Ethno Med, 12(2): 96-105 (2018)
DOI: 10.31901/24566772.2018/12.02.544

Review of Pap Smears Cell Segmentation and Classification Techniques for Cervical Cancer Analysis

N. Kumaresan and D. Somasundaram*

Department of Electronics and Communication Engineering, Anna University, Regional Centre, Coimbatore, Tamil Nadu, India
**E-mail: somgce@gmail.com*

KEYWORDS Cervical Cancer. Classification. Microscopic Image. Pap Smear Cells. Segmentation

ABSTRACT Cervical cancer is the second most common gynecologic cancer among women. Cervical cancer can be prevented completely if found at the initial stages. Pap smear test is a simple screening procedure for detecting the cervical cancer at the early stage. Pap smear image analysis is an open issue which does not show satisfactory results. The limitation of the analysis is due to the complexity of the cell structure. The smear image analysis struggles with issues like overlapping and folding of cells. Segmentation of individual cytoplasm and nuclei from the cluster of overlapped cervical cell helps in detecting the cancer. Appropriate classifiers are used to classify the smear image as normal/abnormal using the identified parameters. This paper provides collective reviews on cervical cancer analysis which includes various segmentation approaches for overlapping problems, parameter identification methodologies and classifiers used for smear image analysis.

INTRODUCTION

Cervical cancer is considered as the fourth most common cancer among women and the seventh most common cancer overall. Ninety percent of cervical cancer deaths occur in developing countries due to the lack of screening programs. Cervical cancer begins in the uterine cervix, which forms the lower part of the uterus. The uterus and the vagina (which is called as birth canal) are connected by cervix. There are two parts of cervix: (i) end cervix, (ii) ectocervix. End cervix is the part which is closest to the body of the uterus. Ectocervix is the part next to the vagina. There are two main cells covering the cervix: (i) squamous cell which covers the ectocervix and, (ii) glandular cells that covers the end cervix. There two cells meet at a point which is the transformation zone. The location of the transformation zone gets altered due to aging and after pregnancy. Cervical cancer mostly occurs in cells that are present in the transformation zone. The affected cells in this zone will not suddenly change as cancer cells. Initially the cells develop into pre-cancerous cells then it will change to a cancer cell. The pre-cancerous cell stages are described as cervical intraepithelial neoplasia (CIN), squamous intraepithelial lesion (SIL) and dysplasia. Pap test helps in detecting these changes at the early stage which can be treated to prevent cancer. Pap smear images are microscopic images. Pre-cancer/cancer cells are identified by the nature of the cell structured viewed under a light microscope. Squamous cell carcinoma and adenocarcinoma are the two major types of cervical cancer. Aden squamous carcinoma is another type of cancer which has the mixed features of squamous cell carcinomas and adenocarcinomas which is less common.

The Papanicolaou test which is abbreviated as Pap test also known as Pap smear/cervical smear/smear test. Pap test is a cervical screening procedure that helps in detecting the pre-cancerous and cancerous cells in the cervix. A regular Pap test helps in preventing the cervical cancer at the early stage.

Pap test is done by collecting the cells at the transformation zone. These cells are then examined under light microscope to find the abnormalities. The factor that affect the sensitivity of Pap test are: (i) number of sampled cells, (ii) overlapping among cells, (iii) poor contrast of cytoplasm, (iv) presence of mucus and, (v) blood and inflammatory cells. Automated cell deposition techniques are used which help in mono layer preparation, removal of blood, mucus and debris and help to get smear image with less overlap. These automated cell deposition techniques help in improving the sensitivity and specificity and ease in detecting, segmenting and classification of microscopic cells. Cell detection and segmentation are the major issues for automated analysis of cervical cytology specimen.

Recently, Bora et al. (2017) proposed an automated method to classify normal and abnormal cells. In this process Ripplet Type I transform, Histogram first order statistics and Gray Level Co-occurrence, Matrix have been used. Dey et al. (2017) discussed this similar problem for habitual smokers. Saha et al. (2017) proposed a method, which is circular shape constrained fuzzy clustering (CiscFC) for nucleus segmentation. This method provides precision of 0.96. Zhang et al. (2017), discussed a Graph-Based Segmentation method for abnormal Pap smear cells.

Research work on Pap smear images focus mainly on segmentation of overlapping cells, parameter identification and classification for detecting the presence of cervical cancer. This paper is a survey done on various segmentation approaches, parameters identified and various classifiers used for identifying the cancer. The paper is organized as Section I: Various segmentation approaches for segmenting the overlapping smear images, Section II: Parameters identification and classification, Section III: Comparative analysis, Section IV: Discussion and Section V: Conclusion.

METHODOLOGY

Profile Based Method

In this approach, Bengtsson's extracted features from profile of cell images. This algorithm used for both boundary and nuclei segmentation. Regions are evaluated based on the density profile of image. Difference chain code method is used for analysis. From the analysis, size, relative location of nuclei and overlapping position are identified. Overlaps in the cell image separated based on the concavity properties. This test applied more than 200 images and attained notable accuracy in segmentation.

