,2);
ne(7)=se(3,1); ne(8)=se(3,1);
    se=ne;
    ELTP_P=se; ELTP_P(se==-1)=0;
    ELTP_N=se; ELTP_N(se==-1)=1; ELTP_N(se==1)=0;
    ELTP_N = ELTP_N(pattern); ELTP_P = ELTP_P(pattern); P=8; ELTP_N=sum(ELTP_N==1);
ELTP_P=sum(ELTP_P==1);
    ELTP(row,col)=ELTP_P *(P+2)-(ELTP_P* (ELTP_P+1))/2 + ELTP_N;
    end end
```

B. LBP Code for extraction

```
function lbps= LBP(img,pattern)
    %dimensions of img [r,c]=size(img);
    lbps=[]; %// For each pixel in our image, ignoring the borders... for row = 2 :
r - 1 for col = 2 : c - 1
    %//neighboorhood pixels = double(img(row-1:row+1,col-1:col+1));
    t=pixels(2,2); % Get ranges and determine LTP se= zeros(3, 3);
for i=1:3
for j=1:3
gp=pixels(i,j);
if gp >= t
se(i,j)=1;
else
se(i,j)=0;
end
```

```
end
end
lbp(1)=se(1,1); lbp(2)=se(1,2); lbp(3)=se(1,3);
lbp(4)=se(2,3); lbp(5)=se(3,3); lbp(6)=se(3,2);
lbp(7)=se(3,1); lbp(8)=se(2,1);
lbp=lbp(pattern);
lbps(row,col)=bi2de(lbp);
end
end
end
```

C. Code for training SVM model

```
close all
clear all
clc
rng(0)
%//////////////////////////////////////////////////////////// %Cross Validation Evalkuation load('featureset.mat')
load('target.mat')
k=5; %Here put the number for k-folds
%[featureset,idx]=featselect(featureset,target,.5);
featureset=pca(featureset','NumComponents',25);
indices = crossvalind('Kfold',target,k);
test = (indices == k);
tran1 =  test;
tt = featureset(test,:);
tr = featureset( tran1,:);
tts= target(test);
```

```
trs= target(tran1);
Mdl = fitcsvm(tr,trs,’Standardize’,true,’KernelFunction’,’RBF’,... ’KernelScale’,’auto’);
compactSvm=compact(Mdl);
svmProb=fitPosterior(compactSvm,tr,trs);
[result,scores] = predict(svmProb,tt);
C=confusionmat(result,tts)
accuracy=sum(diag(C))/sum(C(:))
[truepositive,truenegative,falsepositive,falsenegative,precision,recall,fscore,accuracy,MAE,sensitivity,spe
[X,Y,T,AUC]=perfcurve(tts,scores(:,2),1);
area_under_curve=AUC
figure(1)
plot(X,Y)
hold on
xlabel(’False positive rate’)
ylabel(’True positive rate’)
title(’ROC for classification by SVM’)
```

D. Code for training Neural Network Model

```
clc
clear all2
close all
load(’featureset.mat’)
load(’target.mat’)
rng(0)
%[featureset,idx]=featselect(featureset,target,.5);
featureset=pca(featureset’,’NumComponents’,25);
input=featureset;
```

```
k=5;
indices = crossvalind('Kfold',target,k);
test = (indices == k);
tran1 =  test;
tt = featureset(test,:);
tr = featureset( tran1,:);
tts= target(test);
trs= target(tran1);
hiddenLayerSize = 15;
trainFcn = 'trainscg'; %Training function for classification
net = patternnet(hiddenLayerSize, trainFcn);
net.divideParam.trainRatio = 100/100;
net.trainParam.max_fail=1000;
net.performFcn = 'crossentropy';
net.plotFcns = 'plotperform','plottrainstate','ploterrhist';
[net, ] = train(net,tr',trs); % train the network
view(net) % view neural network
y = net(tt');
prediction=y;
prediction(y>.5)=1;
prediction(y<.5)=0;
testactual=tts;
[truepositive,truenegative,falsepositive,falsenegative,precision,recall,fscore,accuracy,MAE,sensitivity,spe
[X,Y,T,AUC]=perfcurve(testactual,y,1);
area_under_curve=AUC
figure(1)
plot(X,Y)
hold on
xlabel('False positive rate')
```

