Automatic Detection and Classification of Cervical Cancer in Pap Smear Images using ETCM & CFE methods based Texture features and Various Classification Techniques

S.Athinarayanan[1], M.V.Srinath[2]

Research Scholar, Manonmaniam Sundaranar University, Abhisekapatti, Tirunelveli.

Director, Department of MCA, STET Women's College, Mannargudi, Trichy.

Abstract

Abstract: *Classification of the cervical cell is one of the most important and crucial tasks in the medical image analysis. Due to its importance, the aim of the paper is to investigate about the classification of Cervical Cell as Normal Cell or Abnormal Cell by using individual feature extraction method and combining individual feature extraction features method with the classification technique. In this paper four Feature Extraction methods were used: from that four, two were existing individual feature extraction methods namely Gray Level Co-Occurrence Matrix (GLCM) & Texton Co-Occurrence Matrix (TCM) and the remained two were proposed novel methods. From that proposed two, one was individual feature extraction method, that is Enriched Texton Co-Occurrence Matrix (ETCM) and other was combining individual feature extraction features method, that is Concatenated Feature Extraction (CFE). The CFE method represents all the individual feature extraction methods of GLCM, TCM & ETCM features are combining together to one feature to assess their joint performance. Then, these four feature extraction methods are tested over three classifiers such as Support Vector Machine (SVM), Radial Basis Function (RBF) and Feed Forward Neural Network (FFNN). This Examination was conducted over a set of single cervical cell based pap smear images. The dataset contains two classes of images, with a total of 952 images. The distribution of number of images per class is not uniform. Then, the performance of the proposed system was evaluated in terms of the statistical parameters of sensitivity, specificity & accuracy in both the individual feature extraction method with the classification techniques and combining individual feature extraction methods with the classification techniques. Hence, the performance of individual combination method described, the proposed ETCM features with SVM Classifier combination had given the better results than the other combinations such as ETCM with RBF Classifier, ETCM with FFNN Classifier, GLCM with SVM Classifier, GLCM with RBF Classifier, GLCM with FFNN Classifier, TCM with SVM Classifier, TCM with RBF Classifier & TCM with FFNN Classifier. Then the performance of the combining individual feature extraction features method described, proposed Concatenated Feature Extraction (CFE) method with SVM Classifier had given the better results than all other remained CFE method with classifier combinations and all other individual feature extraction and classification combinations.*

Keywords: Cervical Cancer, Feature Extraction, Classification.

1.Introduction

Cancer is a class of diseases characterized by out-of-control cell growth [1]. Cervical Cancer is a second highest cause of cancer death among women in the world [2]. The automatic diagnosis of cancer using the various computational technique is the most important and major utilization of medical image processing and computer-aided diagnosis system. Nowadays, cervical cancer is a very common concern of the women of the developing countries. The cure of this cancer is related to early detection and proper treatment according to the malignancy level. An automated method to analyze the uterine cervical images and to extract the diagnostic features of the images can be helpful to detect the cancer [3].

Cervical cancer has become one of the major causes of death among women worldwide. It can be cured when it is detected and treated in its earlier stage. But for most of the cases it throws symptoms only in the advanced stages. The traditional visual procedures are time consuming and error prone. Further it is impossible for a handful pair of eyes to sit and screen each and every woman on the planet. To solve this problem, we need some automated process that could accelerate the process and also produce accurate results. The automated system would consist of four phases [4] namely pre-processing phase for noise removal, segmentation phase to identify the cells and to separate nucleus from its cytoplasm, feature extraction phase to identify and extract the feature of normal cell & abnormal cell and the final classification phase classifies the given cervical cell as normal or abnormal cell based on the extracted features.

Analysing individual cervical cells in microscopic cytology images, obtained from the Pap smear test, is an important task for early diagnosis of cervical cancer [5]. The most effective way of detecting the cervical cancer is Pap smear screening test which was introduced by Dr. Georges Papanicolaou in 1940s. This Pap test is a manual screening procedure which is used to detect pre-cancerous changes in cervical cells on the basis of color and shape properties of cell nuclei and cytoplasm regions. There are two primary approaches to cervical cell segmentation. Traditional methods segment nuclei from single cells and collection of cells. For instance, Wu et al. [6] detect the boundary of nuclei by solving an optimal thresholding problem. Morphological analysis is also used to detect nucleus from cervical cell images [7].

2.Related Work

Classification of medical images using textural classification have been successfully performed in various medical images such as breast cancer, liver cancer, lung disease & cervical cancer [8]. Initially, texture of the cell was analysis based on the first and second order statistics of textures. The co-occurrence matrix features were first proposed by Haralick [9]. Weszka [10] compared texture feature extraction schemes based on the Fourier power spectrum, second order gray level statistics, the co-occurrence statistics and gray level run length statistics. The co-occurrence features were found to be the best of these features.

To extract the features of the single cervical cell pap smear image using Grey Level Co-occurrence Matrix and to classify that image by using Support Vector Machine approach [11]. SVM Classifier classify the texture features extracted into a set of supervised learning methods which analyze data and recognize patterns, used for statistical classification and regression analysis. Since an SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. Intuitively, an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

One of the most important changes in a cell when it becomes precancerous is a change in chromatin texture. The field of nuclear texture analysis gives information about the spatial arrangement of pixel gray levels in a digitized microscopic nuclei image. A well-known method for quantifying textures in digital images is the gray level co-occurrence matrix(GLCM). This method tries to quantify specific pairwise gray level occurrence at specific relative positions [12].

Cervical Cell features are extracted by first order statistical features and second order gray level co-occurrence matrix (GLCM). Then, the SVM classifier will predict the treatment outcome of the patient is cancer affect or not. Haralicks's gray level co-occurrence matrices (GLCM's) have been used very successfully in texture classification [13].

Cheb et al. [14] have developed an algorithm for segmenting nucleus and cytoplasm components. This system classifies the pap smear cells into anyone of four different types of classes using SVM. Two experiments were conducted to validate the classification performance which showed the best performance outputs. The output of SVM is found to be best for the most of the classes and better results for the remaining classes [15].

3.Proposed Methodology

The Framework of the Proposed Methodology is shown in given figure.1.

figure.1. Proposed Methodology
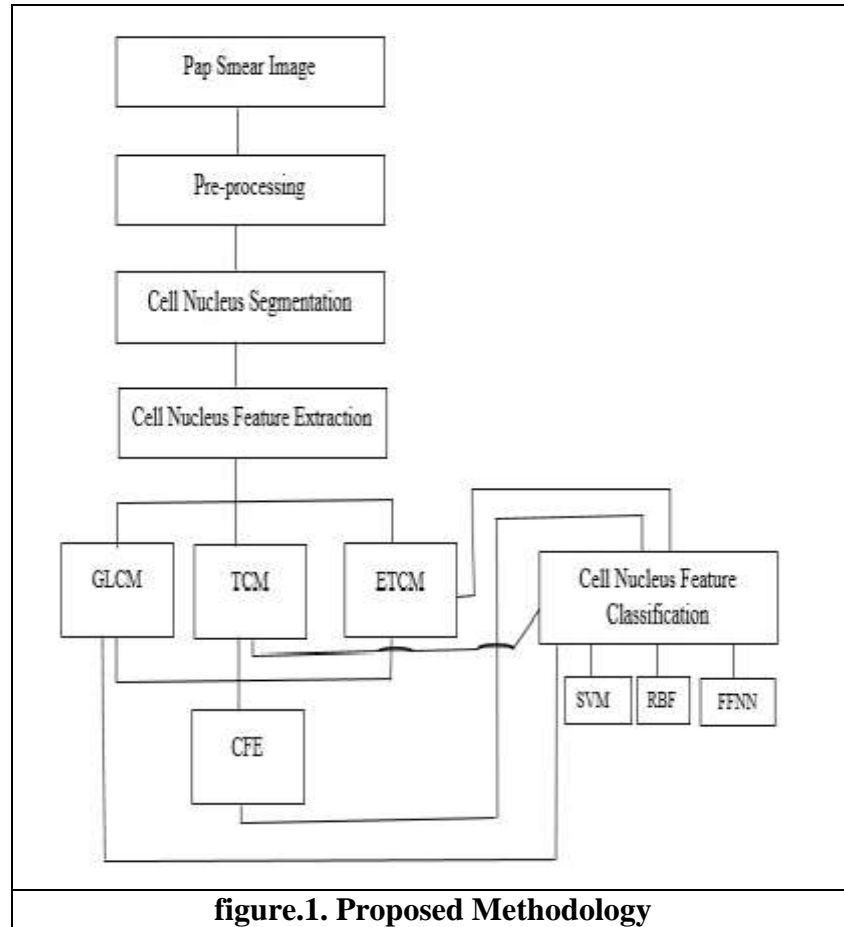
The Proposed Methodology Consists of four number of process.

