

Multi Class Cervical Cancer Classification by using ERSTCM, EMSD & CFE methods based Texture Features and Fuzzy Logic based Hybrid Kernel Support Vector Machine Classifier

S.Athinarayanan¹, Dr.M.V.Srinath², R.Kavitha³

¹(Research Scholar, Mononmaniam Sundaranar University, India)

²(Director, Department of MCA, STET Women's College, Sundarakkottai, India)

³(Assistant Professor, Department of BCA, The MDT Hindu College, Pettai, India)

Abstract: Cervical cancer is the highest rate of incidence after breast cancer, gastric cancer, colorectal cancer, thyroid cancer among all malignant that occurs to females ; also it is the most prevalent cancer among female genital cancers. Manual cervical cancer diagnosis methods are costly and sometimes result inaccurate diagnosis caused by human error but machine assisted classification system can reduce financial costs and increase screening accuracy. In this research article, we have developed multi class cervical classification system by using Pap Smear Images according to the WHO descriptive Classification of Cervical Histology. Then, this system classifies the cell of the Pap Smear image into anyone of five types of the classes of normal cell, mild dysplasia, moderate dysplasia, severe dysplasia and carcinoma in situ (CIS) by using individual and Combining individual feature extraction method with the classification technique. In this paper three Feature Extraction methods were used: From that three, two were individual feature extraction method namely Enriched Rough Set Texton Co-Occurrence Matrix (ERSTCM) and Enriched Micro Structure Descriptor (EMSD) and the remained one was combining individual feature extraction method namely concatenated feature extraction method (CFE). The CFE method represents all the individual feature extraction methods of ERSTCM & EMSD features are combining together to one feature to assess their joint performance. Then these three feature extraction methods are tested over Fuzzy Logic based Hybrid Kernel Support Vector Machine (FL-HKSVM) Classifier. This Examination was conducted over a set of single cervical cell based pap smear images. The dataset contains five classes of images, with a total of 952 images. The distribution of number of images per class is not uniform. Then the performance was evaluated in both the individual and combining individual feature extraction method with the classification techniques by using the statistical parameters of sensitivity, specificity & accuracy. Hence the resultant values of the statistical parameters described in individual feature extraction method with the classification technique, proposed EMSD+FLHKSVM Classifier had given the better results than the other ERSTCM+FLHKSVM Classifier and combining individual feature extraction method with the classification technique described, proposed CFE+FLHKSVM Classifier had given the better results than other EMSD+FLHKSVM & ERSTCM+FLHKSVM classifiers.

Keywords: Cervical Cancer, Pap Smear Images, Feature Extraction, Classification.

I. Introduction

Cervical cancer is a second most common cancers affecting women in worldwide [1]. It can be cured, if it is detected at early stages and also identify the stage for providing the proper treatment. In the meanwhile, causes of the death rate of this cancer is high. It is reported that yearly 1 lakh new cases were diagnosed among these 75 thousand were deaths in India, which is nearly one-third of the global cancer deaths [2]. Pap test is a effective screening test of cervical cancer, which is said to be the gold standard forever. Due to the subjective discrepancy of different cytologists, this test results show more of irregularities [3]. The test output shows more of false positive and false negative results, which make the trustworthiness of the screening process is a question mark [4]. Also the manual screening process, hundreds of images are analyzed daily; the classification of cells become difficult and the chance of human errors become high. Many automatic and semiautomatic methods have been proposed in several times to detect various stages of cervical cancer. Many of these methods were not supported in achieving the objectives of providing measured variables which could eliminate the interpretation errors and inter observer discrepancy [5]. Pap smear images are rich in various features like color, shape, and texture. The process of accurate extraction of unique visual features from these images would very well help in developing an automated screening device. This can be achieved only through texture feature than others since all the cellular changes are observed only using these features. Since the texture parameters are the simple mathematical representations like smooth, rough, and grainy, then the analysis becomes easier [6]. Analyzing all the above issues, two important challenges are considered. First, selection of unique texture features suitable for classification. Second, selection of the most efficient and scalable classifier improves the accuracy more.

II. Related Works

Plissiti et al. [7] have developed the fully automated method to detect the nucleus in Pap smear images. Using morphological analysis, the nuclei centroids are detected by applying distance dependent rule and classification algorithms on resulted centroids, the undesirable artifacts were removed from the cell. By considering nucleus as the most informative region of the cell, Sobrevilla et al. [8] have been presented an algorithm for automatic nuclei detection of cytology cell. This algorithm combines color, cyto pathologist's knowledge, and fuzzy systems which show high performance and more computational speed.

Bergmeir et al. [9] have developed an algorithm used to detect cell nuclei and cytoplasm. This algorithm used the combination of voting scheme and prior knowledge to locate the cell nuclei and elastic segmentation to determine the shape of nucleus. The noise is removed with mean-shift and median filters, and edges were extracted with canny edge detection algorithm. The automatic classification of Pap smear images focuses on marking of single cells into any one of binary classes (normal and abnormal) or multiple classes (based on severity). Multiple classification of Pap smear images became popular when CIN based classification on cervical cytology images was proposed.

Holmquist et al. [10] developed a binary classification method to distinguish between normal and abnormal cells. Marinakis et al. [11] proposed a metaheuristic approach to classify the Pap smear cells. Uniquely described twenty features are extracted from each cell image and classified as normal and abnormal type. Chen et al. [12] have developed an algorithm for segmenting nucleus and cytoplasm counters. 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.

III. multi class cervical cancer classification system

Current manual screening methods are costly and sometimes result in inaccurate diagnosis caused by human error. The introduction of machine assisted screening will bring significant benefits to the community, which can reduce financial costs and increase screening accuracy. In this research article, we have developed, Multi class cervical cancer classification system based on individual and combining individual texture features and Fuzzy Logic based hybrid kernel support vector machine using pap smear images. Two major contribution of the proposed system is feature extraction and feature classification. The architecture of the proposed classification system is depicted in Figure 1. The system involves the following steps: preparation of Pap smear images and pre-processing, segmentation of nucleus, extraction of texture features, and classification.