Pixel Base Method

A fast boundary finding algorithm is discussed which works without threshold operation and without any interactive control. The procedure can be described as a hierarchical two-step algorithm. In the first step the image is divided into two disjunctive regions, one of them including the whole object of interest. In the second step the problem of boundary finding is suggested as a classification problem, which means that for any pixel a four-dimensional feature vector is computed which allows classification of pixels into contour elements and any other pixels.

Fourier Transform Method

This method extracted features of mean, variance and entropy. This feature varies from normal to abnormal cells. For classification fourfold validation method is used. For classification, Bayesian classifier, linear discriminant analysis (LDA), K-nearest neighbor (KNN) algorithms, artificial neural network (ANN), and support vector machine (SVM). The result from the support vector machine provides the best accuracy and the lowest false rate. It achieves more than ninety-two percent correct classification rate on a set of 276 cervical single-cell images containing 138 normal cells and138 abnormal cells.

Segmentation of Nucleus and Cytoplasm

Segmentation helps in extracting the specific region of interest (ROI) of any object/image. Pap smear cell has nucleus and cytoplasm. The location of nuclei is a challenging issue due to cell overlapping, inconsistent staining process and the presence of artefacts. Automated analysis of Pap smear image deals with the location of nuclei and extracting the boundary. Initially single Pap smear cell which has single nuclei surrounded by cytoplasm is taken for analysis by many researchers. Many segmentation methodologies are proposed for extracting the nuclei and cytoplasm regions. When real time Pap smear images are considered, the segmentation algorithm should consider many addition information about the overlaps and certain artefacts which can make the segmentation to fail at some point. More sophisticated algorithms should be considered for automated analysis of Pap smear to isolate the cell clusters and find the nuclei and cytoplasm regions.

Thresholding is the most commonly used process for segmenting the foreground from the background. The foreground is generally considered to be the object of interest. The segmentation of cervical cell is based on intensity histogram where the segmentation is done by characterizing the pixel. The nuclei are separated from the cytoplasm and the background as the nuclei has the darkest intensity pixels compared to another region. Threshold selection methods, segmentation of nuclei from a low-resolution Pap smear image (Poulsen and Pedron 1995), an optimization process for segmenting the nuclei region, these methods help in extracting the nuclei from the smear image. Classifying the nuclei as normal/abnormal using fuzzy radial basis function neural network (Kim et al. 2007) where the thresholding help in image binarization for nuclei detection. The nuclei extraction is done based on multiscale local adaptive threshold (Li and Najarian 2007) with shape stability. Thresholding done on mean filtered and rescaled pixels helped in detecting the nuclei region.

Nucleus Segmentation

The nucleus segmentation is the extraction of nuclei region (Plissiti et al. 2011) based on colour information of the image. The nuclei is extracted based on the colour characterization. The method is suitable for both single and overlapping cell clusters. The method undergoes preprocessing and yields binary mask. Finding the candidate nuclei centroid using morphological operators and refinement of candidate centroid by spatial characteristics and nucleus circumference is found. Finally, the nuclei is extracted by two steps. The empirical rule helps to find the distance between the centroid for reducing the false positive value. Fuzzy C means and SVM (Plissiti et al. 2011) classifiers are used to detect the nuclei centroid. The performance analysis of K means, spectral clustering, SVM using discriminative features are done based on minimum redundancy maximum relevance criteria. The approach helps in extracting the nuclei and its corresponding neighbourhood regions.

Fuzzy grayscale morphological operation (Kim et al. 2006) is used to detect the nucleus boundary using watershed transformation. Pixel classification techniques (Lezoray and Cardot 2002) detects nuclei markers and helps in reducing the over segmentation problems. K means and Bayesian classifiers are used to extract the nuclei. Nuclei detection is done based on fuzzy rules and pixel classification. Segmentation of nuclei is based on colour information using fuzzy system and expert knowledge (Vaschetto et al. 2009).

Wu et al. (1998) proposed parametric fitting algorithm for segmentation with shape and region image information. The algorithm is related to template matching process. Parametric elliptical model is developed for nucleus shape detection and the obtained parameters are adjusted to fit in the nucleus region. Coarse and fine optimization search is done to find the optimal fit. The shape information is considered which reduces the number of parameter and helps in detecting the boundary of the nucleus.

A deformable template approach (Garrido and Perez de la Blanea 2000) for segmenting the nuclei region was developed. The approach includes location of nucleus, elliptical approximation of nuclei boundary and refinement of boundary using deforming models. Generalized Hough transform is used to detect the nuclei location. Deformable template approach is used to extract the exact nucleus boundary. Active contour model (Bam ford and Lovell 1998) is used to extract the nuclei boundary. The contour model is initialized by search space construction by which the possible points to locate the nuclei boundary is done. The search space is done by Viterbi search based dual active contour algorithm.