That is

1. Pre-processing.

2. Nucleus Segmentation.

3. Nucleus Feature Extraction.

4. Nucleus Feature Classification.

The brief details of the above process is given.

3.1. Pre-processing:

It is the crucial stage in the entire process. Pre-processing on the input image is extremely essential, so that the image gets altered to be related to the further processing. In this paper, first of all the experimental images are color images and it converted into grayscale images for further processing. Then these images cannot be given directly applying the segmentation process. Since, these images are passed through an anisotropic filter which diminishes the noise and enhances the image quality. Anisotropic filter is used for reducing image noise without removing significant parts of the image content, particularly the edges, lines or other details that are important for the interpretation of the image [16].

3.2. Nucleus Segmentation:

The cervical cell nucleus segmentation process consists of two stages such as

i)   Image binarization using thresholding.

ii)   Sharpening the region using morphological operations.

3.2.1. Original image is convert into a binary image by thresholding

Initially, the input image is transformed into a binary image. An image of up to 256 gray levels is translated to a black and white image using the threshold value. The gray level value of every pixel in the improved image is considered at this stage. All the pixels with values above the threshold are set as white and the remaining pixels are set as black in the image during the binarization process. In this paper, the threshold value is selected based on the contrast of the image as given in eqn (1).

$$\text{Binarized Image}, B_{Binary}(k,y) = \begin{cases} 0, & if \ B_{grey}(k,y) = < Threshold \\ 1, & Otherwise \end{cases} \quad \dots (1)$$

3.2.2. Sharpening the nucleus region using Morphological operation

After transforming into binary images, the morphological process is applied for sharpening the regions and filling the gaps. The main processes of the morphological operations are opening, closing, erosion and dilation. In this paper, erosion operation is applied for removing the hurdle, noise and enhances the image. *Erosion:* In the erosion operation on an image F having labels 0 and 1 with structuring element Y, the value of pixel i in F is changed from 1 to 0, if the result of convolving Y with F, centered at i, is below some predefined value. We have set this value to be the area of Y, which is principally the number of pixels that are 1 in the structuring element itself. The structuring element, also known as the erosion kernel, finds out the details of how particular erosion thins boundaries as given in eqn (2).

.

$$IE = imerode(F,Y) \quad \dots\dots (2)$$

Then the final sharpened nucleus mask of the binary image is mapped with the original image means to get the segmented nucleus from the original pap smear image.

3.3. Nucleus feature Extraction:

The process of extracting the features of the high contrast image sequence in a temporal frame with gray scale reference information for text block detection in both horizontal and vertical edge scanning of adjacent text block in a multi-resolution fashion are considered as feature extraction. It extracts information grounded on maximum gradient difference. The purpose of feature extraction is to reduce the original data set by measuring certain properties, or features, that distinguish one input pattern from another pattern. The extracted feature is expected to provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into a feature space. In this paper four Feature Extraction methods were used: from that four, two were existing individual feature extraction methods namely Gray Level Co-Occurrence Matrix (GLCM) & Texton Co-Occurrence Matrix (TCM) and the remained two were proposed novel methods. From that proposed two, one was individual feature extraction method, that is Enriched Texton Co-Occurrence Matrix (ETCM) and other was combining individual feature extraction features method, that is Concatenated Feature Extraction (CFE). The CFE method represents all the individual feature extraction methods of GLCM, TCM & ETCM features are combining together to one feature to assess their joint performance. The types of individual and combining individual feature extraction features method is given below.

<u>**Individual Feature Extraction Methods:**</u>

* Computation of Feature Vector F(V1) using GLCM.

* Computation of Feature Vector F(V2) using TCM.

* Computation of Feature Vector F(V3) using ETCM.

Combining Individual Feature Extraction Features Method:

- Computation of Feature Vector F(V4) using CFE.

The detailed Process of the above feature extraction method is given below.

Individual Feature Extraction Methods:

3.3.1. Computation of Feature Vector F(V1) using GLCM:

Histogram based features are local in nature. These features do not consider spatial evidence into deliberation. Consequently for this persistence gray-level spatial co-occurrence matrix hd(i,j) based features are well-defined which are recognized as second order histogram based features. These features are based on the joint probability dispersal of duos of pixels. Distance d and angle θ within a given neighbourhood are used for calculation of joint likelihood spreading between pixels. Generally, d=1,2 and θ=0°,45°, 90°,135° are used for calculation. Texture features can be labelled using this co-occurrence matrix [17,18]. The Illustration of GLCM is shown in given figure.2. The GLCM F(V1) utilizes contrast, homogeneity, correlation and energy to describe cervical cell features
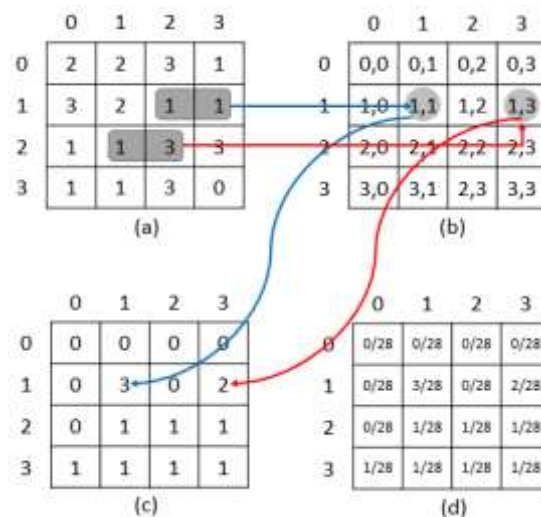
Figure.2: GLCM constructions. (a) An original image, (b) GLCM indices, (c) Co-occurrence matrix, (d) Probability value

3.3.2. Computation of Feature Vector F(V2) using TCM:

The term "Texton" is conceptually proposed by Julesz [19]. It is a very useful concept in texture analysis and has been utilized to developed efficient models in the context of texture recognition or object recognition [20,21]. TCM can represent the spatial correlation of textons, and it can discriminate color, texture and shape features simultaneously [22].