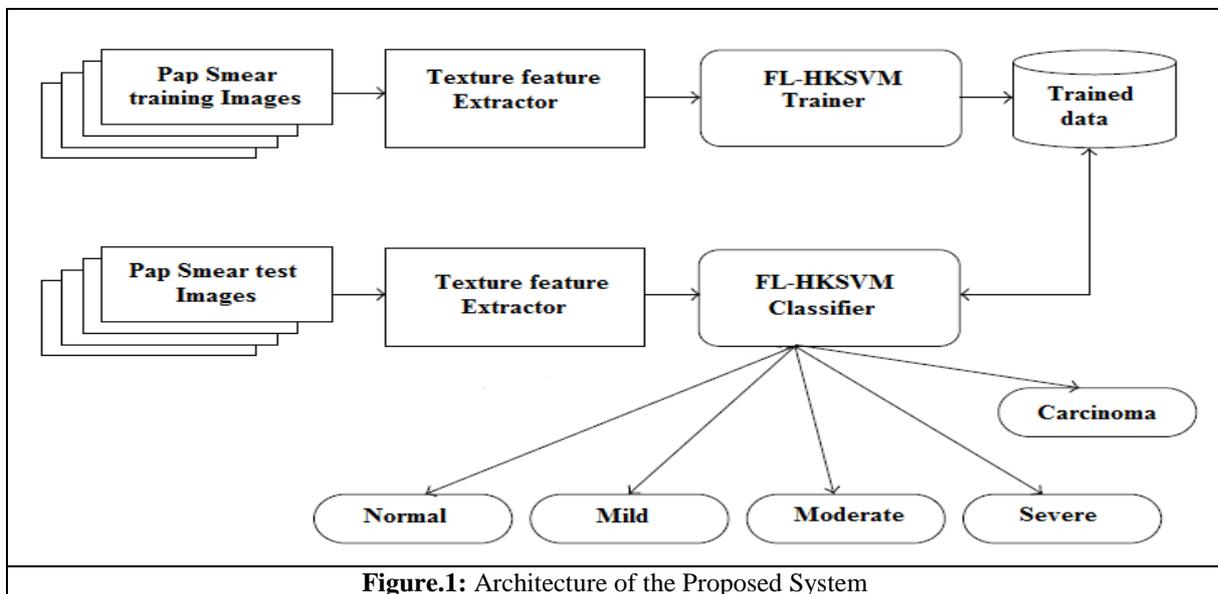


Figure.1: Architecture of the Proposed System

3.1 Feature Extraction

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, we proposed three novel feature extraction methods. From that three, two were individual feature extraction methods, they are Enriched Rough Set Texton Co-Occurrence Matrix (ERSTCM) & Enriched Micro Structure Descriptor (EMSD) and the remained one was combining

individual feature extraction features method named as Concatenated Feature Extraction (CFE). The CFE method represents all the individual feature extraction methods of ERSTCM & EMSD 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.

Individual Feature Extraction Methods:

- Computation of Feature Vector $F(V1)$ using ERSTCM.
- Computation of Feature Vector $F(V2)$ using EMSD.

Combining Individual Feature Extraction Features Method:

- Computation of Feature Vector $F(V3)$ using CFE.

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

3.1.1. Individual Feature Extraction Methods:

3.1.1.1. Computation of Feature Vector $F(V1)$ using ERSTCM:

As per Julesz [13] description a texton is a pattern which is shared by an image as a common property all over the image. Textures will be formed only if the adjacent elements lie within the neighbourhood. Texton Image has the discrimination power of color, texture and shape features. Based on the Julesz's [13] textons theory the texton co-occurrence matrices (TCM) algorithm, can describe the spatial correlation of textons for image retrieval. With the use of limited number of selected pixels the algorithm computes different features. Extraction of effective texture information will be increased by calculating texton using Rough texture. This method is represented Enriched Rough Set Texton Co-occurrence Matrices. In this Method. As a texture descriptor, good retrieval accuracy can be achieved especially for directional textures. The texton can be inclined with the use of critical distances between texture elements. Texture element size determines the critical distance. Texture can be resolved into minute units, like orientation, Texton classes of colors, elongated blobs of specific width, aspect ratios and terminators of elongated blobs. If texture elements are expanded to a large extent in one orientation discrimination reduces. If the elongated elements are not jittered in orientation texture gradients increase at boundaries. Hence by using a sub image of size 3×3 a texton gradient can be obtained. Here 20 textons of 3×3 grids are proposed as shown in Figure 2. Even the co-occurrence probability of same valued pixels in 3×3 grids is smaller than that of 2×2 grid, but the textons developed using 2×2 grid may not give complete information regarding direction. The computational complexity for using the overlapped components of 20 textons is also less to obtain final texton image. The 20 textons of 3×3 grid can detect textons in all directions and also the corners of the textures. If three pixels are highlighted and have same value then, grid will form a texton as shown in Figure 3.

Texton Detection

The Working Mechanism of Texton detection is illustrated in figure.4. In the final segmented nucleus image, we move 3×3 block from left to right and top to bottom throughout the image to detect textons with 3 pixel as step length. If a texton is detected, the same value of the three pixels in the 3×3 grids are kept unchanged and the remained pixels are zero value. Finally we will obtain a Texton image, denoted by $T(x,y)$. After the formation of final texton image, the feature vector $F(V1)$ (Seven features such as contrast, inverse difference moment, correlation, variance, cluster shade, cluster prominence and Homogeneity) is extracted from it.

