Classification of Microscopic Cervical Cancer Images using Regional Features and HSI Model

Robert P, Celine Kavida A

Abstract: ***The main purpose of this paper is to classify the microscopic cervical images in order to identify the true impact of cancer that helps the patient to be treated properly. The Pap smear test is most efficient medical test, but it generates problem at the time of interpretation under the microscope. In order to unravel this drawback, automatic cancer detection is developed. This detection process includes few techniques of the image processing such as segmentation, and enhanced SVM classification algorithm. The final outcome of this proposed technique is compared to previous classification techniques such as ANN (Artificial Neural Network), KNN (K-Nearest Neighbor). The proposed algorithm is found to yield a good result from the experimental results & performance evaluation.***

Keywords: Cervical Cancer, Microscopic Images, Classification, CIN

I. INTRODUCTION

Cervical cancer is the most common cancer in developing countries among women under 35. Cervical cancer is a disease caused by Human Papilloma virus. It may be prevented and treated early by means of smear test. Pap test, a colposcopy is widely used to verify the cervix and the vagina [1]. The test identifies the irregular cells in the cervix and they are classified from normal to abnormal. Many research papers on segmentation, classification over the past 35 years, have been published [2]. A segmentation method usually deals with localization of nucleus. Classification methods works only with already separated cervix cells, ignoring the rest of the image. Early sexual contact, multiple sexual partners and pills for prevention of conception are the different causes for cervical cancer.

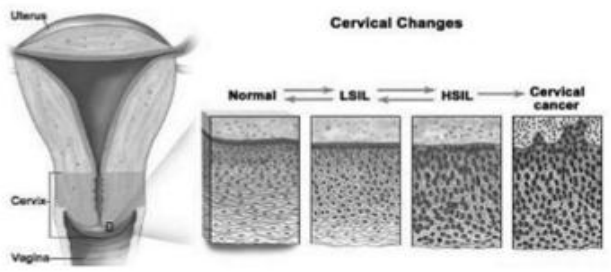


Fig. 1 Squamous Cell Dysplasia

Revised Manuscript Received on May 22, 2019.

Robert P, Research Scholar, Department of Information Technology, Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, Tamil Nadu, India

Celine Kavida A, Associate Professor, Department of Physics, Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, Tamil Nadu, India

Pap test has several disadvantages including: expertise dependent, low sensitivity and the need for retesting. An effective segmentation & classification technique is needed for identifying the real stage of cervical cancer. For the manual classification of normal and abnormal cervical cells, the collected cervical cells are examined under a microscope. However, due to human errors in cell classification, this method suffers from high false positive rates. This method is very cost-effective and only 4 to 5 slides per day can be classified by a pathologist [3]. Due to the cytoplasm and nucleus present in the cell structure, it is difficult to perform the process at a faster rate. Any computer aided cervical cancer screening system includes two fundamental processes: segmentation and classification. Classification is used to classify cells into various stages [CIN 0, CIN 1, CIN 2, CIN 3].

II. RELATED WORK

Lots of research on the classification of cervical cells has been done. Classifying cells into normal or abnormal is known as two-class problem, and classifying cells into one of the seven individual classes is known as seven class problem. The nucleus region is represented with nine features whereas cytoplasm region is represented with eleven features.

M.K.Soumya et al. suggested a classification technique using magnetic resonance image for cervical cancer staging. The result was compared to non-linear SVM classification models based on the tumors' second-order texture as well as transforming characteristics [4]. The texture features and statistical features are playing vital role in classifying cancer into normal and abnormal. ObrayanH.Gomez et al., used assembled algorithms to construct classifier. The results of the experiments show that using the combinations of algorithms bagging cum multilayer perceptron gives high percentage of correctly classified instances [5]. Martin implemented both supervised and unsupervised learning mechanism for categorizing cervical cells into various stages. He used three classifiers for performance with K-fold cross validation, namely Hard C-means, Fuzzy C-means, and Gustafson-Kessel (GK). The proposed system's performance is better than previous classification techniques [6].

Y.M.S.AlWesabi et al., analyzed and presented various classification techniques and demonstrated the benefits of feature selection approaches for the best prediction of cervical cancer disease. This paper was very much helpful in understanding advantages and disadvantages of classifiers [7].