```
ylabel('True positive rate')
title('ROC for classification by ANN')
```

E. Code for training KNN Model

```
close all
clear allff
clc
rng(0)
%///////////////////////////////////////////////////////////////////
%Cross Validation Evaluation
load('featureset.mat')
load('target.mat')
k=5; %Here put the number for k-folds
%[featureset,idx]=featselect(featureset,target,.5);
featureset=pca(featureset','NumComponents',25);
indices = crossvalind('Kfold',target,k);
test = (indices == k);
tran1 =  test;
tt = featureset(test,:);
tr = featureset( tran1,:);
tts= target(test);
trs= target(tran1);
Mdl = fitcknn(tr,trs,'NumNeighbors',5,'Standardize',1);
[result,scores] = predict(Mdl,tt);
C=confusionmat(result,tts)
accuracy=sum(diag(C))/sum(C(:))
[truepositive,truenegative,falsepositive,falsenegative,precision,recall,fscore,accuracy,MAE,sensitivity,spe
```

```
[X,Y,T,AUC]=perfcurve(tts,scores(:,2),1);
area_under_curve=AUC
figure(1)
plot(X,Y)
hold on
xlabel(’False positive rate’)
ylabel(’True positive rate’)
title(’ROC for classification by KNN’)
```

F. Code for Cross-Validation Evaluation

```
close all
clear all
clc
rng(0)
%////////////////////////////////////////////////////////////
%Cross Validation Evaluation
load(’featureset.mat’)
load(’target.mat’)
k=5; %Here put the number for k-folds
featureset=pca(featureset’,’NumComponents’,25);
indices = crossvalind(’Kfold’,target,k);
test = (indices == k);
tran1 =  test;
tt = featureset(test,:);
tr = featureset( tran1,:);
tts= target(test);
trs= target(tran1);
```

```
Mdl = fitcknn(tr,trs,'NumNeighbors',5,'Standardize',1);
[result,scores] = predict(Mdl,tt);
C=confusionmat(result,tts)
accuracy=sum(diag(C))/sum(C(:))
[truepositive,truenegative,falsepositive,falsenegative,precision,recall,fscore,accuracy,MAE,sensitivity,spe
[X_knn,Y_knn,T,AUC]=perfcurve(tts,scores(:,2),1);
area_under_curve=AUC
%///////////////////////////////////////////////////////////////////
%Cross Validation Evaluation
load('featureset.mat')
load('target.mat')
k=5; %Here put the number for k-folds
featureset=pca(featureset','NumComponents',25);
indices = crossvalind('Kfold',target,k);
test = (indices == k);
tran1 =  test;
tt = featureset(test,:);
tr = featureset( tran1,:);
tts= target(test);
trs= target(tran1);
Mdl = fitcsvm(tr,trs,'Standardize',true,'KernelFunction','rbf',...
'KernelScale','auto');
compactSvm=compact(Mdl);
svmProb=fitPosterior(compactSvm,tr,trs); [result,scores] = predict(svmProb,tt);
C=confusionmat(result,tts)
accuracy=sum(diag(C))/sum(C(:))
[truepositive,truenegative,falsepositive,falsenegative,precision,recall,fscore,accuracy,MAE,sensitivity,spe
[X_svm,Y_svm,T,AUC]=perfcurve(tts,scores(:,2),1);
area_under_curve=AUC
```

```
load(’featureset.mat’)
load(’target.mat’)
featureset=pca(featureset’,’NumComponents’,25);
input=featureset;
hiddenLayerSize = 10;
trainFcn = ’trainscg’; %Training function for classification net = patternnet(hiddenLayerSize,
trainFcn);
net.divideParam.trainRatio = 80/100;
net.divideParam.valRatio = 10/100;
net.divideParam.testRatio = 20/100;
net.trainParam.max_fail=1000;
net.performFcn = ’crossentropy’;
net.plotFcns = ’plotperform’,’plottrainstate’,’ploterrhist’;
[net,tr] = train(net,input’,target); % train the network view(net)
% view neural network testinput=featureset(tr.testInd,:);
y = net(testinput’);
prediction=y;
prediction(y>.5)=1;
prediction(y<.5)=0;
testactual=target(tr.testInd);
[truepositive,truenegative,falsepositive,falsenegative,precision,recall,fscore,accuracy,MAE,sensitivity,spe
[X_ann,Y_ann,T,AUC]=perfcurve(testactual,y,1);
area_under_curve=AUC
figure(1)
plot(X_knn,Y_knn)
hold on
plot(X_svm,Y_svm)
hold on
plot(X_ann,Y_ann)
```