Texton Detection

Five special types of texton templates are used to detect the textons, which are shown in figure.3. The flow chart of texton detection is illustrated in figure.3a. In an image, we move the 2 x 2 grid from left-to-right and top-to-bottom throughout the image to detect textons with one pixel as the step-length. If the pixel values that fall in the texton template are the same, those pixels will form a texton, and their values are kept as the original values. Otherwise they will be set to zero. Each texton template can lead to a texton image (an example of texton detection result is shown in figure.3a), and the five texton templates will lead to five texton images. We combine them into a final texton image, as shown in figure.3b. The TCM F(V2) utilizes energy, contrast, entropy and homogeneity to describe cervical cell features.
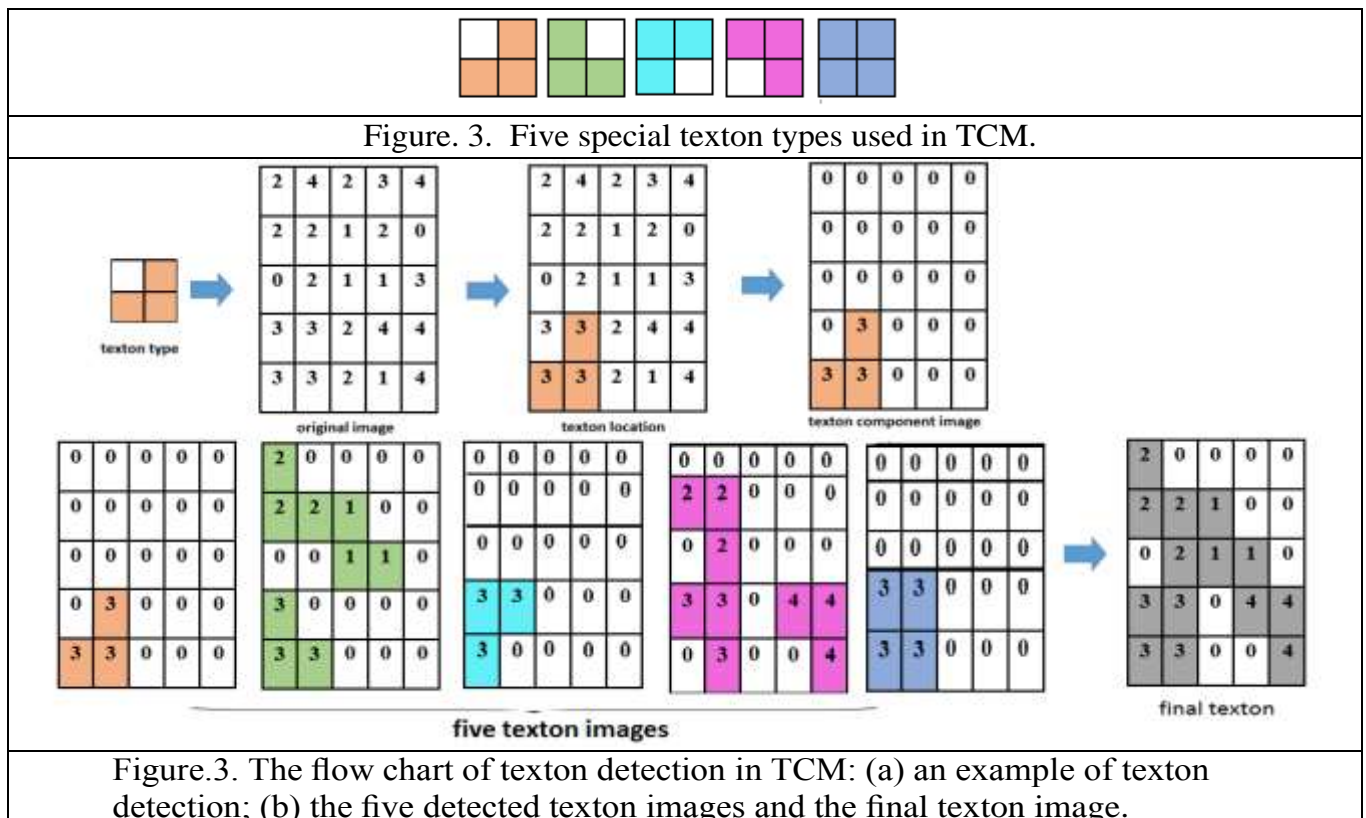
3.3.3. Computation of Feature Vector F(V3) using ETCM:

Texton is one of the very important concepts for texture analysis, it was developed 20 years ago [19]. It is a set pattern sharing a common property all over the image. According to the neuropsychological findings, different types of incentive are processed disjoint, yet concurrently, by dissimilar neural mechanism previously to the stimulus are intentionally perceived as a whole. In this proposed method, feature extraction process is done with the help of Enriched Texton Co-occurrence Matrix (ETCM). In this method, both Histogram and co-occurrence matrix are used for feature extraction process. The relationship between the values of neighbouring pixels is characterized by TCM. Histogram based techniques are simple to compute, but highest indexing performance. The co-occurrence matrix directly uses a feature representation of the image. If the dimension of the image is high, then the performance is decreased. The spatial information is lost when the histogram is used only for feature representation of the image [23]. Hence combine both histogram and co-occurrence matrix for feature extraction and representation.

Texton Detection

The Texton template defined in ETCM are different from those in TCM (refer to figure.3). In this paper, four special texton types are defined on a 2 x 2 grid, as shown in figure.4. Denote the four pixels as V1, V2, V3 and V4. If the two pixels highlighted in the gray color have the same value, the grid will form a texton. Those 4 texton types are denoted as T1, T2, T3 & T4, respectively.

The Working Mechanism of Texton detection is illustrated in figure.5. In the final segmented nucleus mapped with the original image, we move 2 x 2 block from left to right and top to bottom throughout the image to detect textons with 1 pixel as step length. If a texton is detected, the detected pixel pair values in the 2 x 2 grids are kept unchanged. Otherwise it will have the zero value. Finally we will obtain a Texton image, denoted by T(x,y). This process results was shown in figure.5(a-e). After the formation of final texton image, the feature vector F(V3) (Five features such as, ASM, entropy, IDM, contrast and Maximum probability) is extracted from it.

Figure. 3. Five special texton types used in TCM.

Figure.3. The flow chart of texton detection in TCM: (a) an example of texton detection; (b) the five detected texton images and the final texton image.
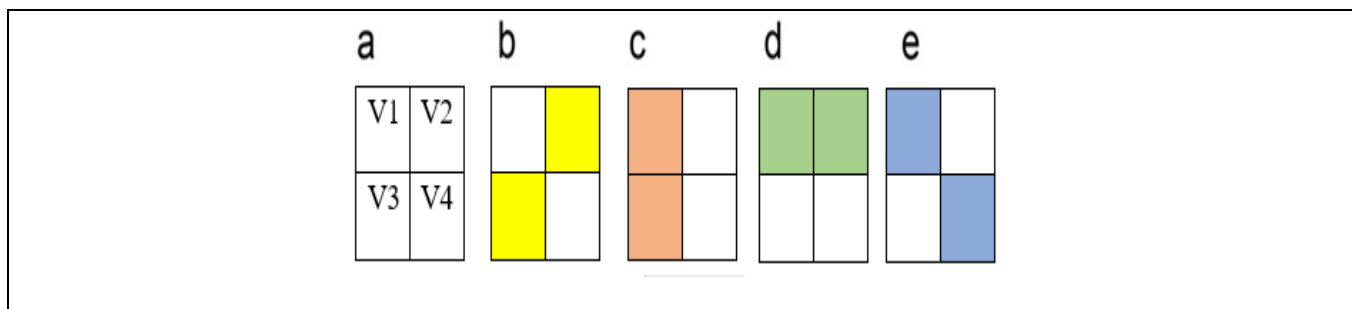
Figure.4. Special texton types of EETCM (a) original 2x2 matrix (b)Texton T1 (c) Texton T2 (d) Texton T3 (e) Texton T4
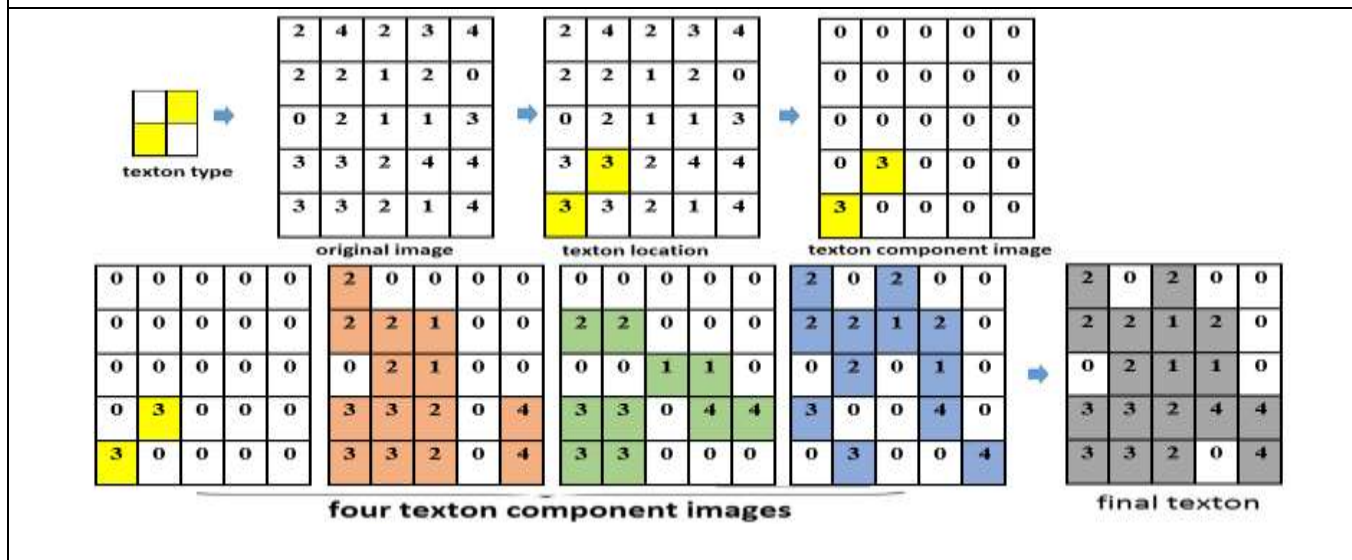