Jantzenetal., proposed linear network model for the classification of cervical cells. It is the easiest classifier which incorporates linear activation function to minimize the squared error. This method classified cervical cells into various stages. The output is measured by using K-fold cross validation with no feature reduction [8]. Athinarayan et al., proposed model for techniques of image processing such as segmentation, enhancement, extraction and classification of features. This system finds the details of the region of interest very clearly seen from the initial cervical cells and arc with more precision than the original image. Cervical cells of cervical cancer classified as CIN 0(Normal Dysplasia), CIN 1(Mild Dysplasia), CIN 2 (Moderate Dysplasia), CIN 3(Severe Dysplasia). Edwin et al., proposed using selected texture features to pap smear images. This method is very useful in selecting features that are most suitable for all classes such as LSIL (Low-grade squamous intraepithelial lesion) and HSIL (High-grade squamous intraepithelial lesion)[9].

III. MOTIVATION AND JUSTIFICATION OF THE PROPOSED WORK

Expert manual identification of cervical cells for cancer detection is a time-consuming and laborious task. Due to the presence of large numbers of cells on the glass slide, the problem becomes more complex. Other factors which may result in error by experts include poor contrast and inconsistent staining [10]. In the initial stage, cervical cancer cannot be diagnosed with any symptoms. And the difficulty lies in the process of analyzing the microscopic images where the degree of certainty is low in determining the impact of cancer. The medical experts say that “the detection of cervical cancer is an easy task and the diagnosis can be accurate, but the discrimination of cervical cancer is highly subjective and requires detailed discussion among experts” [11].

In view of all these factors, it is clear that the image classification algorithm is required to classify cells in a short time and to reduce screening errors for cervical cancer diagnosis.

IV. METHODOLOGY

The proposed system consists of segmentation, extraction of features, enhanced classification of K means, and the input image storage database. Figure 2 shows the architecture of the proposed classification system.

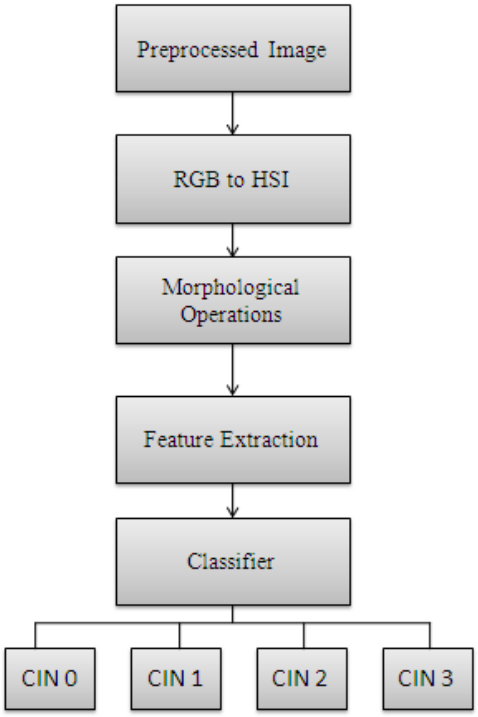


Fig. 2 Block Diagram for Classification

The system includes the following steps: preparation and pre-processing of microscopic cervix images, nucleus segmentation, extraction and classification of texture features. The pre-processed images are converted to HSI images from color. In order to segment nucleus and cytoplasm, morphological operations were then applied. The K means technique of clustering is used to classify the cells into various stages.

Preprocessing

Denoising images is one of the fundamental image processing techniques. Qualities of images are reduced due to heat generated by the electrons, bad sensors & vibrations. In order to keep the fine features of image a good noise removal technique is required.

Step 1: Input image X.

Step 2: Define the kernel size[3,3].

Step 3: Use genetic algorithm to find the best coefficients of the kernel size [3*3].

Step 4: Do repeat 50 to 100 iterations for finding the best coefficients.

Step 5: Use convolution process , multiply noise image with kernel to get denoised image.

The final results of this test case showed that if the parameters are properly optimized, the expected and valid denoising performance is provided by this bilateral filter with genetic algorithm [12].