```
xlabel(’False positive rate’)
ylabel(’True positive rate’)
title(’ROC for classification by KNN,SVM and ANN’)
legend(’KNN’,’SVM’,’ANN’)
```

G. Code for ROI extraction

```
clc
clear all
%specify the malignant image and rois path
maglinantimgpath = ’C:;́
maglinantroipath = ’C:;́
%specify the benign image and rois path
benignimgpath = ’C:;́
%we get the tif and ovl files in the malignant image and rois path respectively
maglinantimgdir=dir(fullfile(maglinantimgpath,’*.tif’)); maglinantroidir=dir(fullfile(maglinantroipath,’*.
%we get the tif and ovl files in the benign image and rois path respectively
benignimgdir=dir(fullfile(benignimgpath,’*.tif’));
benignroidir=dir(fullfile(benignroipath,’*.ovl’));
%determine the number of files in the maglinant and begnin directory lengthmelignantdirFiles=length(maglinantimgdir);
lengthbegnindirFiles=length(benignimgdir);
%specify the the size of roi to be extracted
Size=51;
%change directory to created malignant folder to save all malignant rois
%extracted
cd maglinant/
%loop through the code for i=1:lengthmelignantdirFiles-1
```

```
%image file and roi file in the loop
imagefile=strcat(maglinantimgpath,’/’,maglinantimgdir(i).name);
roifile=strcat(maglinantroipath,’/’,maglinantroidir(i).name);
%read the image file and convert it to uint8 (8bits)
img=im2uint8(imread(imagefile));
%import roi associated with the image file from the ovl file
A=importdata(roifile);
%extarct roi portions from the image rioimg=img(A(:,2),A(:,1));
%set a start point
startp=1;
%get the base name of the file
base=maglinantimgdir(i).name;
%get the total number,ed of the pixels size that can be extracted
ed=floor(size(A,1)/Size);
%loop through ed . crop the pixel block of the size and save to the cropped pixel
block in the folder created
for j=1:ed
I=rioimg(startp:startp+Size-1,startp:startp+Size-1);
name=strcat(base(1:end-4),num2str(j),base(end-3:end));
imwrite(I,name);
startp=startp+51;
end
end
%we repeat the same thing for begnign
cd ../
cd benign/
for i=1:lengthbegnindirFiles
imagefile=strcat(benignimgpath,’/’,benignimgdir(i).name);
roifile=strcat(benignroipath,’/’,benignroidir(i).name);
```

```
img=im2uint8(imread(imagefile));
A=importdata(roifile);
rioimg=img(A(:,2),A(:,1));
startp=1;
base=benignimgdir(i).name;
ed=floor(size(A,1)/Size);
for j=1:ed
I=rioimg(startp:startp+Size-1,startp:startp+Size-1);
name=strcat(base(1:end-4),num2str(j),base(end-3:end));
imwrite(I,name);
startp=startp+51;
end
end
cd ../
```

H. Code for selecting features

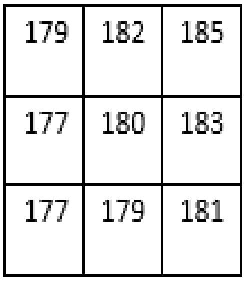
```
 function [input,idx]=featselect(input,target,perc)
rng(0)
mdln=fscnca(input,target);
f=mdln.FeatureWeights;
m=max(f(find( isnan(f) & f =Inf)));
idx=find( isnan(f) & f =Inf & f¿=perc*m);
input=input(:,idx);
end
```

Code for true positive and false positive calculation function