Plissiti (2011) presented a method to segment the nuclei region which is a combination of physical based model with active shape model (ASM). The training is done based on the shape of the nuclei represented by vibration of a spring mass system. The deformable model which is developed based on physical model is used for training. The deformable model with shape model is used for training and helps to identify the nuclei region. The morphological operation is performed to define the candidate nuclei centroid. GVF is applied for segmenting the candidate nuclei boundary. Fuzzy C means clustering help in the final decision for determining the candidate nuclei with the GVF snake which gives a better accuracy in nuclei detection. The shape of the nuclei is segmented using elastic segmentation algorithm (Bergmeir et al. 2012). The noise in the image is removed using mean shift and median filtering and the edges of the nucleus region is detected by canny edge detector. The candidate nuclei is generally elliptical shape which is initially identified using randomized Hough transform and then processed using level set algorithm. A database of 207 images are taken for analysis and the algorithm gives promising results compared with another algorithm.

Segmentation of Cytoplasm and Nucleus

A pixel classification method (Lassouaoui and Hamami 2003) was developed based on multi fractal algorithm. The algorithm is used to classify the pixels of the image into background, cytoplasm or nucleus. Genetic algorithm is an optimization technique used to reclassify the pixels of the image as the above three categories. The optimization helps in preserving the detected area.

The nucleus, cytoplasm and background is detected using predefined criterion function based on statistical object structure. The local spatial likelihood functions with prior probabilities are considered as criterion function. K mean algorithm with iterative procedure helps in grouping the pixel into any three categories. Modified seed based region grouping is used for cervical cell segmentation. K mean clustering algorithm for classifying the pixel into three region cytoplasm, nucleus and background. Seed pixel location and calculation of moments are used to relate the pixel to any of the three regions.

A geometric active contour without re-initialization (GACWR) (Harandi et al. 2010) helps in classifying the image into three region free lying cells, connected cells and irrelevant objects. The original image is divided into sub images. Binary mask is developed for each sub-images which helps to remove the irrelevant objects. An initial contour is created using GACWR for each cell. This method helps in segmenting the cytoplasm and nuclei boundaries.

In linear contrast enhancement method is used for segmenting the nucleus and cytoplasm. The noise is removed using median filter and preserves the edge and boundary of the nucleus and cytoplasm from the background. The nucleus contour is extracted by maximum grey gradient difference method. Radiating gradient vector flow (RGVF) snake algorithm for extracting the nucleus and cytoplasm. K means clustering algorithm is used to extract the region of nucleus, cytoplasm and background separately. RGVF snake is then applied which include edge map computation and stack based refinement for boundary detection of nucleus and cytoplasm. Herlev dataset with 917 images is used for study.

A gradient direction edge enhancement based contour (GDEEBC) detector method is used for segmenting nucleus and cytoplasm. Trim meaning filter is used to remove impulse and Gaussian noise in cervical images and preserve the edges in the image. The nucleus and cytoplasm regions are segmented using bi group enhancer. The gradient direction enhancer is used to suppress the noise gradient and helps to improve the brightness of the nucleus and cytoplasm contour. The algorithm is also applicable for other types of images. FCM clustering algorithm is used for segmenting nucleus and cytoplasm. Two types of random number generators are considered as the membership matrix. Few parameters are measured. They are partition coefficient, partition entropy, compactness and separation function. The segmentation is performed based on the shape and the calculated parameters.

Tsai et al. (2008) presented the work to detect the nucleus and cytoplasm using CNC (Cytoplasm and nucleus detector). The CNC uses median filter to remove the noise. The noise is suppressed and the object boundary is brightened by bi group enhancer. The cytoplasm is differentiated from the background using K means clustering and maximal colour difference method is employed for extracting the contour of the nucleus. An unsupervised approach for segmenting the cervical cell is proposed. The method uses Herlev and Hacettepe dataset. The segmentation involves two stages. The first segmentation stage uses homogeneity and circularity property to partition the image into sub region called as multi scale hierarchical segmentation. The cells are separated from background using thresholding based on morphological operation. The second stage is about classifying the region into nucleus and cytoplasm. The binary classifier is used for this purpose. The other classifiers used for the classification are Bayesian, decision tree, support vector machine.