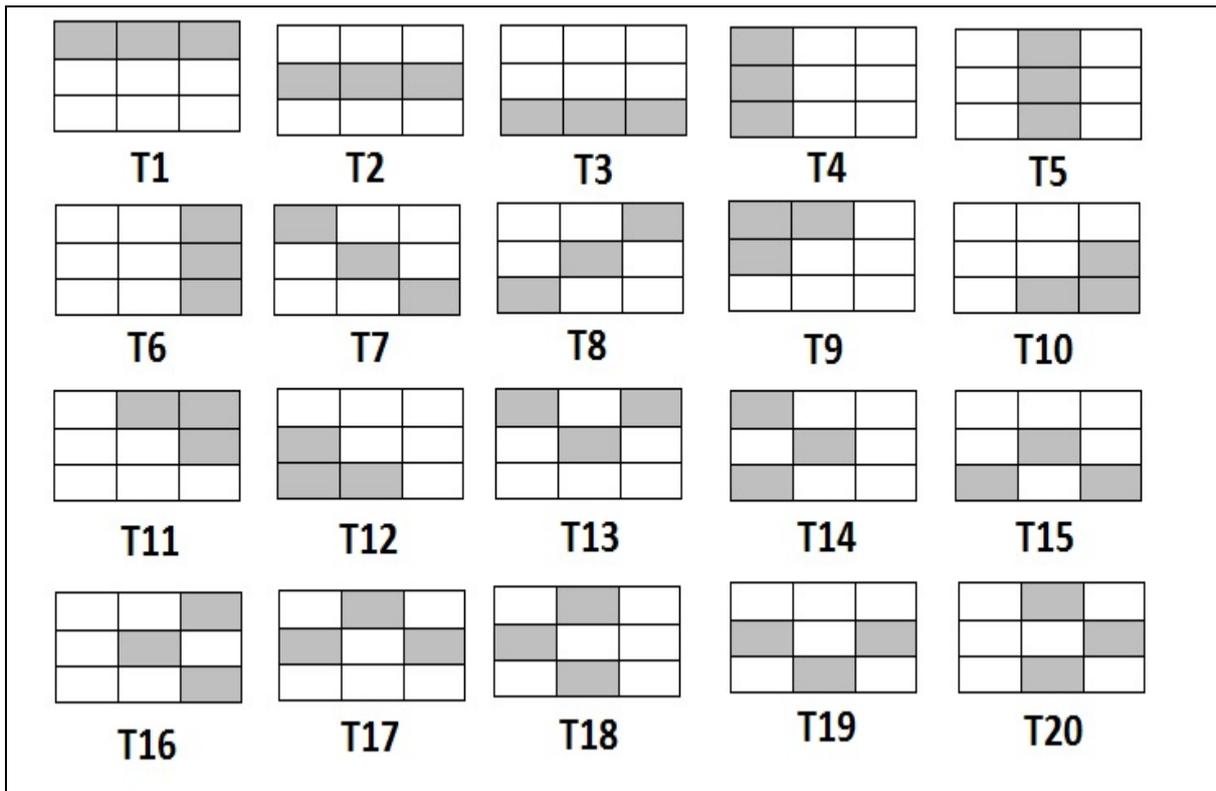


Figure.2 : 20 types of 3 x 3 Textons

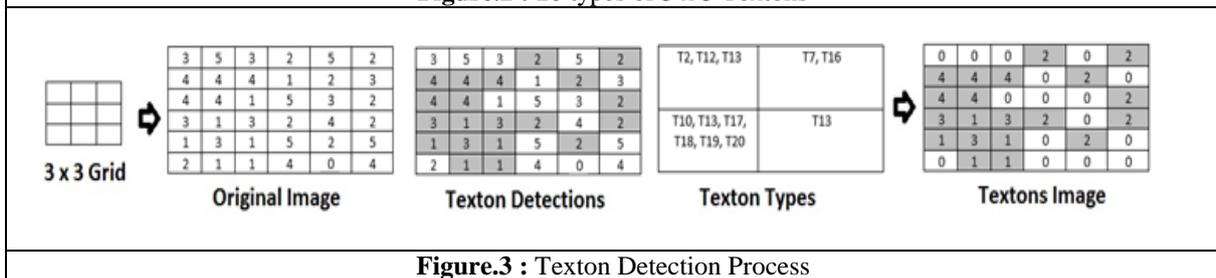


Figure.3 : Texton Detection Process

3.1.1.2. Computation of Feature Vector F(V2) using EMSD:

Julesz’s texton theory mainly focuses on analyzing regular textures, while the micro-structures can be considered as the extension of Julesz’s textons or the color version of textons. Since micro-structures involve color, texture and shape information, they can better present image features for image retrieval. In our method in order to find the micro-structures, we partition the original image into 3 x 3 block. Then We move the 3x3 block from left-to-right and top-to-bottom throughout the image to detect micro-structures. In the 3x3 block, if one of the eight nearest neighbors has the same value as the center pixel, then it is kept unchanged; otherwise it is set to empty. If all the eight nearest neighboring pixels are empty, then the 3x3 block is not considered as a micro-structure and all pixels in the 3x3 block will be set to empty. The pattern resulted from this operation is called a fundamental micro-structure. Figure. 4(a-d) shows an example of the micro-structure detection process.

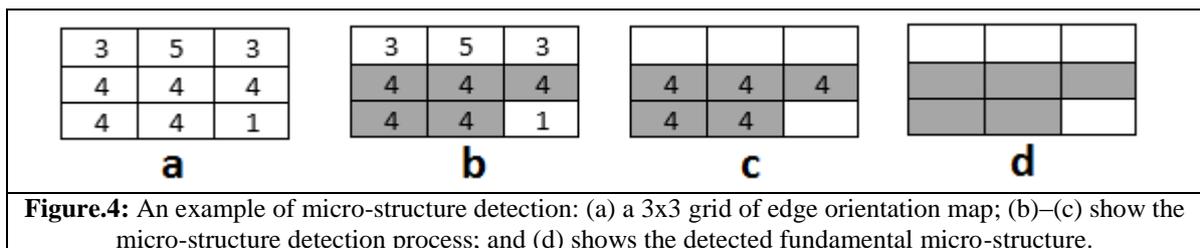


Figure.4: An example of micro-structure detection: (a) a 3x3 grid of edge orientation map; (b)–(c) show the micro-structure detection process; and (d) shows the detected fundamental micro-structure.

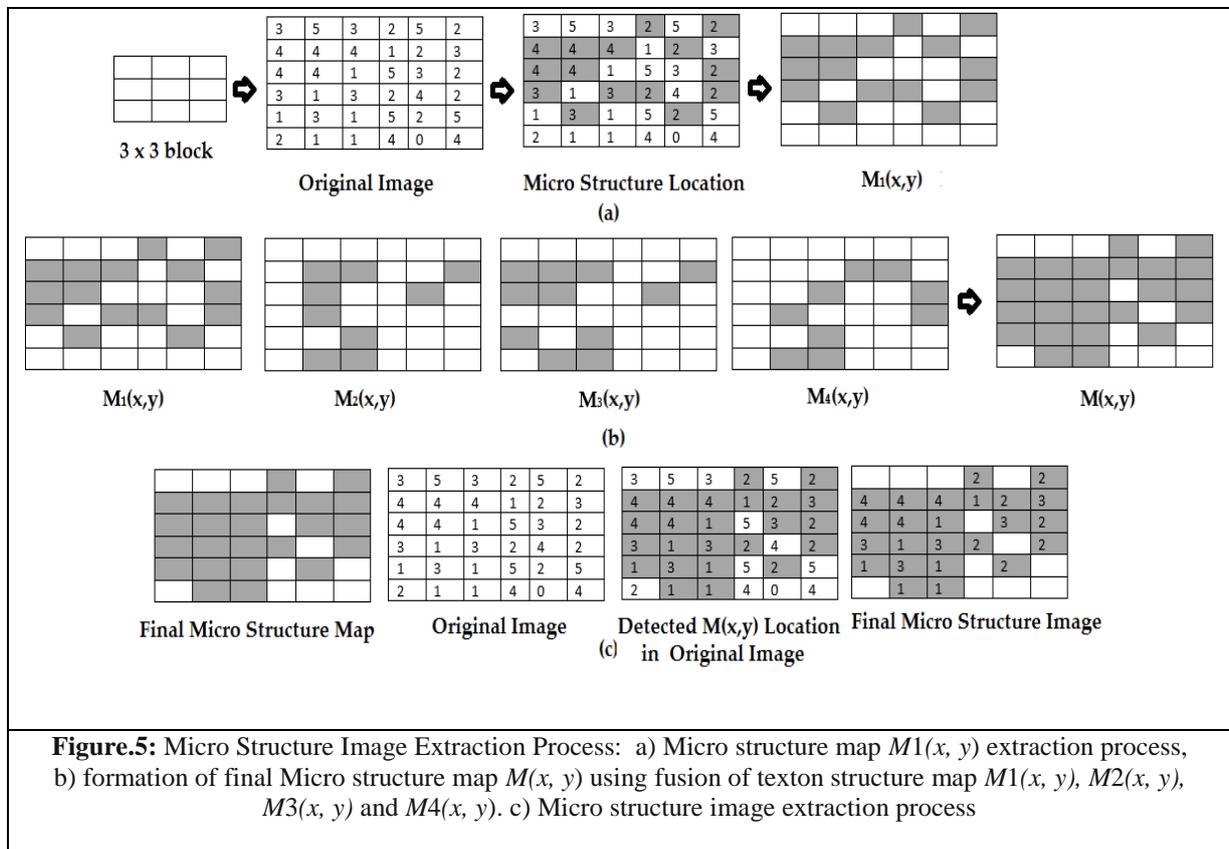
The micro structure map extraction process is a four-step process as described below and is shown in Fig. 5a.