```
[truepositive,truenegative,falsepositive,falsenegative,precision,recall,fscore,accuracy,MAE,sensitivity,spe
truepositive=0;
truenegative=0;
if numel(predict) =numel(actual)
disp(’predict and actual must have equal items’)
else
falsenegative=0;
falsepositive=0;
for i =1:numel(predict)
if predict(i)==1 && actual(i)==1
truepositive=truepositive+1;
elseif predict(i)==0 && actual(i)==0
truenegative=truenegative+1;
elseif predict(i)==1 && actual(i)==0
falsepositive=falsepositive+1;
elseif predict(i)==0 && actual(i)==1
falsenegative=falsenegative+1;
end
end
error=actual-predict;
MAE=mae(error);
RMSE=sqrt(immse(actual,predict));
precision=100*truepositive/(truepositive+falsepositive);
recall=100*truepositive/(truepositive+falsenegative);
fscore=2/((1/precision) +(1/recall));
C=confusionmat(predict,actual);
accuracy=100*sum(diag(C))/sum(C(:)); sensitivity=100*(truepositive)/(truepositive+falsenegative);
specificity=100*(truenegative)/(truenegative + falsepositive);
```

```
fprintf('truepositive= %d ',truepositive)
fprintf('truenegative= %d',truenegative)
fprintf('falsepositive= %d',falsepositive)
fprintf('falsenegative= %d',falsenegative)
fprintf('accuracy= %3.2f',accuracy)
fprintf('precision= %3.2f',precision)
fprintf('recall= %3.2f',recall)
fprintf('fscore= %3.2f',fscore)
fprintf('sensitivity= %3.2f',sensitivity)
fprintf('specificity= %3.2f',specificity)
fprintf('Mean absolute error= %3.2f',MAE)
fprintf('Root Mean square error= %3.2f',RMSE)
end
end
```

J. Calculation for ELTP



179	182	185
177	180	183
177	179	181

Step1: Find the Mean

=179+182+185+183+181+179+177+177

=1443/8

=180.4

Step2: find median

1- Ascending order

177 177 179 179 180 183 183 185

2- Median= 179+180/2

= 180

Step 3: Find the MAD (median absolute deviation)

1- By subtract each pixel from median

3 3 1 1 1 2 3 5

2- Ascending order

1 1 1 2 3 3 3 5

3- MAD= 2+3/2

= 2.5

Step 4: subtracting each pixel from Mean

(179-180.4)= -1.4

(177-180.4)= -3.4

(177-180.4)= -3.4

(179-180.4)= -1.4

(181-180.4)= 0.6

(183-180.4)= 2.6

(185-180.4)= 4.6

(182-180.4)= 1.6

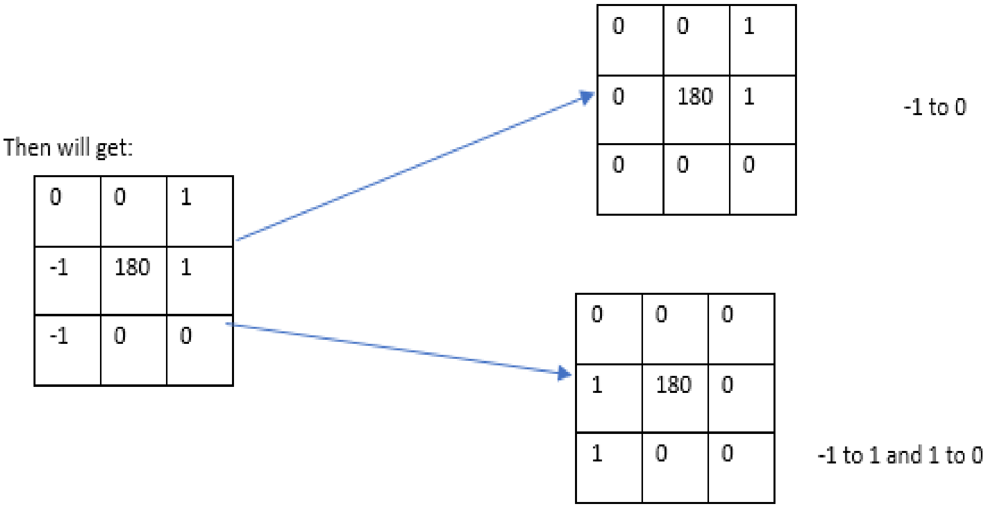
Comparing this result with 2.5 which is the MAD

$-1 < \text{MAD} < 1$ otherwise = 0

Then,

179	→	0
177	→	-1
177	→	-1
179	→	0
181	→	0
183	→	1
185	→	1
182	→	0

Then will get:



0	0	1
-1	180	1
-1	0	0

0	0	1
0	180	1
0	0	0

-1 to 0

0	0	0
1	180	0
1	0	0

-1 to 1 and 1 to 0

According to ELTP algorithm, -1 should to be converted two times the first time to zero.

Then to 1 and convert the 1 to zero.

By applying ELTP algorithms:

=2*(8+2) – (2*(2+1))/2 +2

=19

9 References

[1] Breast Cancer. ‘U.S. Breast Cancer Statistics’ (2019). http://www.breastcancer.org/symptoms/understand bc/statistics. Accessed 31 October 2019.

[2] American Cancer Society, Cancer Facts and Figures 2019. Available at https://www.cancer.org/research/cancer-facts-statistics/all-cancer-factsfigures/cancer-facts-figures-2019.html, accessed October, 2017.

[3] Kelsey, Jennifer L., Marilie D. Gammon, and Esther M. John. ”Reproductive factors and breast cancer.” Epidemiologic reviews 15.1 (1993): 36.

[4] Gotzsche, Peter C., and Karsten Juhl Jørgensen. ”Screening for breast cancer with mammography.” Cochrane database of systematic reviews 6 (2013).