The cytoplasm and nucleus is segmented from the clusters of overlapping cell using level set optimization approach (Lu et al. 2013). The researchers have developed Adelaide Pap smear dataset with real free lying cells and synthetic overlapping cell images. The initial step is to extract the free lying cells and their nucleus from the background. The level set optimization algorithm is used to segment the nucleus and cytoplasm. The optimization algorithm helps in calculating the area and length of each cell, the amount of overlapping, and the grey values inside the overlapping regions. Polar gradient vector field (PGVF) snake algorithm (Chuanyun et al. 2013) helps in segmenting the nucleus and cytoplasm boundary. The dataset used for the analysis is Herlev dataset. The input image is taken, the Descartes coordinates of the image is converted to polar coordinates. The boundary of the cell is obtained using the Sobal operator and the interference elements present on the boundary is identified by sand inhibition method. Now the polar coordinates of edge map are again converted back to Descartes coordinates. Gradient vector flow snake algorithm is then applied to segment the nucleus and cytoplasm regions.

The patch based fuzzy C-means clustering method (Chankong et al. 2014) helps in segmenting the nucleus and cytoplasm. The dataset used for the analysis is Herlev dataset. The input image is converted to grayscale image. The median filter is applied on the image to remove the noise and smoothen the image. The FCM clustering is now applied on the smoothen image to segment the image into three regions, they are nucleus, cytoplasm and background. FCM helps in sorting the patches by its center value. From this the patches with darker value is considered as nucleus and if the patch value lies between nucleus and cytoplasm threshold is the cytoplasm and other values are the background.

The graph cut approach (Zhang et al. 2014) is presented for segmenting the cervical cell based on global and local schemes. Multi-way graph cut approach is used to segment the cytoplasm region from background and considered as the global scheme. The local adaptive graph cut is used as local scheme for segmenting the nucleus region. The touching nuclei are separated by identifying the concave points.

Segmentation of Overlapping and Touching Nuclei and Cytoplasm

An overlap detection algorithm (Bengtsson et al. 1981) for segmenting the overlapping nuclei base on nucleus contour and density profile is proposed. The nuclei contour is extracted by thresholding. The concavity nature along the contour is identified and features are extracted for verifying the nuclei as single or overlapping nuclei. The overlapping nuclei region is extracted based on distance transform and topographic surface determined by distance transform. The parameters of the nucleus are determined using EM algorithm. The overlapping nucleus is separated using separation lines defined by a criterion function. Constrained ellipse fitting technique is used to reconstruct each occluded nucleus area.

Segmentation of overlapped nuclei is considered as optimization problem and a marker based H-minima transform (Jung and Kim 2010) is introduced which help in obtaining the distance map to segment the nuclei region. Segmentation of cervical cells is done using local and global scheme based on graph cut segmentation (Zhang et al. 2014). Multi way graph cut is performed with global scheme on a^* channel enhanced images for segmenting the cytoplasm region. The approach is effective when the image histogram is a non-bimodal distribution. Local scheme based graph cut approach is useful to segment the nuclei region especially when they are abnormal which needs the information on intensity, texture, boundary of the nuclei region. The touching nuclei is segmented based on concavity points.

Classification Techniques

The extracted features are given as inputs to the classifier. The classifier gives the information related to the cervical cell. The classifier result helps in classifying the cell into normal and abnormal cell.

Oliver et al. (1979) presented the work to classify the cervical cell as normal/abnormal by Bayes rule which is based on two decision rules. Texture features are extracted using parametric Bhattacharya distance formulation. These features are considered as input to the classifier. The classifier is used for classifying the Pap smear samples using Bayes classifier. The Bayesian classifier is used by other researchers (Chankong et al. 2009) for classifying the cervical cell.

The neural network based classifier is used for classifying Pap smear cell, hierarchical hybrid multilayer perceptron network, where the cervical cell is classified into three types normal, low and high grade squamous intraepithelial lesion. The mentioned classifier is compared with RBF, MLP and HMLP networks with 1 to 50 varying hidden nodes. The work is done with 550 annotated Pap smear images. Gallegos-Funes et al. (2009) proposed rank M type radial basis function neural network for classifying the Pap smear image. The parameters are estimated using median M type estimator. The sensitivity and specificity are calculated and compared with simple median RBF neural network.

A feed forward neural network with back propagation as training method helps to classify the Pap smear cell. The training method employs variable learning rate and network uses single hidden layer for classification. Multilayer sigmoid neural network is used for classifying the Pap smear images. The network uses Levanberg Marquardt back propagation training algorithm which is comparatively better than fuzzy classifiers. Fuzzy RBF classifier (Kim et al. 2007) is used for classifying the nuclei as normal and abnormal. Fuzzy adaptive resonance theory algorithm (Kim et al. 2006) is used for classifying the Pap smear image into three category normal, abnormal and cancerous cells.

The Pap smear cell image is classified into normal and abnormal using support vector machine (SVM) (Chankong et al. 2009). SVM (Zhang and Liu 2004) based classifier is used for classifying multispectral Pap smear images. A system is used with dual wavelength method for separating cytoplasm and nuclei. The method involves four steps segmentation, feature extraction, cell and smear classification. Linear discrimination analysis is used for classification. The nucleus detection is done by nucleus definition failure logic based on the regions like surface cell, intermediate cell, parabasal cell, cylinder cell, dyscaryotic surface, intermediate and parabasal cell. The discriminant analysis helps in categorizing the cell into surface and intermediate as one group, parabasal and cylinder cell into the second group and the third group includes malignant cells. The smear classification helps in classifying the single cell as normal by counting the number of atypical cell and others as abnormal. The limitation of the approach is the wavelength selection of cell based on thresholding.