Figure.5: Texton image formation process using ETCM (a) 2x2 Matrix (b) Original image intensity value (c) Texton location of the original image (d) Four texton types (e) Final texton image of ETCM.

The four texton types used in ETCM contain rich information than those in TCM because the co-occurrence probability of two same-valued pixels is bigger than that of three or four same-valued pixels in a 2 x 2 grid. As for the texton detection procedure, ETCM is also faster than TCM. In the texton detection of TCM, the 2 x 2 grid moves throughout the image with one pixel as the step-length, and the detected textons based on the 5 textons and the same Texton detection in ETCM, the 2 x 2 grid moves throughout the image with one pixel as the step-length, and the detected textons based on 4 textons. So, the each pixels of an image 5 textons were applied in TCM but the ETCM the 4 textons only applied for texton detection process. So, the ETCM will reduce the computational complexity than TCM. And the detected texton in TCM & ETCM methods, neighborhood may overlap. So, The final texton image needs to be fused by the overlapped components of textons.

Difference between Proposed ETCM with TCM & GLCM in Texton Detection and Computational Complexity:

The four texton types used in ETCM contain rich information than those in TCM because the co-occurrence probability of two same-valued pixels is bigger than that of three or four same-valued pixels in a 2 x 2 grid. This Result was shown in the figure.6.As for the texton detection procedure, ETCM is also faster than TCM.
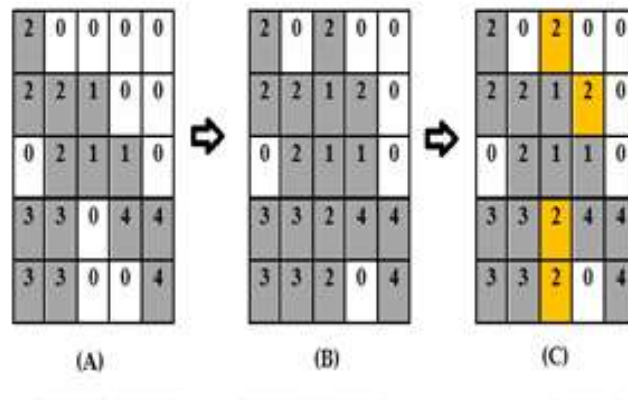
*Figure.6. The Result of Texton detection A) TCM B) ETCM C) Difference
between TCM & ETCM.*

.So, the ETCM will reduce the computational complexity than TCM. Then the ETCM method compared with the GLCM method, it's computational complexity is very low. Because the ETCM method, 5 texture features only were extracted, But for the GLCM method four texture features were extracted in two distances with four directions (i.e. 4 x 2 x 4 = 32 features). So, the extraction of more features in GLCM, could be take more computational time than ETCM.

Combining Individual Feature Extraction Features Method:
3.3.4. Computation of Feature Vector F(V4) using CFE:
The name CFE Stands for Concatenated Feature Extraction. The CFE method represents all the individual feature extraction methods of GLCM, TCM & ETCM features are combining together to one feature to assess their joint performance. Hence, total Feature vector uses F(V4) = F(V1) + F(V2) + F(V3) dimensional vector as the concluding image features in cervical cell classification.

Table.1. Summary of Texture Features
The four feature extraction methods and its features details are given in the following table.1.

Feature category	Features	Number of features
GLCM	Following GLCM features distance at 1 & 2 then angle value(Θ) at 0°, 45°, 90°, and 135° are calculated: (1) contrast, (2) homogeneity, (3) correlation, (4) Energy	2*4*4 =32 features
TCM	energy, contrast, entropy and homogeneity	4 features
ETCM	ASM, Entropy, IDM, Contrast, Maximum Probability.	5 features
CFE	The features of CLCM, TCM & ETCM	41 features

Table.1. Features Summary Table.

In the texton detection of TCM, the 2 x 2 grid moves throughout the image with one pixel as the step-length, and the detected textons based on the 5 textons and the same Texton detection in ETCM, the 2 x 2 grid moves throughout the image with one pixel as the step-length, and the detected textons based on 4 textons. So, the each pixels of an image 5 textons were applied in TCM but the ETCM the 4 textons only applied for texton detection process

4.Nucleus feature Classification:

The individual feature extraction methods of GLCM, TCM, ETCM & combining individual feature extraction method (CFE) features are tested over three classifiers such as Support Vector Machine (SVM), Radial Basis Function (RBF) and Feed Forward Neural Network (FFNN). The result of this classifiers represent classifies the cervical cell as normal cell or abnormal cell. This Examination was conducted over a set of single cervical cell based pap smear images. The dataset contains two classes of images, with a total of 952 images. From that 952 images normal cell images were 440 & abnormal cell images were 512 images.

4.1. Feature Classification using SVM:

The SVM has been widely used in pattern recognition applications due to its computational efficiency and good generalization performance. It is widely used in object detection and recognition, content-based image retrieval, text recognition, biometrics, speech recognition, etc. It creates a hyperplane that separates the data into two classes with the maximum margin. Originally it was a linear classifier based on the optimal hyperplane algorithm. A support vector machine searches an optimal separating hyper-plane between members and non-members of a given class in a high. In SVMs, the training process is very sensitive to those training data points which are away from their own class. A Support Vector Machine (SVM) is a classification model that finds an optimal separating hyperplane that discriminates two classes. An SVM is a linear discriminator; however, it can perform non-linear discriminations as this is a kernel method [24]. A special property of SVM is that it can simultaneously minimise the empirical classification error and maximise the geometric margin of a classifier. It is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation, leading to many applications including image interpretation, data mining, biometric authentication, biotechnological investigation and other electrical applications [25]. The Support Vector Machine has become one of the most popular classification methods in medical applications. The main objective of SVM is to orient the separating of the hyperplane to a direction that maximises the distance between the support vectors of each class and the hyperplane [26].

Theory of linear separable binary classification

In this study Pap smear images were categorized into two classes normal and abnormal, this kind of situation is described as classification problems. Classification of two class data is called binary classification and can be described as follows.

Assume that we have set of labelled objects denoted by the order pairs (x_i,y_i), $i = 1\ldots n$ where $(x_i \in R)$ are known as feature vectors and $y_i \in \{-1,+1\}$ are class labels. The classification task is to generate the rule that assigns any new object (point) x to one of the classes.

The support vector Machine has become one of the most popular classification methods in medical applications.