[5] Cheng, H.D., Shi, X.J., Min, R., et al.: ‘Approaches for automated detection and classification of masses in mammograms’, Pattern Recognit., 2006, 39, pp. 646–668.

[6] Chokri, F., Farida, M.H.: ‘Mammographic mass classification according to Bi-RADS lexicon’, IET Comput. Vis., 2017, 11, pp. 189–198(9).

[7] Sajeev, S., Bajger, M., Lee, G.: ‘Segmentation of breast masses in local dense background using adaptive clip limit-CLAHE’, International Conference on Digital Image Computing: Techniques and Applications (DICTA), Adelaide, Australia, 2015.

[8] Abdel-Nasser, M., Rashwan, H.A., Puig, D., et al.: ‘Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern’, Expert Syst. Appl., 2015, 42, (24), pp. 9499–9511.

[9] Kooi, T., Litjens, G., van Ginneken, B., et al.: ‘Large scale deep learning for computer aided detection of mammographic lesions’, Med. Image Anal., 2017, 35, pp. 303–312.

[10] de Oliveira, Fernando Soares Sérvulo, et al. ”Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes

and SVM." Computers in biology and medicine 57 (2015): 42-53.

[11] Khan, Salabat, et al. "Optimized Gabor features for mass classification in mammography." Applied Soft Computing 44 (2016): 267-280.

[12] Sajeev, Shelda, Mariusz Bajger, and Gobert Lee. "Superpixel texture analysis for classification of breast masses in dense background." IET Computer Vision 12.6 (2018): 779-786.

[13] Charan, Saira, Muhammad Jaleed Khan, and Khurram Khurshid. "Breast cancer detection in mammograms using convolutional neural network." 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). IEEE, 2018.

[14] Ahmed, Hafeez, and Abdul Haseeb. "LMS Based Adaptive Algorithm for Breast Cancer Detection using Mammogram Images." American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) 43.1 (2018): 169-177.

[15] Nguyen, M. P., et al. "An alternative approach to reduce massive false positives in mammograms using block variance of local coefficients features and support vector machine." Procedia computer science 20 (2013): 399-405.

[16] R. Nithya, B. Santhi, Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer, Int. J. Comput. Appl. 28 (6) (2011) 21–25 (published by Foundation of Computer Science, New York, USA).

[17] F. Moayedi, Z. Azimifar, R. Boostani, S. Katebi, Contourlet-based mammography mass classification using the SVM family, Comput. Biol. Med. 40 (4) (2010) 373–383.

[18] D. Costa, L. Campos, A. Barros, Classification of breast tissue in mammograms using efficient coding, BioMed. Eng. OnLine 10 (1) (2011) 55.

[19] P. GöRgel, A. Sertbas, O.N. Ucan, Mammographical mass detection and classification using local seed region growing-spherical wavelet transform (LSRG-SWT) hybrid scheme, Comput. Biol. Med. 43 (6) (2013) 765–774.

[20] Sajeev, S., Bajger, M., Lee, G.N.: ‘Improving breast mass segmentation in local dense background: an entropy based optimization of statistical region merging method’. Breast Imaging – 13th Int. Workshop (IWDM), Malmo, Sweden, 2016, pp. 635–642.

[21] Sajeev, S., Bajger, M., Lee, G.: ‘Segmentation of breast masses in local dense background using adaptive clip limit-CLAHE’, International Conference on Digital Image Computing: Techniques and Applications (DICTA), Adelaide, Australia, 2015.