- i. Divide the original image $f(x, y)$ into 3×3 blocks.
- ii. Move the 3×3 block horizontally and vertically from left to right and top to bottom throughout the original image $f(x, y)$ with a step length of three pixels from the origin (0, 0). Then, generate the Micro structure map $M1(x, y)$, where $0 \leq x \leq W-1, 0 \leq y \leq N-1$.
- iii. Repeatedly doing the step (ii) from the origin (0, 1), (1, 0) and (1, 1) and generate the micro structure maps $M2(x, y)$, where $0 \leq x \leq W-1, 1 \leq y \leq N-1$, $M3(x, y)$, where $1 \leq x \leq W-1, 0 \leq y \leq N-1$ and $M4(x, y)$, where $1 \leq x \leq W-1, 1 \leq y \leq N-1$, respectively.
- iv. Generate the final micro structure image $M(x, y)$ using the fusion of micro structure map as per Eq.1.

$$M(x, y) = \text{Max} \{M1(x, y), M2(x, y), M3(x, y), M4(x, y)\} \text{ ----(1)}$$

The formation of final micro structure map $M(x, y)$ using fusion of micro structure map $M1(x, y), M2(x, y), M3(x, y)$ and $M4(x, y)$ is shown in Fig. 5b.

After the final micro-structure map $M(x,y)$ is extracted from the original image $f(x,y)$, The micro structure map is applied to the same original image. The pixels that do not match the map are set to empty. Based on this process, we get the final micro structure image. This process was shown in Fig. 5c. After the formation of final texton image, the feature vector $F(V2)$ (Seven features such as contrast, inverse difference moment, correlation, variance, cluster shade, cluster prominence and Homogeneity) is extracted from it.



Difference between ERSTCM and EEETCM:

The formation of Micro Structure Map based process of EMSD contains rich information than 20 types of Texton based ERSTCM. This was shown in Figure 6. Then the Computational Complexity of EMSD is also less to obtain final image than ERSTCM, Because ERSTCM contains 20 textons, so each 3×3 block of the image it needs to apply that 20 textons for getting the final image, but in EMSD 3×3 block is similar, but the process applying direction is less when compared to ERSTCM.

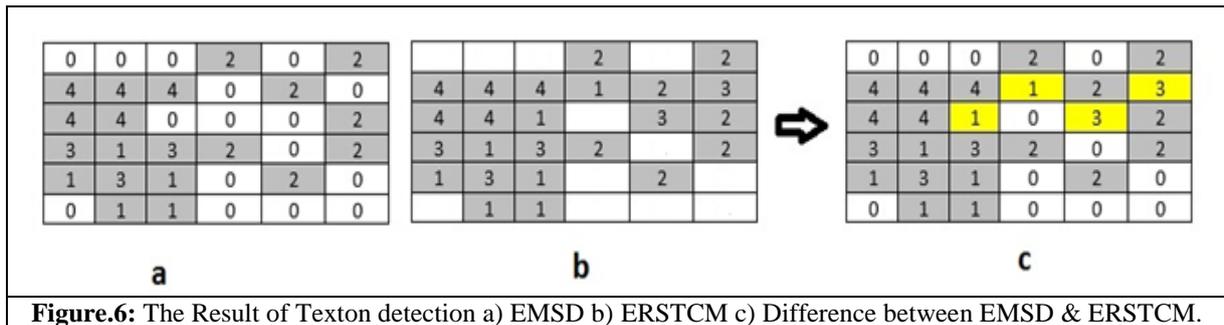


Figure.6: The Result of Texton detection a) EMSD b) ERSTCM c) Difference between EMSD & ERSTCM.

3.1.2. Combining Individual Feature Extraction Features Method:

3.1.2.1 Computation of Feature Vector F(V3) using CFE:

The name CFE Stands for Concatenated Feature Extraction. The CFE method represents all the individual feature extraction methods of ERSTCM & EMSD features are Combining together to one feature to assess their joint performance. Hence, total Feature vector uses $F(V3) = F(V1) + F(V2)$ dimensional vector as the concluding image features in cervical cell classification.

Table.1. Summary of Texture Features

The three feature extraction methods and its features details are given in the following Table 1.

Feature category	Features	Number of features
ERSTCM	Contrast, inverse difference moment, correlation, variance, cluster shade, cluster prominence and Homogeneity	7 features
EMSD	Contrast, Inverse difference moment, correlation, variance, cluster shade, cluster prominence and Homogeneity	7 features
CFE	The features of ERSTCM & EMSD	14 features

Table.1: Features Summary Table.

3.2. Nucleus feature Classification:

The individual feature extraction methods of ERSTCM & EMSD & Combining individual feature extraction features method (CFE) features are tested over Fuzzy Logic based Hybrid Kernel Support Vector Machine (FL-HKSVM) classifier. The result of this classifier represent classifies the cervical cell of the pap smear image into any one of the classes of normal, mild dysplasia, moderate dysplasia, severe dysplasia and Carcinoma in Situ (CIS). This Examination was conducted over a set of single cervical cell based pap smear images. The dataset contains four classes of abnormal cell images, with a total of 952 images. From that 912 images, normal – 440, mild dysplasia - 107, moderate dysplasia - 124, severe dysplasia - 139, & Carcinoma in Situ (CIS) were 142 images.