Segmentation using morphological operation

For nucleus segmentation of biopsy images such as threshold-based, region-based, and clustering-based algorithm, several segmentation strategies have been used.

A good segmentation technique makes an impact in classification of picture.

The following steps will implement the algorithm. The microscopic biopsy image is initially taken as the input. A bilateral filter and genetic algorithm are used to remove the noise from the image [12]. As a next step, we convert the enhanced image into an HSI model. The nucleus from the saturation component of an input image is segmented with the help of thresholding. Global thresholding is applied to an intensity component of an image and the morphological operations are applied on the saturation and intensity components separately followed by decreasing the saturation and intensity components of input image into binary images. Then in order to get a new mask, the binary images are put together. After combining the binary images, connected component concept is used to perform Labeling. The criteria for performing labeling is that done the pixels must be connected with each other or the pixels must share similar intensity values in a connected component.

Lastly, the nuclei are precisely segmented and used to extract and classify features. The proposed method is based on morphological threshold operation and associated component labeling to remove cytoplasm nuclei and minimize over segmentation. Due to the variation in size, shape of nucleus and staining process, detection of nucleus become a challenging process. The important aspect of proposed approach is ability to find all nucleus cells in the image and remove all the other components. This provides correct information about whether nucleus is normal or abnormal. The improvement in segmentation accuracy is established with the help of quantitative measures. This method produces good accuracy in segmenting the nucleus compared to edge and region based segmentation. Further future work may be extended to increase the accuracy of segmentation to further and classify cervix tissue as normal and abnormal.

Feature Extraction

From the segmented images, regional features are extracted. Five characteristics are taken in our proposed method: energy, correlation, entropy, contrast and homogeneity,

Using the segmented pap smear images, these features are extracted. The feature energy displays pixel pair replication in an image.

It is given as

$$K_1=\sum_{x=0}^{N-1}\sum_{y=0}^{K-1}P_{\mu}^{2}(x,y) \quad (4.1)$$

In contrast, local variation in the image is measured. The high contrast value reveals large variations in the image.

$$K_2=\sum_{x=0}^{N-1}t^{2}\left\{\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}P_{\mu}(x,y)\right\} \quad (4.2)$$

Correlation is a measure of linear gray value dependency in a matrix. It is a 2D histogram in which separate pairs of pixels are allocated on the basis of a particular. Predefined vector of displacement as indicated.

$$K_3=\sum_{x=0}^{K-1}\sum_{y=0}^{K-1}\frac{(x,y)P(x,y)-\mu_1\mu_2}{\sigma_1^2\sigma_2^2} \quad (4.3)$$

Where the mean and standard deviation values of μ1, μ2, σ1, σ2 are individually related in the x and y directions.

Entropy is a non-uniformity measurement is an image based on the probability of co-occurrence values. It also indicates the image's complication.

$$K_4=-\sum_{x=0}^{K-1}\sum_{y=0}^{K-1}P_{\mu}(x,y)\log(P_{\mu}(x,y)) \quad (4.4)$$

Homogeneity in contrast to constant energy is inversely proportional, whereas it is inversely proportional to energy.

$$K_5=\sum_{x=0}^{K-1}\sum_{y=0}^{K-1}\frac{P_{\mu}(x,y)}{1+(x-y)^2},\ x\neq y \quad (4.5)$$