Assume that the data is linear separable, so a line can be drawn in a graph of x_1 vs. x_2 just having only two features, likewise a hyperplane can be drawn on a graph of $x_1, x_2 \ldots x_D$ … when D>2. That hyperplane is given by

$$w^t x + b = 0 \qquad\qquad (3)$$

Where w is the normal vector of the hyperplane, and the term $\frac{b}{\|w\|}$

is the perpendicular distance between the hyperplane and the origin. The Support Vectors in essence are those objects closet to the separation hyperplane. The main objective of Support Vector Machine (SVM) is to orient the separating the hyperplane to a direction that maximizes the distance between support vectors

of each class and the hyperplane. The figure below shows an illustration of two linear separable classes of data by hyperplane [27, 28].
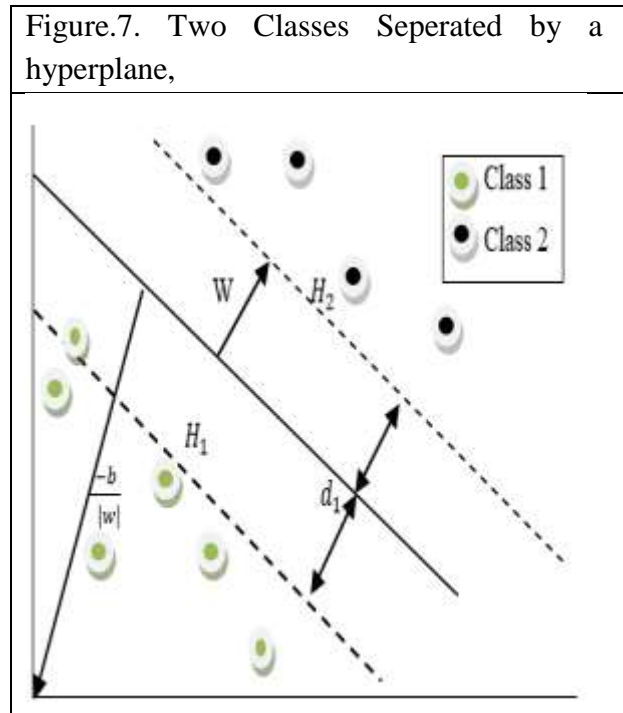

Figure.7. Two Classes Seperated by a hyperplane,

Figure.7 depicts that implementing a support vector machines (SVM) can be cut down to selecting the appropriate values of the variables w and b. Hence the training data for two classes can be described as follows:

$$w^T.x_i+b \geq +1 \tag{4}$$

For $y_i = +1$

For $w^T.x_i+b \leq -1 \tag{5}$

$y_i = -1 \tag{6}$

The generalization of these formulas can be given as follows:

$y_i (w^T.x_i+b)-1 \geq 0$ □$_i$

The two planes containing the support vectors for the classes 1 and 2 are named as H_1 and H_2 respectively, these two planes can be computed as follows:

$$w^T.x_i+b = +1 \tag{7}$$

For H_1

$$w^T.x_i+b = -1 \tag{8}$$

For H_2

The variable d_1 and d_2 are the perpendicular distances of separating hyperplanes to h_1 and h_2 respectively and are known as SVM margins. The SVM aims to orient the hyperplane in a way that it maximizes the margin. This margin is equal to $\frac{1}{\|w\|}$. To maximize this margin it is necessary to

minimize   ∥ w ∥, the maximization of the margin of separation between binary classes is equivalent to minimizing the Euclidean norm of weight vector w [28].

4.2. Feature Classification using RBF:

Kagan Tumer et al.[29] propose a radial basis function (RBF) networks with ensemble algorithms for the detection of cervical pre cancer. The RBF networks are feed forward neural networks with a single hidden layer. One class of the RBF consists the Gaussian kernels for activation function. The RBF network uses the k means algorithm to train the dataset. Francisco J. Gallegos-Funes et al.[30] propose a radial basis function (RBF) with rank M-type for classifying the Pap smear slides for detecting cervical cancer by extracting the features of the nucleus.

4.3. Feature Classification using Feed Forward Neural Network

The feed forward neural network with a single hidden layer is used with the back propagation method where the input is given as n units with an n dimensional input vector. A feed forward ANN with back propagation learning model is implemented with input, hidden and output layer. The back propagation algorithm is used in layered feed-forward ANNs. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. Jennifer Hallinan[31] proposes a new method using ANN for detecting the cell of the pap smear image as normal cell or abnormal cell.

5. Experimental Results and Discussion

5.1 Experimental image Data set.

For any machine learning algorithm, the database with which it is trained plays an important role. It is said that a machine can be made to learn and reproduce any human behaviour, provided it is trained with suitably precise database. The database prepared in this work consists of two classes of normal and abnormal cell based 952 single cervical cell presented in pap smear images. In this study two classes of single cell cervical cyto smear images totally 200 images (160 abnormal Images and 40 normal Images) have been used, which were derived from Pap smear tests collected in 2013 at Muthamil Hospital, Tirunelveli District, Tamilnadu State, India. From these 160 abnormal and 40 normal images, 80 abnormal images and 20 normal images were used for training phase. Remaining 80 abnormal and 20 normal images used for testing phase. The images were taken with 100X lens magnification using Olympus ch20i Microscope. Each image was examined and diagnosed by pathologists of that hospital before being used as reference for this study. The Sample Image data sets are shown in the table.2. The Implementation was done in the tool of matlab.

Sample Data Sets:

Table.2.Sample Image Data Set

Note: For the purpose of Images included in this table were rescaled.

5.2 Experimental Results

In this Experimental Result Section as per the three process of this paper which was referred in our previous section (2.1 & 2.2), collected images could be processed. The Result was given in following table.3.
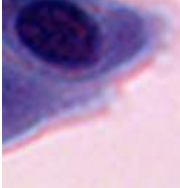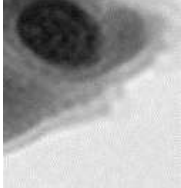
Image No	Image Name	Original Image	Gray Scale Image	Binary Image	Nucleus Segmented Gray Scale image
1	Papwork 021				
2	Papwork 022				
3	Pap work 023				
4	Pap work 024				
5	Pap work 025				
Table.3. Experimental Results					

Comparative Analysis

Individual feature Extraction and Classification Combination Method:

In this paper three individual feature extraction methods features are used. From that three, two were existing feature extraction methods namely GLCM & TCM and one was proposed ETCM method and the features of these three methods are evaluated with the three classification techniques such as SVM, RBF & FFNN. Hence, this individual method features with Classifier performance evaluation of this work is conducted with widely used statistical measures, sensitivity, specificity, accuracy and error rate [32]. TP = True Positive is correctly classified the cell as abnormal cell. FN = False Negative is incorrectly classified the cell as no cancer. FP = False Positive is incorrectly classified the cell as cancer. TN = True Negative is correctly classified the cell as no cancer. Higher values of sensitivity, the proportion of correctly classified positives, indicate better performance of the method in predicting positives. Specificity measures how well the system can predict the negatives. Accuracy measures the overall correctness of the classifier in predicting both positives and negatives, and overall error rate is calculated as per the following eqn(9-12).

$$Sensitivity = TP/(TP + FN) \quad \text{..... (9)}$$

$$Specificity = TN/(TN + FP) \quad \text{........(10)}$$

$$Accuracy = (TN + TP)/(TN + TP + FN + FP) \quad \text{......(11)}$$

$$Error\ rate = 1 - Accuracy\text{........(12)}$$

The performance analysis has been made by plotting the graphs of evaluation metrics such as sensitivity, specificity, accuracy details of the GLCM features with Classifier, TCM features with Classifier and ETCM features with Classifier are shown in respective tables.3, 4 & 5 and also these results based graphical charts are shown in the figures.(8-12).

From the results was shown in the tables.(3-5) and graphs (8-12), the proposed ETCM features with SVM Classifier combination had given the better results in the statistical parameters of sensitivity, specificity, accuracy and error rates than the other combinations such as ETCM with RBF Classifier, ETCM with FFNN Classifier, GLCM with SVM Classifier, GLCM with RBF Classifier, GLCM with FFNN Classifier, TCM with SVM Classifier, TCM with RBF Classifier & TCM with FFNN Classifier.