3.2.1. Feature Classification using FL-HKSVM:

The proposed classification system classify the Pap smear images into any one of five classes, namely normal, mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ. The manual classification of the entire dataset has been already done by the experts. In the proposed classification system, we used Fuzzy Logic based Hybrid Kernel SVM algorithm to classify the Pap smear images.

There are two phases in the support vector machine namely.

(1) Training phase and (2) Testing phase.

Fuzzy logic based SVM (FSVM) is an effective supervised classifier and accurate learning technique, which was first proposed by Lin and Wang [14]. In FSVM is to assign each data point a membership value according to its relative importance in the class. Since each data point x_i has an assigned membership value μ_i , the training set s_f and is given by Eq.2.

$$s_f = \{x_i, y_i, \mu_i\}_{i=1}^n \tag{2}$$

For positive class ($y_i = +1$), the set of membership values are denoted as μ_i^+ , and are denoted as μ_i^- for negative class ($y_i = -1$), they are assigned independently. The main process of fuzzy SVM is to maximize the margin of separation and minimize the classification error.

The optimal hyperplane problem of FSVM can be defined as the following problem [15] and is given by Eq.3.

$$\min_{w, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n f_i \varepsilon_i \tag{3}$$

Subject to $y_i(w \cdot x_i + b) \geq 1 - \varepsilon_i$
 $\varepsilon_i \geq 0, i = 1, \dots, n$

Where $f_i (0 \leq f_i \leq 1)$ is the fuzzy membership function, $f_i \varepsilon_i$ is a error of different weights and C is a constant

The inputs to FSVM algorithm are the feature subset selected via ERSTCM, EMSD & CFE. It follows the structural risk minimization principle from the statistical learning theory. Its kernel is to control the practical risk and classification capacity in order to broaden the margin between the classes and reduce the true costs. A Fuzzy support vector machine searches an optimal separating hyper-plane between members and non-members of a given class in a high dimension feature space.

The Lagrange multiplier function of FSVM is Eq.4.

$$L(w, b, \xi, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n f_i \xi_i - \sum_{i=1}^n \alpha_i (y_i (wz_i + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \tag{4}$$

Which satisfies the following parameter condition

$$w - \sum_{i=1}^n \alpha_i y_i z_i = 0$$

$$- \sum_{i=1}^n \alpha_i y_i = 0$$

$$f_i C - \alpha_i - \beta_i = 0$$

Then the optimization problem can be transferred and defined in Eq.5.

$$\begin{aligned} \text{Max } W(\alpha) &= \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j H(x_i, x_j) \\ \text{Subject to: } & \sum \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq f_i C, i = 1, 2, \dots, n \end{aligned} \tag{5}$$

Where the parameter α_i can be solved by the sequential minimal optimization (SMO) quadratic programming approach. We have analyzed the kernel equation from the existing work (Admasua et al., 2003) and used them in the proposed work, namely, RBF and quadratic function.

Radial basis function:

The Support Vector will be the centre of the RBF and σ will determine the area of influence. This support vector has the data space, it is given in Eq. 6:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \text{ ----- (6)}$$

Quadratic Kernel function:

Polynomial kernels are the form $K(\vec{X}, \vec{Z}) = (1 + \vec{X}^T \vec{Z})^d$. Where d=1, a linear kernel and d=2, a quadratic kernel are commonly used. Let K_1 (RBF), K_2 (Quadratic) be kernels over $\mathbb{R}^p \times \mathbb{R}^p$ and K_3 be a kernel over $\mathbb{R}^p \times \mathbb{R}^p$. Let function $\varphi: \mathbb{R}^p \rightarrow \mathbb{R}^p$. The two kernels based formulations are represented in Eq. 7 and 8:

$$K(x, y) = k_1(x, y) + k_2(x, y) \text{ is a kernel. ----- (7)}$$

$$K(x, y) = k_1(x, y)k_2(x, y) \text{ is a kernel. ----- (8)}$$

Substitute Eq. 7 and 8 lagrange multiplier Eq. 5 and get the proposed hybrid kernel. It is exposed in Eq.9:

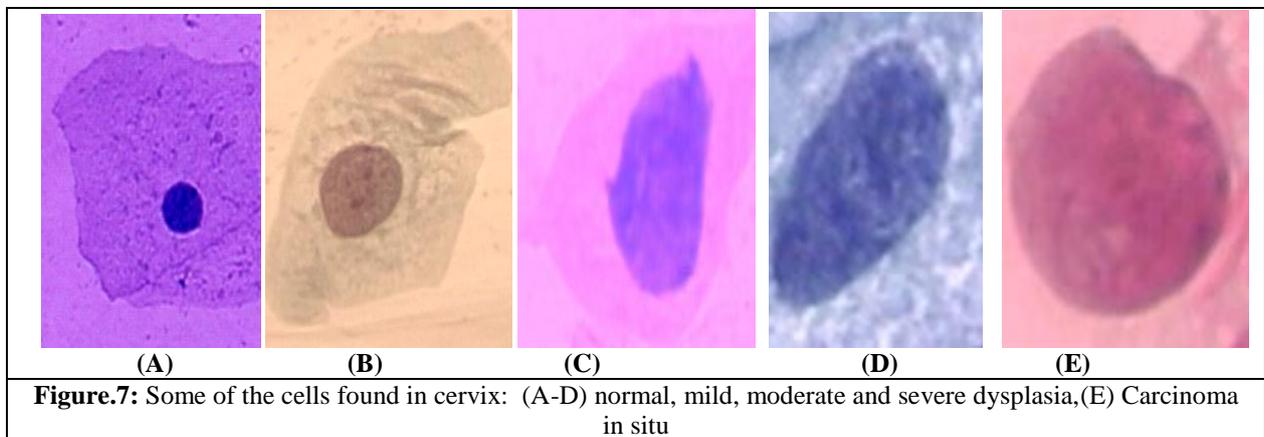
$$\begin{aligned} \text{Max } W(\alpha) &= \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j (k_1(x_i, x_j) + k_2(x_i, x_j)) \\ \text{Max } W(\alpha) &= \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j (k_1(x_i, x_j) \cdot k_2(x_i, x_j)) \text{ ----- (9)} \end{aligned}$$

IV. Experimental Results And Discussion

4.1 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 five classes of normal & abnormal cell based 952 single cervical cell presented in pap smear images. These Images were collected from Muthamil Hospital, Tirunelveli District, Tamilnadu State, India in 2013 and these 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 sample Pap smear experimental images of various classes is shown in Figure 7. The data set contains 952 images and these images were histologically diagnosed and graded based on World Health Organization (WHO) criteria as the following abnormal class distribution:

- Normal. 440 images.
- Mild dysplasia, 107 images.
- Moderate dysplasia, 124 images.
- Severe dysplasia, 139 images.
- Carcinoma in Situ. 142 images.