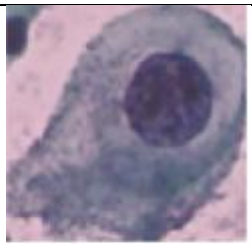
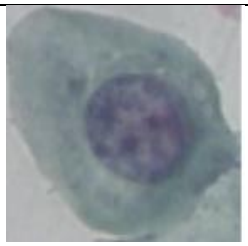
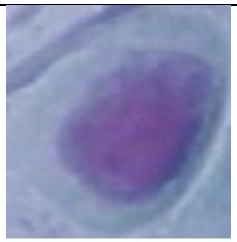
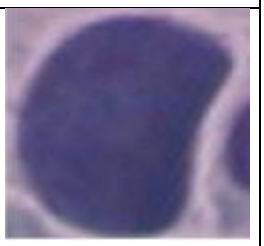
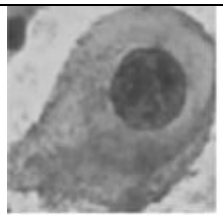
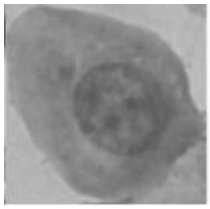
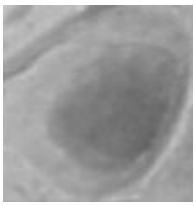
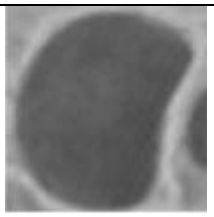
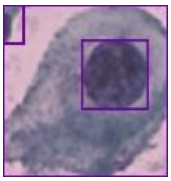
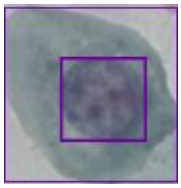
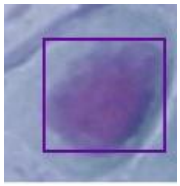
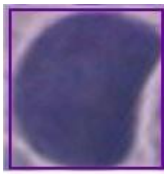
V. CLASSIFICATION

It is one of the clustering techniques used in related fields and medical imaging. Clustering is a learning that is not supervised. The algorithm gathers information into groups of k. It focuses on partitioning 'n' observations into k-clusters where each observation with the nearest mean belongs to the cluster. Features are grouped into the three categories in the proposed system: colour, regional, and texture. The letters in the figures S1, S2, S3, and S4 denote Normal (CIN-0), LSIL (CIN-1), HSIL (CIN-2), and Cervical Cancer (CIN-3). To gain a clear understanding of the relationship between the Color-Related Features inputs, we used linear classifiers over nonlinear ones. Features related to the subregion describe the color's geometric distribution. This enhanced K-means algorithm properly classifies the image input into normal or abnormal based on the image extracted characteristics.

VI. EXPERIMENTAL RESULTS

The cervical cancer images are obtained from Tuticorin government medical college. Images undergoes preprocessing and segmentation. The segmented images are classified using enhanced k-means clustering algorithm. The sample of four images (mild, moderate, severe, carcinoma) result are given in table .1.

Table. 1 Experimental Results

Input Image Model	Type-1	Type-2	Type-3	Type-4
Input Image				
Gray Image				
Segmented Nucleus				
Marked Object				
Cancer Stage	**Mild (CIN-0)**	**Moderate (CIN-1)**	**Severe (CIN-2)**	**Carcinoma (CIN-3)**

In the table, the first row contains four types input images (mild, moderate, severe, carcinoma). The second row contains gray image of input images. In subsequent rows, segmented nucleus and nucleus is marked with rectangle box. In last row, impact of cancer is specified. The proposed algorithm is tested against single cell and multi cell microscopic cervical cancer images.

VII. PERFORMANCE ANALYSIS

The performance of the proposed method is measured on the basis of precision, sensitivity, specificity. For classification, different types of cervical cell images are considered. Compared to Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Artificial Neural Network (ANN), the proposed classification technique is shown in table.2.

TP - True positive, classifies the abnormal (cancer) cells correctly.

TN - True negative, classifies the normal cells correctly.

FP - False positive, classifies normal cells wrongly.

FN - False negative, classifies abnormal(cancer) cells wrongly.

Accuracy- It is the percentage of cells from the total number of cells in the test set correctly classified. It is as follows:

$$Accuracy = \frac{(TP + TN)}{TP + FN + TN + FP}$$

Sensitivity- It is the percentage of abnormal class cells properly classified in the test set as abnormal. It is as follows:

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

Specificity- It is the percentage of normal class cells properly classified in the test set as normal. It is as follows:

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

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table. 2 Comparison of proposed experimental results with other systems

Evaluation Metrics	Regional Features with Enhanced K-means clustering (Proposed)	SVM	ANN	KNN
TP	76	74	75	70
TN	18	16	13	10
FP	4	6	7	11
FN	2	4	5	9
Sensitivity	97.43	94.8	93.75	88.60
Specificity	81.81	72.7	65	47.61
Accuracy	94	90	88	80