The classification of Pap smear image data analysis is based on numerical measure (Jantzen and Dounias 2006). The data analysis follows manual classification of cell classes, feature extraction and cell classification. The sample images are initially categorized into seven groups/classes: superficial squamous epithelial, columnar epithelial, dysplasia, mild, moderate and severe squamous non- keratinizing, and squamous cell carcinoma. The superficial squamous, columnar and dysplasia are normal cells and others are abnormal cells. The classifier is trained with twenty features like nucleus and cytoplasm area, nucleus to cytoplasm ratio, brightness of nucleus and cytoplasm region, shortest and longest diameter of nucleus and cytoplasm, elongation of nucleus and cytoplasm, roundness of nucleus and cytoplasm, nucleus position, maximum and minima nucleus and cytoplasm. The degree of overlap is obtained by finding the distance between each class center. This simple minimum distance classifier is compared with other classifiers. The classifier helps in classifying the cell into mild, moderate and severe dysplasia.

Genctav et al. (2012) proposed an unsupervised approach for segmenting nucleus and cytoplasm. Two stages of segmentation is carried out. Multi-scale hierarchical segmentation is used to partition the image into various sub regions. The partition is done based on homogeneity and circularity. The foreground is separated from background using thresholding followed with morphological operation. The second step is to discriminate the regions into nucleus and cytoplasm which is done using binary classifier. The other classifier used is Bayesian classifier, decision tree, SVM, combination of all three-classifier using sum and product of posterior probability. Adaptive network based fuzzy inference system is compared with various clustering based classifier like fuzzy C means, nearest neighbourhood and Gustafson Kessel classifiers. The classification performance can be improved by feature reduction obtained using simple simulated annealing operation. The classification of cervical cell is done using supervised and unsupervised learning based on clustering algorithm. Three different classifiers are used for the work are hard and fuzzy C means, Gustafson Kessel with K fold. Comparing the three classifiers supervised Gustafson Kessel gives better performance with simulated annealing process.

The feed forward neural network is trained with Leven Berg Marquardt with adaptive momentum (LMAM) and Optimized Leven Berg Marquardt with adaptive momentum (OLMAM) (Ampazis et al. 2004). LMAM and OLMAM is the two second order algorithm applied on Herlev dataset for classification. The performance of the classifiers is done using 10-fold cross validation and then it is compared with intelligence approaches. The discrimination of classifiers as two and seven class is done using inductive and transductive classifiers (Norup 2005). The classifier used are inductive and transductive model classifier, linear least square network, nearest, K nearest and weighted K nearest neighbour classifiers. Neuro fuzzy inference method for transductive reasoning is done and the results obtained are not promising compared to the work previously done. The classification was done with 20 fixed features without any feature reduction algorithm.

Marinakis and Dounias (2006) presented various nearest neighbourhood classifiers by features obtained by Tabu search algorithm and the performance of the classifiers are analysed. The various classifier used are 1-nearest neighbor, 3-nearest neighbor, 5-nearest neighbor, weighted 3 and 5 nearest neighbour classifiers. The k fold cross validation is used for analysing the performance. The classification of cervical cell into normal or abnormal is done using nucleus area (Plissiti and Nikou 2012). Two unsupervised classifiers namely spectral clustering and fuzzy C means are considered for classification. Non-linear techniques like kernel PCA, Laplacian Eigenmaps, locally linear embedding techniques and Isomap are used for feature reduction. The analysis reported that the considered nucleus area as single feature gives a better result than the combined 20 features of both nucleus and cytoplasm region.

Chankong et al. (2014) presented artificial neural network for classifying the cervical cell. The two class and seven class problem is addressed. The classification is done using nine features considering both nucleus and cytoplasm regions. Six features are related to nucleus and remaining features are related to cytoplasm. The k fold cross validation is considered for performance analysis. Finally, the artificial neural network is compared with other classifiers like K-NN, Bayesian, SVM and LDA. ANN with 5 nucleus based feature with ERUDIT achieved 96.2 percent and 4 and 2 class achieved 97.83 percent accuracy. The accuracy was 93.78 percent and 99.27 percent for 9 cell based features for 7 and 2 class problem. For LCH dataset the accuracy 9 cell based feature was 95.00 percent and 97.00 percent for 4 and 2 class problem. Mariarputham et al. (2015) proposed a classification system for seven class problem called as nominated texture based cervical cancer classification system. The features used for the classification are relative size of nucleus and cytoplasm, relative displacement of nucleus, tamura features, grey level co-occurrence matrix, dynamic range of nucleus, first four moments of intensities of nucleus, edge oriented histogram. The classifiers used for the analysis are SVM and neural network. Various combination of the features is considered for the analysis. The k fold cross validation is used for the performance validation. From the analysis the linear kernel SVM is considered better.