The Implementation was done in the tool of matlab.

4.2. Experimental Results & Comparative Analysis

This section describes the experimental results of the proposed classification method using Pap smear images with different types of cervical cancer. In the proposed method, the experimental image data set is divided into two sets such as training set and testing set. The details of this set was shown in Table 2. The classifiers are trained with the training images and the classification accuracy is calculated only with the testing images. In the testing phase, the testing dataset is given to the proposed technique to find the cancers type in smear images and the obtained results are evaluated through evaluation metrics namely, sensitivity, specificity and accuracy [16], it is given by (eqn.10-12)

$$Sensitivity = TP / (TP + FN) \text{ -----(10)}$$

$$Specificity = TN / (TN + FP) \text{ -----(11)}$$

$$Accuracy = (TN + TP) / (TN + TP + FN + FP) \text{ ----- (12)}$$

For example, Where TP corresponds to True Positive, TN corresponds to True Negative, FP corresponds to False Positive and FN corresponds to False Negative. These parameters for a specific category, say, Mild dysplasia are as follows: TP is True Positive (an image of 'Mild dysplasia' type is categorized correctly to the same type), TN = True Negative (an image of 'Non- Mild dysplasia' type is categorized correctly as 'Non- Mild dysplasia' type), FP =False Positive (an image of 'Non- Mild dysplasia' type is categorized wrongly as 'Mild dysplasia' type) and FN is False Negative (an image of 'Mild dysplasia' type is categorized wrongly as 'Non- Mild dysplasia' type). 'Non- Mild dysplasia' actually corresponds to any of the three categories other than 'Mild dysplasia'. Thus, 'TP & TN' corresponds to the correctly classier images and 'FP & FN' corresponds to the misclassified images.

Table.2: Experimental image dataset for classification

Cancer Type	Training data	Testing data	Total no of images
Normal	200	240	440
Mild dysplasia	50	57	107
Moderate dysplasia	50	74	124
Severe dysplasia	50	89	139
carcinoma in situ	50	92	142
Total	400	552	952

4.2.1. Individual feature Extraction and Classification Combination Results and Comparison:

Thus, different parameter values are obtained for each class and also for the different classifiers. These parameters are estimated from the confusion matrix which provides the details about the false and successful classification of images from all categories for each classifiers. The confusion matrix of the ERSTCM+FL-HKSVM method is illustrated in Table 3.

Table.3: Confusion matrix of proposed method (ERSTCM+FL-HKSVM)

Class predicted	Ground Truth Class (Assigned by Pathologist)				
	Normal	Mild dysplasia	Moderate dysplasia	Severe dysplasia	Carcinoma in situ
Normal	179	16	19	15	11
Mild dysplasia	1	50	1	3	2
Moderate dysplasia	4	4	62	2	2
Severe dysplasia	2	1	2	82	2
Carcinoma in situ	1	2	1	2	86

In the Table 3, the row-wise elements correspond to the four categories and the column-wise elements correspond to the target class associated with that abnormal category. Hence, the number of images correctly classified (TP) under each category is determined by the diagonal elements of the matrix. The row-wise summation of elements for each category other than the diagonal elements corresponds to the ‘FN’ of that category. The column-wise summation of elements for each category other than the diagonal element corresponds to the ‘FP’ of that category. Similarly, ‘TN’ of the specific category is determined by summing the elements of the matrix other than the elements in the corresponding row and column of the specific category.

For example, among the 57 Mild dysplasia testing images, 50 images have been successfully classified (TP) and the remaining 7 images (first row-wise summation) have been misclassified to any of the non- Mild dysplasia categories (FN). Similarly 23 images (first column-wise summation) from the other three categories (non- Mild dysplasia) have been misclassified as Mild dysplasia category (FP). In the Table 4, the classification accuracy of ERSTCM with FL-HKSVM in class 1 (Normal) is 87.50%, class 2 (Mild dysplasia) type cancer is 94.57%, class 3(Moderate dysplasia) type cancer is 93.66%, class 4(Severe dysplasia) type cancer is 94.75% and class 5(carcinoma in situ) type cancer is 95.83%. then the overall sensitivity, specificity & accuracy of this ERSTCM + FL-HKSVM is 83.15%, 95.79% & 93.26%. The performance of overall calculation of sensitivity, specificity and accuracy by using this Eq.13-15.

The miss classification rate of class 1(Normal) non cancer type, Class 2 (Mild dysplasia), class 3 (Moderate dysplasia) & class 4 (Severe dysplasia) of cancer type is high compared to the other class of class 5 (Carcinoma in Situ).

Table.4: Performance measure of ERSTCM with FL-HKSVM

Class predicted	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)
Normal	179	304	8	61	74.58	97.44	87.50
Mild dysplasia	50	472	23	7	87.72	95.35	94.57
Moderate dysplasia	62	455	23	12	83.78	95.19	93.66
Severe dysplasia	82	441	22	7	92.13	95.25	94.75
carcinoma in situ	86	443	17	6	93.48	96.30	95.83
Overall Sensitivity, Specificity & Accuracy Results					83.15	95.79	93.26

$$\text{Overall classification Sensitivity} = \left(\frac{\sum_i TP(\text{Class}(i))}{\sum_i TP(\text{Class}(i)) + FN(\text{Class}(i))} \right) * 100 \text{ ----- (13)}$$

$$\text{Overall classification Specificity} = \left(\frac{\sum_i TN(\text{Class}(i))}{\sum_i TN(\text{Class}(i)) + FP(\text{Class}(i))} \right) * 100 \text{ -----(14)}$$

$$\text{Overall classification accuracy} = \left(\frac{\sum_i TP(\text{Class}(i)) + TN(\text{Class}(i))}{\sum_i TP(\text{Class}(i)) + TN(\text{Class}(i)) + FP(\text{Class}(i)) + FN(\text{Class}(i))} \right) * 100 \text{ -----(15)}$$

Based on the similarly approach of the Table 3, the confusion matrix of the EMSD+FL-HKSVM method is illustrated in Table 5.

Table.5: Confusion matrix of proposed method (EMSD+FL-HKSVM)

Class predicted	Ground Truth Class (Assigned by Pathologist)				
	Normal	Mild dysplasia	Moderate dysplasia	Severe dysplasia	Carcinoma in situ
Normal	195	12	13	12	8
Mild dysplasia	0	52	2	1	2
Moderate dysplasia	1	3	67	1	2
Severe dysplasia	2	1	2	83	1
Carcinoma in situ	1	1	2	1	87

In the Table 5, the row-wise elements correspond to the four categories and the column-wise elements correspond to the target class associated with that abnormal category. Hence, the number of images correctly classified (TP) under each category is determined by the diagonal elements of the matrix. The row-wise summation of elements for each category other than the diagonal elements corresponds to the ‘FN’ of that category. The column-wise summation of elements for each category other than the diagonal element corresponds to the ‘FP’ of that category. Similarly, ‘TN’ of the specific category is determined by summing the elements of the matrix other than the elements in the corresponding row and column of the specific category.