Combining Individual Feature Extraction features and Classification Combination Method:

In this combining individual feature extraction methods of GLCM, TCM and ETCM features are concatenated to each other and evaluating the performance with the classification methods of SVM, GLCM & TCM. This method is called as Concatenated Feature Extraction (CFE) method. The performance of this features with classification methods based Evaluation Metrics table is given in table.6 and its graphical representation is given figure.13 & 14. Then this results shows CFE is a best method when compared to all other method.

Table.3. GLCM features with SVM, RBF & FFNN Classifiers.

Evaluation Metrics		GLCM Features with SVM	GLCM Features with RBF	GLCM Features with FFNN
Input Pap Smear Image Data set	**TP**	64	62	58
	TN	8	6	4
	FP	12	14	16
	FN	16	18	22
	Sensitivity(%)	80	77.5	72.5
	Specificity(%)	40	30	20
	Accuracy(%)	72	68	62
Total Error (%)		28	32	38

Figure.8. Graphical Analysis for GLCM features and its Classification:

Table.3. TCM features with SVM, RBF & FFNN Classifiers.

Evaluation Metrics		TCM Features with SVM	TCM Features with RBF	TCM Features with FFNN
Input Pap Smear Image Data set	TP	68	66	62
	TN	10	8	8
	FP	10	12	12
	FN	12	14	18
	Sensitivity(%)	85	82.5	77.5
	Specificity(%)	50	40	40
	Accuracy(%)	78	74	70
Total Error (%)		28	22	26

Figure.9. Graphical Analysis for TCM features and its Classification
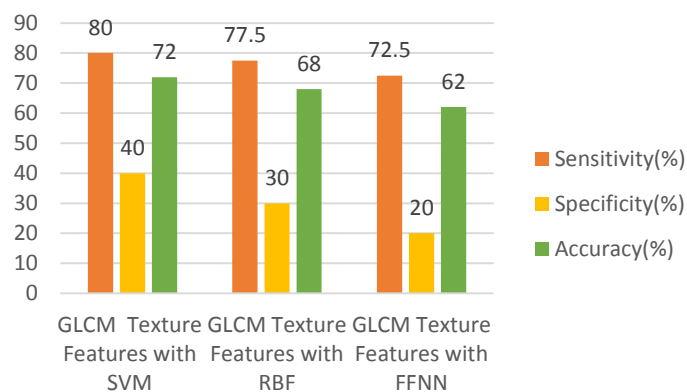

Table.3. ETCM features with SVM, RBF & FFNN Classifiers.

Evaluation Metrics		ETCM Features with SVM	ETCM Features with RBF	ETCM Features with FFNN
Input Pap Smear Image Data set	TP	72	70	66
	TN	14	12	12
	FP	6	8	8
	FN	8	10	14
	Sensitivity(%)	90	87.5	82.5
	Specificity(%)	70	60	60
	Accuracy(%)	86	82	78
Total Error (%)		28	14	18

Figure.10. Graphical Analysis for ETCM features and its Classification:
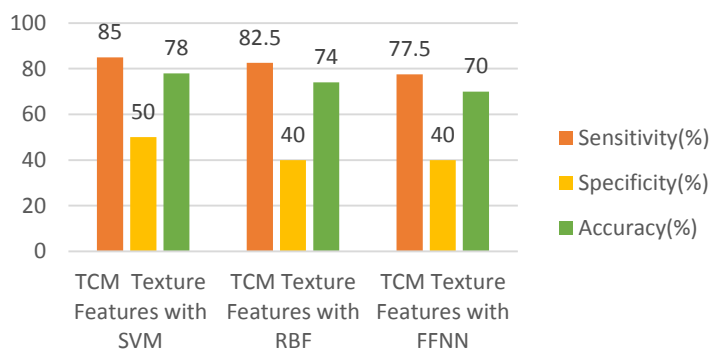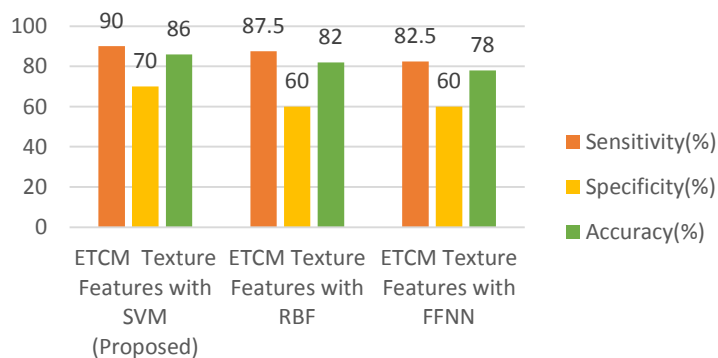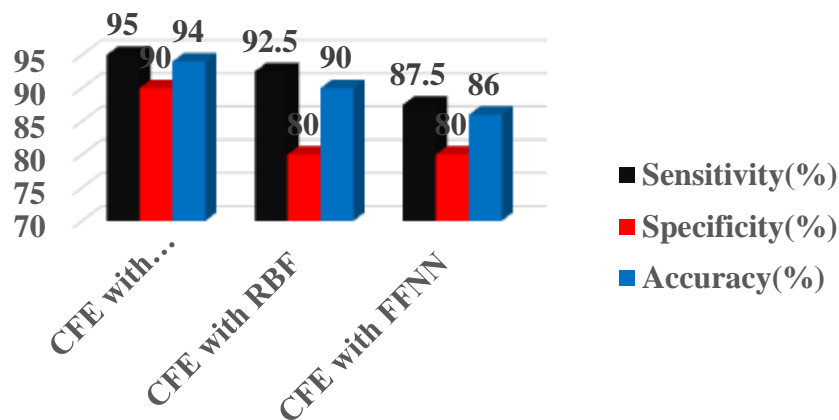
Table.3. CFE features with SVM, RBF & FFNN Classifiers.

Evaluation Metrics		CFE Features with SVM	CFE Features with RBF	CFE Features with FFNN
Input Pap Smear Image Data set	TP	76	74	70
	TN	18	16	16
	FP	2	4	4
	FN	4	6	10
	Sensitivity (%)	95	92.5	87.5
	Specificity (%)	90	80	80
	Accuracy (%)	94	90	86
Total Error (%)		28	6	10

Figure.10. Graphical Analysis for CFE features and its Classification:

Comparative analysis using the K-fold cross-validation method

This section presents the performance analysis of the proposed system (Individual features with classification system and concatenated Features with Classification System) using K-fold cross-validation method [33]. According to this, the original data set of 400 images is divided into k subsets (k=10) of data and for every validation, a single subset is used as the testing data and the remaining subsets are utilized as training data. This procedure is repeated until all the subsets of data utilized as testing data. Here, we have chosen k=10 so that, the input data are divided into '40' sub-samples to extensively analyze the proposed system. The obtained experimental results of the proposed individual and combining individual features methods were compared with the existing results in terms of the resulting parameters of sensitivity, specificity and accuracy using k-fold cross validation method.