[22] Choi, J.Y., Ro, Y.M.: ‘Multiresolution local binary pattern texture analysis combined with variable selection for application to false-positive reduction in computer-aided detection of breast masses on mammograms’, Phys. Med. Biol., 2012, 57, pp. 7029– 7052.

[23] Sajeev, S., Bajger, M., & Lee, G. (2017, November). Structured micro-pattern based LBP features for classification of masses in dense breasts. In 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.

[24] Yuan, Jing-Hua, Hao-Dong Zhu, Yong Gan, and Li Shang. ”Enhanced Local Ternary Pattern for Texture Classification.” In International Conference on Intelligent Computing, pp. 443-448. Springer, Cham, 2014.

[25] Rampun, Andrik, et al. ”Breast density classification using local ternary patterns in mammograms.” International Conference Image Analysis and Recognition. Springer, Cham, 2017.

[26] Heath, M., Bowyer, K., Kopans, D., et al.: ‘The digital database for screening mammography’. Proc. Fifth Int. Workshop on Digital Mammography, Toronto, Canada, 2001, pp. 212–218.

[27] Liao, W. H. (2010, August). Region description using extended local ternary patterns. In 2010 20th International Conference on Pattern Recognition (pp. 1003-1006). IEEE.

[28] Rampun, Andrik, et al. ”Breast density classification using local ternary

patterns in mammograms." Int. Conference Image Analysis and Recognition. Springer, Cham, 2017.

[29] Berbar, Mohamed A., Yaser A. Reyad, and Mohamed Hussain. "Breast mass classification using statistical and local binary pattern features." 2012 16th International Conference on Information Visualisation. IEEE, 2012.

[30] Liu, Jun, et al. "Improved local binary patterns for classification of masses using mammography." 2011 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2011.

[31] Liao, W. H. (2010, August). Region description using extended local ternary patterns. In 2010 20th International Conference on Pattern Recognition (pp. 1003-1006). IEEE.

[32] Yuan, J. H., Zhu, H. D., Gan, Y., & Shang, L. (2014, August). Enhanced Local Ternary Pattern for Texture Classification. In International Conference on Intelligent Computing (pp. 443-448). Springer, Cham.

[33] Hussain, M. (2014). False-positive reduction in mammography using multiscale spatial Weber law descriptor and support vector machines. Neural computing and Applications, 25(1), 83-93.

[34] https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

[35] https://www.dtreg.com/solution/view/20

[36] https://azati.ai/disease-prediction-and-classification-with-neural-networks/

[37] Szeliski, Richard. Computer vision: algorithms and applications. Springer Science & Business Media, 2010.

[38] Davies, E. R. (2017). Computer vision: principles, algorithms, applications, learning. Academic Press.

[39] Singh, Birmohan & Jain, V. & Singh, Sukhwinder. (2014). Mammogram Mass Classification Using Support Vector Machine with Texture, Shape Features and Hierarchical Centroid Method. Journal of Medical Imaging and Health Informatics. 4. 10.1166/jmihi.2014.1312.

[40] Rassem, Taha & Khoo, Bee Ee. (2014). Completed Local Ternary Pattern for Rotation Invariant Texture Classification. The Scientific World Journal. Volume 2014 (2014). 10. 10.1155/2014/373254.

[41] http://www.eng.usf.edu/cvprg/Mammography/Database.html