For example, among the 57 Mild dysplasia testing images, 52 images have been successfully classified (TP) and the remaining 5 images (first row-wise summation) have been misclassified to any of the non- Mild dysplasia categories (FN). Similarly 17 images (first column-wise summation) from the other three categories (non- Mild dysplasia) have been misclassified as Mild dysplasia category (FP). In the Table 6, the classification accuracy of EMSD with FL-HKSVM in class 1 (Normal) is 91.12%, class 2 (Mild dysplasia) type cancer is 96.01%, class 3(Moderate dysplasia) type cancer is 95.29%, class 4(Severe dysplasia) type cancer is 96.20% and class 5(carcinoma in situ) type cancer is 96.74%. then the overall sensitivity, specificity & accuracy of this ERSTCM + FL-HKSVM is 87.68%, 96.92% & 95.07%. The performance of overall calculation of sensitivity, specificity and accuracy by using this Eq.13 -15.

The miss classification rate of class 1(Normal) non cancer type, Class 2 (Mild dysplasia), class 3 (Moderate dysplasia) & class 4 (Severe dysplasia) of cancer type is high compared to the other class of class 5 (Carcinoma in Situ).

Table.6: Performance measure of EMSD with FL-HKSVM

Class predicted	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)
Normal	195	308	4	45	81.25	98.72	91.12
Mild dysplasia	52	478	17	5	91.23	96.57	96.01
Moderate dysplasia	67	459	19	7	90.54	96.03	95.29
Severe dysplasia	83	448	15	6	93.26	96.76	96.20
Carcinoma in situ	87	447	13	5	94.57	97.17	96.74
Overall Sensitivity, Specificity & Accuracy Results					87.68	96.92	95.07

For comparative analysis, the proposed cervical cancer classification system (EMSD + FL-HKSVM) is compared to other classification system (ERSTCM + FL-HKSVM). The overall classification performance measured parameters of sensitivity, specificity & accuracy of the proposed method is 87.68%, 96.92% & 95.07% and this method will be provide the better results than the existing method ERSTCM + FL-HKSVM results in terms of 83.15%, 95.79% & 93.26%. The overall classification results of sensitivity, specificity and accuracy of existing and proposed method are shown in Figure 8.

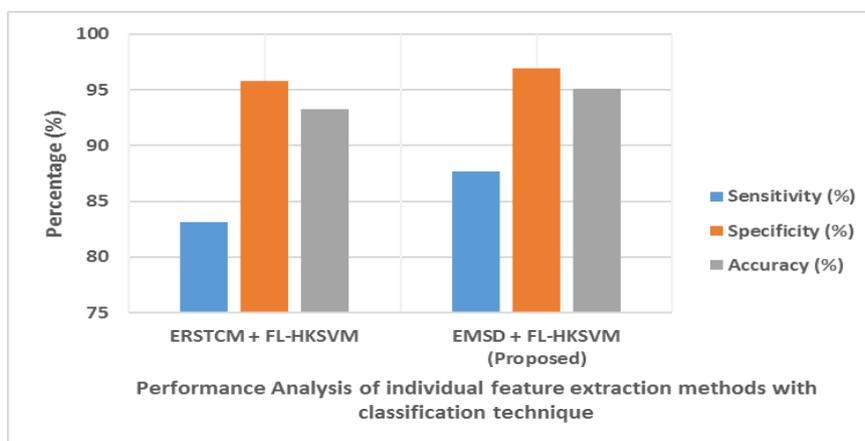


Figure.8: Comparison Results of the Proposed & Existing System in individual feature Extraction methods with Classification Technique.

4.2.2. Individual and Combining individual features & Classification combination comparison:

Based on the similarly approach of the Table.3, the confusion matrix of the CFE+FL-HKSVM method is illustrated in Table 7.

Table.7: Confusion matrix of proposed method (CFE+FL-HKSVM)

Class predicted	Ground Truth Class (Assigned by Pathologist)				
	Normal	Mild dysplasia	Moderate dysplasia	Severe dysplasia	Carcinoma in situ
Normal	224	5	4	3	4
Mild dysplasia	0	56	0	1	0
Moderate dysplasia	2	1	70	0	1
Severe dysplasia	0	1	0	87	1
Carcinoma in situ	0	0	1	0	91

In the Table 7, the row-wise elements correspond to the four categories and the column-wise elements correspond to the target class associated with that abnormal category. Hence, the number of images correctly classified (TP) under each category is determined by the diagonal elements of the matrix. The row-wise summation of elements for each category other than the diagonal elements corresponds to the ‘FN’ of that category. The column-wise summation of elements for each category other than the diagonal element corresponds to the ‘FP’ of that category. Similarly, ‘TN’ of the specific category is determined by summing the elements of the matrix other than the elements in the corresponding row and column of the specific category. For example, among the 57 Mild dysplasia testing images, 56 images have been successfully classified (TP) and the remaining 1 image (first row-wise summation) have been misclassified to any of the non- Mild dysplasia categories (FN). Similarly, 7 images (first column-wise summation) from the other three categories (non- Mild dysplasia) have been misclassified as Mild dysplasia category (FP). In the Table 8, the classification accuracy of CFE with FL-HKSVM in class 1 (Normal) is 96.74%, class 2 (Mild dysplasia) type cancer is 98.55%, class 3(Moderate dysplasia) type cancer is 98.37%, class 4(Severe dysplasia) type cancer is 98.91% and class 5(carcinoma in situ) type cancer is 98.73%. then the overall sensitivity, specificity & accuracy of this CFE + FL-HKSVM is 95.65%, 98.91% & 98.26%. The performance of overall calculation of sensitivity, specificity and accuracy by using this Eq.13-15. The miss classification rate of class 1(Normal) non cancer type, Class 2 (Mild dysplasia), class 3 (Moderate dysplasia) & class 5 (Carcinoma in Situ) of cancer type is high compared to the other class of class 4 (Severe dysplasia).