Individual feature extraction method and classification combination:

Sensitivity Validation:

It can be observed that the sensitivity of the proposed ETCM+SVM method for the dataset 1 represent 90% and this combination when compared to the other combinations, this method has given the highest sensitivity result than all other existing feature extraction and classification combinations. This results was shown in figure.15.
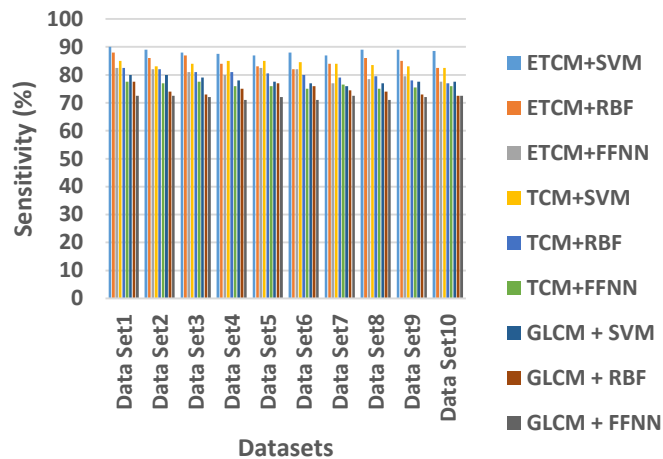
10-fold Cross Validation Method

Fig.15.Cross Validation of Sensitivity Results in proposed (ETCM) and existing (TCM & GLCM) individual feature extraction method and classification combinations.

Specificity Validation:

The same the specificity of the proposed ETCM+SVM method for the dataset 1 represent 70% and this combination when compared to the other combinations, this method has given the highest specificity result than all other existing feature extraction and classification combinations. This results was shown in figure.16.

10-fold Cross Validation Method

Figure.16. Cross Validation of Specificity Results in proposed (ETCM) and existing (TCM & GLCM) individual feature extraction method and classification combinations

Accuracy Validation:

The same the accuracy of the proposed ETCM+SVM method for the dataset 1 represent 86% and this combination when compared to the other combinations, this method has given the highest accuracy result than all other existing feature extraction and classification combinations. This results was shown in figure.17.
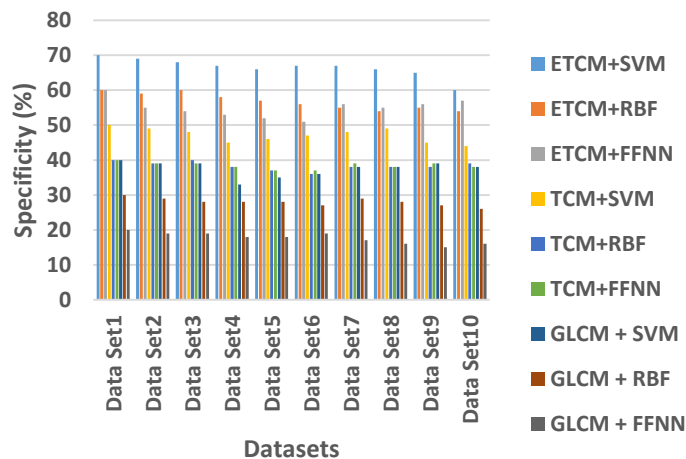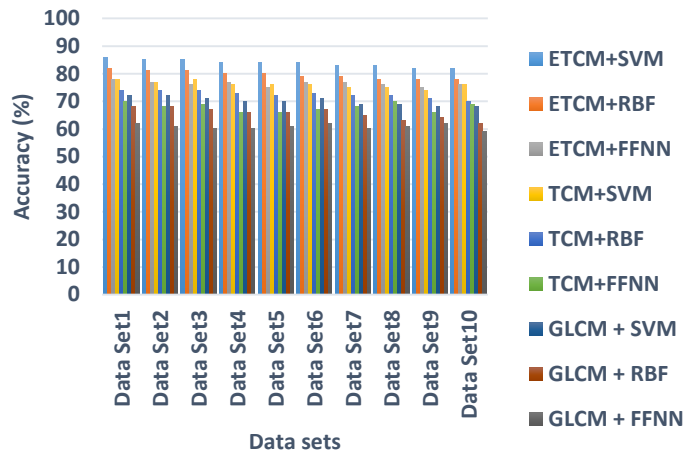
Figure.17. Cross Validation of Accuracy Results in proposed (ETCM) and existing (TCM & GLCM) individual feature extraction method and classification combinations.

Combining Individual feature extraction features and classification combination

Sensitivity Validation

It can be observed that the sensitivity of the proposed CFE+SVM method for the dataset 1 represent 95% and this combination when compared to the other combinations, this method has given the highest sensitivity result than all other CFE + RBF and CFE + FFNN combinations. This results was shown in figure.18.

Figure. 18. Cross Validation of Sensitivity Results in proposed CFE method and classification combinations

Specificity Validation

The specificity of the proposed CFE+SVM method for the dataset 1 represent 95% and this combination when compared to the other combinations, this method has given the highest specificity result than all other CFE + RBF and CFE + FFNN combinations. This results was shown in figure.19.
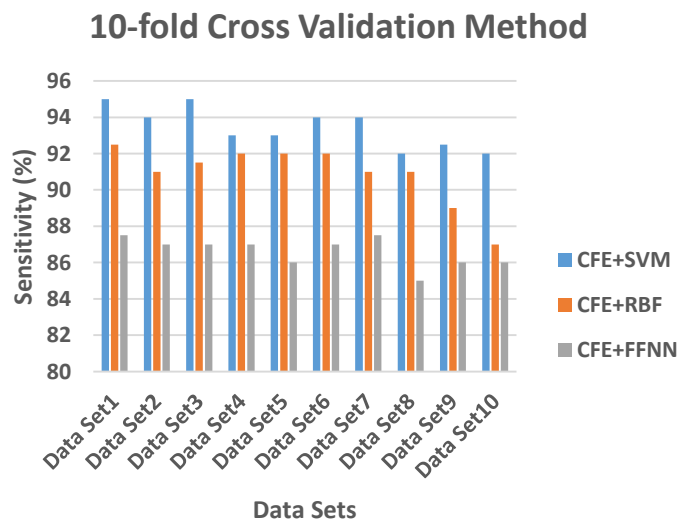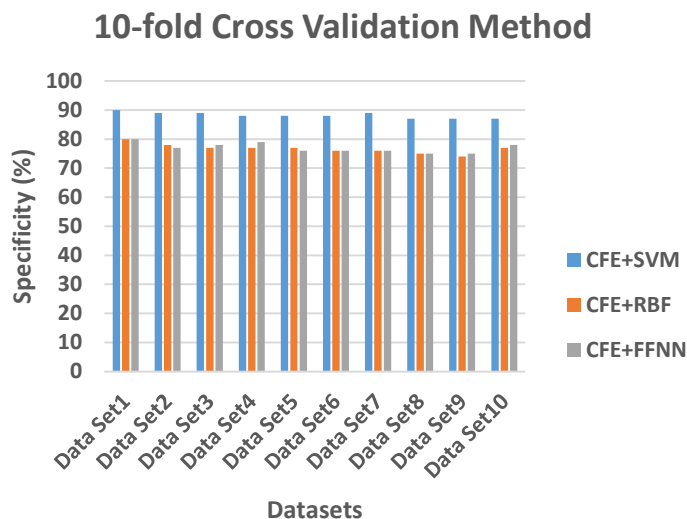
Figure. 19. Cross Validation of Specificity Results in proposed CFE method and classification combinations

Accuracy Validation

The accuracy of the proposed CFE+SVM method for the dataset 1 represent 95% and this combination when compared to the other combinations, this method has given the highest accuracy result than all other CFE + RBF and CFE + FFNN combinations. This results was shown in figure.20.
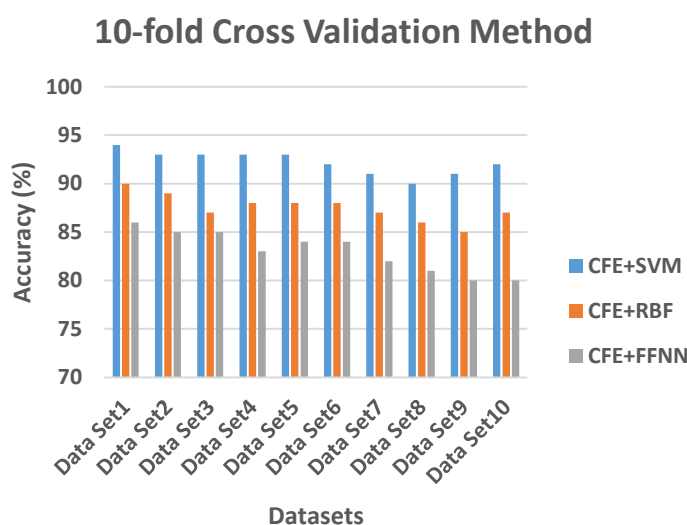

Figure. 20. Cross Validation of Accuracy Results in proposed CFE method and classification combinations

Conclusion

In this paper, we have developed an automated cervical cancer diagnostic system with normal and abnormal classes. The proposed system was designed with the individual and combining individual texture features with the classification techniques. The benefit of the system is to assist the physician to make the final decision without hesitation. According to the experimental results, the proposed method is efficient for the classification of image into normal cell and abnormal cell. For comparative analysis, in an individual feature extraction and classification combinations, our proposed (ETCM + SVM) approach provide the better results in all the statistical parameters when it compared with all other remaining feature extraction and classification combinations. In combining individual features method, the proposed CFE + SVM approach

provide the better results than all other CFE with classifications combination and also all other individual feature extraction and classification combinations. Hence, our proposed method proved that the proposed algorithm graph is good at detecting the cancer in the experimental images.