Sajeena and Jereesh (2015) proposed an automated method for cervical cell segmentation and classification. Radiating gradient vector flow (RGVF) snake is used to segment the cervical cell into nucleus, cytoplasm and background. The dataset used for the analysis is Herlev dataset. Seven cell classes of the dataset are used for classification. The features of cell and nucleus are considered. The classification of cell into normal and abnormal is done using SVM, ANN and Euclidean distance based system. Guven and Cengizler (2014) proposed data cluster analysis based classification of overlapping nuclei of Pap smear images. An unsupervised clustering approach is used for finding nuclei region. The morphological operations are performed for refining the output obtained using clustering method. The clustering helps into segment the nuclei region into overlapping and non-overlapping. The features extracted are two local minima and three shape based features for identifying the overlapping regions. The classification is performed based on feature score, precision, recall values. From the analysis the unsupervised approach with extracted features helps in detecting the over lapping nuclei.

RESULTS AND DISCUSSION

This study consolidates the various methods of segmentation, classification and most probable features to identify the Pap smear cells in semi-automated and automated systems. Table 1 shows the various segmentation methods

Table 1: Comparison between various segmentation methods

Name of the author	*Proposed method*	*Year*	*Features*	*Classification method*	*Accuracy*	*Advantages and disadvantages*
Chankong et al.	Fourier Transform	2009	Mean, variance, and entropy obtained from the frequency components	Bayesian classifier,	88.97	Avoid the problem of cell and nucleus segmentation, This method is manual segmentation
				linear discriminant analysis LDA),	88.60	
				K-nearest neighbor (KNN)	91.18	
				Artificial neural network (ANN)	91.00	
				support vector machine (SVM	92.65	
Chankong et al.	Patch-based fuzzy C-means clustering	2014	Precision, Recall, ZSI, Hausdroff distance	Bayesian classifier,	87.50	Automatic segmentation method. Applied for 3 types of data sets.
				linear discriminant analysis LDA),	93.84	
				K-nearest neighbor (KNN)	96.38	
				Artificial neural network (ANN)	97.83	
				support vector machine (SVM	96.74	
Plissiti et al.	Minimal-Redundancy–Maximal-Relevance criterion	2011	Shape, texture and intensity features	K-means	84.09	Overcome the problem of the detection of nuclei locations.
				Spectral clustering	82.46	
				SVM	82.52	
Ling et al.	Radiating GVF Snake	2012	Edge map computation,	Stack-based refinement	91.00	Automatic segmentation method. Elevated for one types of data sets.
Patrik et al.	Edge based approach using customized Laplacian of Gaussian (LoG) filter	2013	Laplacian	Gaussian operators	94.01	Fast and reliable in segmentation
Ramos et al.	Raman spectroscopy based analysis	2015	Structural analysis	-	-	Well-accepted best-practice method Definitive diagnosis method.
Patrik et al.	Sequential classification scheme	2013	Shape, elliptical deviation, Texture analysis	Mean intensity deviation	90.00	Clear the Debaris in image.
Kuan et al.	Automation-assisted reading (AAR) techniques	2014	Intensity, texture	Boundary and region information	93.00	Improve the sensitivity of AAR in screening forcervical cancer

and its performance comparison. Spectral clustering method provides the least accuracy which is 82.46 percentage. Support Vector Machine (SVM) and Artificial Neural Network (ANN) networks provide better results of ninety-six and ninety-seven percentage of accuracy in the classification process.

More over the effectiveness of the segmentation and classification of pap smear images is depends on the feature which is used for segmentation. In this study shows that, texture and shape based segmentation method provides better results. Filtering and clustering based methods like Laplacian, fuzzy clustering methods provides better results. Segmentation is more accurate using Snake models in Pap smear images. Still these snake models not suitable for automated analysis due to the segmentation time constraints and reliability. This study provides the key values to researchers to carry out their research in the updated models like CNN architectures, deep neural networks in various implementation platforms.

CONCLUSION

The automated analysis of Pap smear image is the challenging issue in computer vision and scientific filed. The segmentation of nucleus, cytoplasm and background regions are challenging task in cervical cancer analysis. The feature extraction is the next step after segmentation. Feature selection is the next problem which should be addressed which is given as the input for the classifiers. Identifying the classifiers for classification is also one of the important tasks for automated analysis of Pap smear images. The paper gives a review report for automated analysis with various methods adopted for nucleus and cytoplasm segmentation, the features used for classification and the various classifiers like LDA, KNN, SVM, PCA, etc., for cervical cell study are also discussed.