Table.8: Performance measure of CFE with FL-HKSVM

Class predicted	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)
Normal	224	310	2	16	93.33	99.36	96.74
Mild dysplasia	56	488	7	1	98.25	98.59	98.55
Moderate dysplasia	70	473	5	4	94.59	98.95	98.37
Severe dysplasia	87	459	4	2	97.75	99.14	98.91
carcinoma in situ	91	454	6	1	98.91	98.70	98.73
Overall Sensitivity, Specificity & Accuracy Results					95.65	98.91	98.26

For comparative analysis, the proposed cervical cancer classification system (CFE + FL-HKSVM) is compared to the other methods (ERSTCM + FL-HKSVM & EMSD + FL-HKSVM), the overall classification performance measured parameters of sensitivity, specificity & accuracy of the proposed CFE + FL-HKSVM method is 95.65%, 98.91% & 98.26% & and this method will be provide the better results than the existing methods of ERSTCM + FL-HKSVM & EMSD + FL-HKSVM Results. The overall classification results of sensitivity, specificity and accuracy of existing and proposed method are shown in Figure 9.

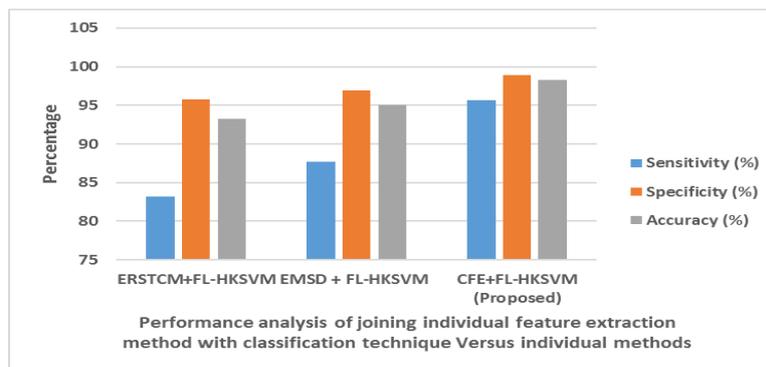


Figure.9. Comparison Results of the Proposed CFE + FL-HKSVM with Existing ERSTCM+FL-HKSVM, EMSD+FL-HKSVM Classification Technique

V. Conclusion

In this paper, a novel approach of individual and combining individual feature extraction method with classification technique developed for identify the class for the cervical cyto cell of the pap smear image. Two major contributions of this paper are feature extraction and classification. In feature extraction, we have taken the advantage of combining individual methods such as ERSTCM & EMSD texture features as one feature to assess the joint performance. In Classification, multiple kernels are combined and developed a Fuzzy Logic based hybrid kernel SVM classifier for improving the classification process. For comparative analysis, our proposed approach is compared with existing works in individual feature extraction with the classification method, the proposed EMSD+FL-HKSVM produce better result than ERSTCM+FL-HKSVM and combining individual feature extraction with the classification method, the proposed CFE+FL-HKSVM produce better result than ERSTCM+FL-HKSVM & EMSD+FL-HKSVM methods in terms of the statistical parameters results sensitivity, specificity & accuracy. Hence, finally our proposed individual and combining individual method are proved good at detecting anyone of the five types of the classes (normal, mild, moderate, severe & carcinoma in situ) of cervical cancer in the pap smear image.

References

- [1] S.Athinarayanan, Dr.M.V.Srinath & R.Kavitha “ Computer aided diagnosis for detection and stage identification of cervical cancer by using pap smear screening test images” *ictact journal of image and video processing*, ISSN: 0976-9102 vol. 6, no.4, May 2016. pp. 1244 – 1251.
- [2] S. Aswathy, M. A. Quereshi, B. Kurian, and K. Leelamoni, “Cervical cancer screening: current knowledge & practice among women in a rural population of Kerala, India,” *Indian Journal of Medical Research*, vol. 136, no. 2, pp. 205–210, 2012.
- [3] S.Athinarayanan, Dr.M.V.Srinath, “Classification of Cervical Cancer Cells in Pap Smea Screenig Test” *ictact journal of image and video processing*, ISSN: 0976-9102 vol. 6, no.4, May 2016. pp. 1234 – 1238.
- [4] B. L. Craine, E. R. Craine, J. R. Engel, and N. T. Wemple, “A clinical system for digital imaging Colposcopy,” in *Medical Imaging II*, vol. 0914 of *Proceedings of SPIE*, pp. 505–511, Newport Beach, Calif, USA, June 1998.
- [5] H. Doornewaard, Y. T. van der Schouw, Y. van der Graaf, A. B. Bos, and J. G. van den Tweel, “Observer variation in cytologic grading for cervical dysplasia of Papanicolaou smears with the PAPNET testing system,” *Cancer*, vol. 87, no. 4, pp. 178–183, 1999.
- [6] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, “Texture analysis of medical images,” *Clinical Radiology*, vol. 59, no. 12, pp. 1061–1069, 2004.
- [7] M. E. Plissiti, C. Nikou, and A. Charchanti, “Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 233–241, 2011.
- [8] P. Sobrevilla, E. Montseny, F. Vaschetto, and E. Lerma, “Fuzzy based analysis of microscopic color cervical pap smear images: nuclei detection,” *International Journal of Computational Intelligence and Applications*, vol. 9, no. 3, pp. 187–206, 2010.
- [9] C. Bergmeir, M. Garc’ia Silvente, and J. M. Ben’itez, “Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework,” *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 497–512, 2012.
- [10] J. Holmquist, E. Bengtsson, O. Eriksson, B. Nordin, and B. Stenkvist, “Computer analysis of cervical cells. Automatic feature extraction and classification,” *Journal of Histochemistry & Cytochemistry*, vol. 26, no. 11, pp. 1000–1017, 1978.
- [11] Y. Marinakis, G. Dounias, and J. Jantzen, “Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification,” *Computers in Biology and Medicine*, vol. 39, no. 1, pp. 69–78, 2009.
- [12] Y.-F. Chen, P.-C. Huang, K.-C. Lin et al., “Semi-automatic segmentation and classification of pap smear cells,” *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 1, pp. 94–108, 2014.
- [13] B. Julesz. Texton gradients: The texton theory revisited. *Biological Cybernetics*, 54(1986) 245-251.
- [14] Wells. W.M. W.E.L. Grimson. R. Kikinis and F.A.Jolez, 1996. Adaptive segmentation of MRI data. *IEEE Trans. Med.Imagin*, 15: 429-442.
- [15] Zijdenbos, A.P., R. Forghani and A.C. Evans, 2002. Automatic pipeline analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis. *IEEE. Trans. Med. Imaging* 21: 1280 – 1291.
- [16] Chen, L.; Chen, C.L.P.; Lu, M.: A multiple-kernel fuzzy C-means algorithm for image segmentation. *IEEE Trans. Syst.Man Cybern. B Cybern.* 41(5), 1263–1274 (2011).