References

[1] Peter Crosta, "Cancer: Facts, Causes, Symptoms and Research", Journal of cancer-oncology, september 2015.

[2]. X. Castlellsagué, S. de Sanjosé, T. Aguado, K.S. Louie, J. M. L. Bruni, M. Diaz, K. Irwin, M. Gacic, O. Beauvais, G. Albero, E. Ferrer, S. Byrne, and F. X. Bosch, "HPV and Cervical Cancer in the World.2007 Report," Vaccine, Report WHO/ICO Information Centre of HPV and Cervical Cancer (HPV Information Centre), vol. 25, 2007.

[3] W. Li, J. Gu, D. Ferris, A. Poirson, " Automated Image Analysis of Uterine Cervical Images ", Medical Imaging 2007: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer, Proc. of SPIE Vol. 6514, 65142P, (2007).

[4] Krishnan Nallaperumal, Krishnaveni. K, et.al "An efficient Multiscale Morphological Watershed Segmentation using Gradient and Marker extraction",INDICON, 2006.

[5] A. Gentav, S. Aksoy, and S. nder. Unsupervised segmentation and classification of cervical cell images. Pattern Recognition, 45(12):4151 − 4168, 2012.

[6] Wu, H.-S., Gil, J., Barba, J.: Optimal segmentation of cell images. IEE Proc. on Vision, Image and Signal Processing. Vol. 145 (1998) 50-56.

[7] Plissiti, M., et al.: Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering. IEEE TITB, V.15, (2011) 233-241.

[8] F. Chabat, G. Yang, and D. Hansell, "Obstructive lungdiseases: texture classification for differentiation," Radiology, vol. 228, pp. 871–877, 2003.

[9] Haralick, r.m., shanmugam, k., and dinstein, "Textural features for image Classification', IEEE Transactions, November 1973, SMCJ, (6), pp. 610621.

[10] J. S. Weszka, C. R. Dyer, A. Rosenfeld, "A comparative study of texture measures for terrain classification", IEEE Transactions in System Man Cybernat,6(4), 1976, pp. 269-285.

[11] International Journal for Research in Applied Science & Engineering Technology (IJRASET), MRI Classification and Segmentation of Cervical Cancer to Find the Area of Tumor S. S. Dhumal1 , S. S. Agrawal2 1Department of EnTC, University of Pune, India Volume 3 Issue VII, July 2015 IC Value: 13.98 ISSN: 2321-9653.

[12] Meng Liang, "3D co-occurrence matrix based texture analysis applied to cervical cancer screening ", 2012.

[13]. P.P.Ohanian and R.C.Dubes,"Performance evaluation for four classes of textural features," Pattern Recognition, vol. 25, no. 8, pp.819–833,1992.

[14]Y.-F.Chen,P.-C.Huang,K.-C.Linetal.,"Semi-automatic segmentation and classification of pap smear cells," IEEE Journal of Biomedical and Health Informatics,vol.18,no.1,pp.94–108, 2014.

[15]. Edwin Jayasingh Mariarputham and Allwin Stephen, "Nominated Texture Based Cervical Cancer Classification" 2014.

[16]. Demirkaya O. "Anisotropic diffusion filtering of PET attenuation data to improve emission images", Phys MED Biol vol.47(20): N271-8.2002.

[17]. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics SMC, 3(6), 610–621.

[18]. Ondimu, S. N., & Murase, H. (2008). Effect of probability-distance based Markovian texture extraction on discrimination in biological imaging. Computers and Electronics in Agriculture, 63, 2–12.

[19].B.Julesz, Textons, The elements of Texture Perception and their interactions, Nature 290 (5802) (1981) 91-97.

[20]. T. Leung, J. Malik, Representing and Recognizing the visual appearance of materials using three-dimensional textons, International Journal of Computer Vision 43 (1) (2001) 29-44.

[21]. J.Winn, A.Crimnisi, T.Minika, Object Categorization by learned universal visual dictionary, in: Proceeding of the 10th IEEE International Conference on Computer Vision, (2005), pp. 1800-1807.

[22]. G.h.Liu, J.Y. Yang, Image Retrieval based on the Texton co-occurrence matrix, Pattern Recognition 41 (12) (2008) 3521 − 3527.

[23]. Liu G.-H., Zhang L., Hou Y.K., Z.Y. Li and Yan J.Y. g 2010 Image retrieval based on multi- texton histogram Pattern Recognition 43 2380-2389.

[24]. Luz Helena Camargo Casallas (2012) Classification of squamous cell cervical cytology, Master's Thesis, Faculty of Medicine – Engineering Faculty, Universidad Nacional de Colombia, Bogota D.C Colombia.

[25]. Yung-Fu Chen, Po-Chi Huang, Ker-Cheng Lin, Hsuan Hung Lin, Li-En Wang, Chung-chuan Cheng, Tsung-Po Chen, Yung-Kuan Chan, John Y. Chiang (2013), Semiautomatic segmentation and classification of Pap smear cells. IEEE Journal of Biomedical and Health Informatics, pp 1–15.

[26]. Ramin Moshavegh, Babak Ehteshami Bejnordi (2013) Chromatin pattern analysis of cell nuclei for improved cervical cancer detection, Master's Thesis in Biomedical Engineering, Department of Signal and systems, Chalmers University of Technology, Sweden.

[27]. S. Theodoridis and K. Koutroumbas (2009) Pattern Recognition, 4th ed. China Machine Press.

[28]. Simon Haykin (2009) Neural networks and learning machines, 3rd ed. China Machine press, Pearson education Asia Ltd.

[29]. K. Tumer, N. Ramanujam, J. Ghosh and R. Richards-Kortum, Ensembles of Radial Basis Function Networks for Spectroscopic Detection of Cervical Precancer, IEEE Transactions on Bio-Medical Engineering, vol. 45(8), pp. 953–961, (1998). http://doi.org/10.1109/10.704864.

[30]. F. J., Gallegos-Funes, M. E. G´omez-Mayorga, J. L., Lopez-Bonilla and R. Cruz-Santiago, Rank M-Type Radial Basis Function (RMRBF) Neural Network for Pap Smear Microscopic Image Classification, Apeiron, vol. 16(4), pp. 542–554, (2009).

[31]. J.Hallinan, Evolving Neural Networks for the Classification of Malignancy Associated Changes, Intelligent Data Engineering and Automated Learning Ideal 2005, Proceedings, vol. 3578, pp. 382–389, (2005).

[32]. Wen Zhu, Nancy Zeng, Ning Wang, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations", Proceedings of the SAS Conference, Baltimore, Maryland, pages: 9, 2010.

[33]. Ounpraseuth S , Lensing SY , Spencer HJ , Kodell RL. "Estimating misclassification error: a closer look at cross-validation based methods", BMC Res Notes (28) : 5:656,2012.