RECOMMENDATIONS

In Pap smear cell analysis, feature selection and classification techniques plays a major role. Selected method need to be fulfilled for various resolution of Pap smear images. SVM and neural networks provides better results and suitable for this analysis. Further, different kind of Convolutional Neural Networks architectures and deep neural networks with large number of datasets will be a better alternate for segmentation and classification of Pap smear images. This method will overcome the problems in classification of Pap smear images in various resolutions.

REFERENCES

Ampazis N, Dounias G, Jantzen J 2004. Pap-smear classification using efficient second order neural network training algorithms. *Methods and Applications of Artificial Intelligence,* 230-245.

Bamford P, Lovell B 1996. A Water Immersion Algorithm for Cytological Image Segmentation. In: *Proceedings of APRS Image Segmentation Workshop,* pp. 75-79.

Bergmeir C, García Silvente M, Benítez JM 2012. 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,* 107(3): 497-512.

Bengtsson E, Eriksson O, Holmquist J, Jarkrans T, Nordin B, Stenkvist B 1981. Segmentation of cervical cells: Detection of overlapping cell nuclei. *Computer Graph Image Process,* 16: 382-394.

Bora K, Chowdhury M, Mahanta LB, Kundu MK, Das AK 2017. Automated classification of Pap smear images to detect cervical dysplasia. *Computer Methods and Programs in Biomedicine,* 38: 31-47.

Chankong T, Theera-Umpon N, Auephanwiriyakul S 2009. Cervical Cell Classification Using Fourier *Transforms. Proceedings of the 13th International Conference on Biomedical Engineering, Part of the IFMBE Proceedings Book Series (IFMBE),* 23: 476–480.

Chankong T, Theera-Umpon N, Auephanwiriyakul S 2014. Automatic cervical cell segmentation and classification in Pap smears. *Computer Methods Programs Biomedicine,* 113(2): 539-556.

Chuanyun X, Yang Z, Sen W 2013. Cell segmentation in cervical smear images using polar coordinates GVF snake with radiating edge map. *Journal of Multimedia,* 8(3): 213-219.

Dey S, Sarkar R, Chatterjee K, Datta P, Barui A, Maity SP 2017. Pre-cancer risk assessment in habitual smokers from DIC images of oral exfoliative cells using active contour and SVM analysis. *Tissue and Cell,* 49(2 Pt B): 296-306.

Garrido A, Perez de la Blanca N 2000. Applying deformable templates for cell image segmentation. *Pattern Recognition,* 33(5): 821-832.

Gallegos-Funes FJ, Gomez-Mayorga ME, Lopez-Bonilla JL, Cruz-Santiago R 2009. Rank M-type radial basis function (RMRBF) neural network for Pap smear microscopic image classification. *Apeiron,* 16(4): 542-554.

Genctav A, Selim A, Sevgen O 2012. Unsupervised segmentation and classification of cervical cell images. *Pattern Recognition,* 45(12): 4151-4168.

Guven M, Cengizler C 2014. Data cluster analysis-based classification of overlapping nuclei in Pap smear samples. *BioMedical Engineering OnLine,* 13: 159.

Harandi NM, Sadri S, Moghaddam NA 2010. An automated method for segmentation of epithelial cervical cells in images of ThinPrep. *Journal of Medical Systems,* 34(6): 1043-1058.

Huang PW, Lai YH 2010. Effective segmentation and classification for HCC biopsy images. *Pattern Recognition,* 43(4): 1550-1563.

Jantzen J, Dounias G 2006. Analysis of Pap-smear Image Data. *Proceedings of the 2nd Annual Symposium on Nature-inspired Smart Information Systems,* 29 November-1 December, Austria.

Jung C, Kim C 2010. Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization. *IEEE Trans Biomed Eng,* 57(10): 2600-2604.

Kim JJ, Goldie SJ 2008. Health and economic implications of HPV vaccination in the United States. *N Engl J Med,* 359(8): 821-832.

Kim JJ, Kuntz KM, Stout NK, Mahmud S, Villa LL, Franco EL, Goldie SJ 2007. Multi-parameter calibration of a natural history model of cervical cancer. *Am J Epidemiol,* 166(2): 137-150.

Kim KB, Kim S, Sim KB 2006. Nucleus classification and recognition of uterine cervical Pap-smears using fuzzy ART algorithm. *Lect Notes Comput Sci,* 4247: 560-567.

Kim KB, Song DH, Woo YW 2007. Nucleus segmentation and recognition of uterine cervical Pap-smears. *Lect Notes Comput Sci,* 4482: 153-160.

Kuan Li, Zhi Lu, Wenyin Liu, Jianping Yin 2012. Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake. *Pattern Recognition,* 45(4): 255-1264.

Lassouaoui N, Hamami L 2003. Genetic Algorithms and Multifractal Segmentation of Cervical Cell Images. *7th International Symposium on Signal Processing and Its Applications,* 1-4 July, Paris, France, 2: 1-4.