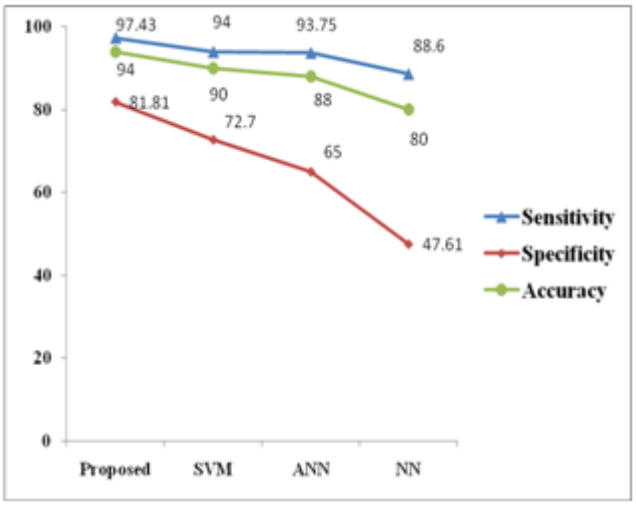


Fig. 3 Comparison result analysis by using sensitivity, specificity, accuracy for proposed system, SVM, KNN, ANN

VIII. CONCLUSION

The microscopic cervical images to identify the real impact of cancer can be classified, which helps in proper treatment of patient. This computer aided cervical cancer detection technique helps the physician to find and decide without hesitation, whether the given macroscopic cervical cancer image contains normal or abnormal cell. The proposed classifier algorithm correctly classifies an input image into mild, moderate, severe, and carcinoma. The proposed approach is compared to other classifications such as KNN, ANN, and SVM classifier for comparative analysis. The accuracy level of the proposed method (94%) showed that the algorithm is good for detecting cancer in the images in real time.

REFERENCES

1. http://www.imaginis.com/cervical-cancer/cervical-cancer-introduction.
2. Sharma.P, Pattanshetty.S.M., 2017, "A Study on Risk Factors of Cervical Cancer Among Patients Attending A Teritary Care Hospital: A Case-Control Study,", Clinical Epidemiology and Global Health (In press).
3. M.Anosouya Devi, S.Ravi, J.Vaishnavi and S.Punitha, "Classification of Cancer using Artificial Neural Network" 1877-0509 © 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.
4. M.K.Soumya, K.Sneha and C.Arunvinodh, "Cervical Cancer Detection and Classification Using Texture Analysis", Biomedical and Pharmacology Journal, Vol. 9(2), 663-671 (2016).
5. Obrayan H. G´omez, Eddy S´anchez-DelaCruz and A. Paulina de la Mata, "Classification of Cervical Cancer Using Assembled Algorithms in Microscopic Images of Paponicolaou", Research in Computer Science, pp. 125–134; rec. 2017-03-14; acc. 2017-05-06.
6. E.Martin, "Pap-smear Classification", Technical University of Denmark:Oersted-DTU, Automation, 2003.
7. Y. M. S. Al-Wesabi, Avishek Choudhur, "Classification of Cervical Cancer Dataset", Proceedings of the 2018 IISE Annual Conference.
8. J.Jantzen, J.Norup, G.Dounias, "Pap-smear benchmark data for pattern classification", in proceedings of Nature inspired Smart Information Systems (NiSIS 2005), 2005, pp. 1-9.
9. S. Athinarayanan and M.V. Srinath, "Classification of Cervical Cancer Cells in Pap Smear Screening Test", ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING, MAY 2016, VOLUME: 06, ISSUE: 04.
10. P.W.Huang andY.H.Lai, "Effective segmentation and classification for HCC biopsy images", Pattern Recognit., vol. 43, no. 4, pp. 1550-1563, 2010.
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, Volume 39, Issue 1, January 2009, Pages 69-78.
12. P.Robert, A.Celine Kavida, Published a paper titled "GENETIC ALGORITHM BASED DENOISING OF MEDICAL IMAGE" in International Journal of Control Theory and Applications ISSN: 0974-5572, 13th June, 2016