Lezoray O, Cardot H 2002. Cooperation of color pixel classification schemes and color watershed: A study for microscopic images. *IEEE Transactions on Image Processing,* 11(7): 783-789.

Li Z, Najarian K 2001. Automated Classification of Pap Smear Tests using Neural Networks. *Proceedings of the International Joint Conference on Neural Networks,* 15 July-19 July, Washington, DC, United States, 4: 2899-2901.

Ling Zhang, Hui Kong, Chien Ting Chin, Shaoxiong Liu, Zhi Chen, Tianfu Wang, Sping Chen 2014. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts. *Computerized Medical Imaging and Graphics,* 38(5): 369-380.

Lu Z, Carneiro G, Bradley AP 2013. Automated Nucleus and Cytoplasm Segmentation of Overlapping Cervical Cells. *Proceedings of Medical Image Computing and Computer-Assisted Intervention–MICCAI,* 22-26 September, Nagoya, Japan, pp. 452-460.

Mariarputham E, Jaya S, Stephen A 2015. Nominated texture based cervical cancer classification. *Computational and Mathematical Methods in Medicine,* 1-10

Marinakis Y, Dounias G 2006. Nearest Neighbour Based Pap-smear Cell Classification Using Tabu Search for Feature Selection. *Proceedings of the NISIS-2006,* 29 November-1 December, Tenerife, Spain, pp. 25-34.

Norup J 2005. *Classification of Pap-smear Data by Transductive Neuro-Fuzzy Methods.* Master Thesis. Denmark: Technical University of Denmark, Oersted-DTU, Automation.

Oliver LH, Poulsen RS, Toussaint GT, Louis C 1979. Classification of atypical cells in the automatic cytoscreening for cervical cancer. *Pattern Recognition,* 11: 205-212.

Patrik Malm, Balakrishnan, Byju N, Sujathan, Vilayil K, Kumar Rajesh, Bengtsson Ewert 2013. Debris removal in Pap-smear images. *Computer Methods and Programs in Biomedicine,* 111(1): 128-138.

Plissiti ME, Nikou C, Charchanti A 2011. Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering. *IEEE Transaction Information Technology on Biomedical,* 15(2): 233–241.

Plissiti ME, Nikou C 2012. Overlapping cell nuclei segmentation using a spatially adaptive active physical model. *IEEE Transaction on Image Processing,* 21(11): 4568-4580.

Poulsen RS, Pedron I 1995. Region of interest finding in reduced resolution colour imagery application to cancer cell detection. *Pattern Recognition,* 28(11): 1645-1655.

Ramos Inxeas Raquel Martins, Alison Malkin, Lyng Fiona Mary, 2015. Current advances in the application of Raman spectroscopy for molecular diagnosis of cervical cancer. *BioMedical Research,* (1): 1-9.

Saha R, Bajgera M, Lee G 2017. Circular shape constrained fuzzy clustering (CiscFC) for nucleus segmentation in Pap smear images. *Computers in Biology and Medicine,* 85: 13-23.

Sajeena TA, Jereesh AS 2015. Cervical cancer detection through automatic segmentation and classification of Pap smear cells. *Int J Appl Eng Res,* 10(18): 39078-39084.

Tsai MH, ChanYK, Lin ZZ, Yang-Mao SF, Huang PC 2008. Nucleus and cytoplast contour detector of cervical smear image. *Pattern Recogn Lett,* 29(9): 1441-1453.

Vaschetto F, Montseny E, Sobrevilla P, Lerma E 2009. THREECOND: An Automated and Unsupervised Three Colour Fuzzy-Based Algorithm for Detecting Nuclei in Cervical Pap smear Images. In: *Proceedings of the 9th International Conference on Intelligent Systems Design and Applications,* pp. 1359-1364.

World Health Organization 2006. *Comprehensive Cervical Cancer Control: A Guide to Essential Practice.* WHO.

Wu HS, Gil J, Barba J 1998. Optimal segmentation of cell images. *IEEE Proceedings Vision, Image and Signal Processing,* 145(1): 50-56.

Zhang L, Hui K, Shaoxiong L, Tianfu W, Siping C, Milan S 2017. Graph-based segmentation of abnormal nuclei in cervical cytology. *Computerized Medical Imaging and Graphics,* 56: 38-48.

Zhang L, Kong H, Chin CT, Liu S, Chen Z, Wang T, Chen S 2014. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts. *Computer Medical Imaging Graphics,* 38(5): 369-380.

Zhang J, Liu Y 2004. Cervical Cancer Detection Using SVM Based Feature Screening. In: *Proceeding of the 7th International Conference on Medical Image Computing and Computer Aided Intervention (MICCAI'04),* Saint-Malo, France, pp. 873–880.

Paper received for publication on July 2017
Paper accepted for publicaiton on November 2017