THINK INDIA JOURNAL
ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

A Review of Pap Smear Based Computer Aided Detection System and Application of Image Analysis For Early Detection of Cervical Cancer

Krishan Bansal

Department of Computer Science and engineering.

Lovely Professional University, Jalandhar 144001, India

Krishan.bansal22@gmail.com

Abstract:

Cancer of Cervical is one of the most widespread cancer form in world. As per WHO cervical cancer is the fourth most widespread form of cancer affecting women throughout the world and its early detection can help in saving life. Automated classification and diagnosis of cervical cancer is very importantrequirement as it allows timely,precise and consistentstudy of thepatient'shealth progress. This survey paper represents an outline of current research going in the field as presented or discussed in numerous research publications on automated classification and diagnosis of cervical(cervix) cancer from Pap smear images. Few of the techniques which include early detection of cervical cancer are segmentation, classification,from Medical Imaging, Machine Learning. The study revealed that some methods or techniques are used more commonly than others: for example thresholding, filtering, KNN and SVM are the most widely exploredmethods for pre-processing, segmentation, filtering and classification of images based on pap-smear. It has been observed that the outcomes based on classification and segmentation algorithms also depends on various other factors as well. Some of the factors which may affect the classification and segmentation are: the set of selected features, segmentation accuracy, and the type of pre-processing techniques used and the type of datasets used. K-Nearest-Neighbours is considered to be an exceptional classifier algorithm for cervical images giving more accuracy over other classification algorithms for a 2-class classification problem.

Keywords:*Cervical cancer, Pap-smear, Medical Imaging, Machine learning*

1. Introduction

Learning about the cervical cancer is one of the most worrying, alarming and dreaded health problems in women. Cervix uteri or the cervical cancer is thefourth most prevalentform of cancer that affected women in year 2018 as per WHO. In this year so far, the 569847 cases of cervical cancer and 311365 deaths due to cervix cancer were noted throughout world [1]. In India alone out of 96922 death rate is 60078[1]. These statistics marked up in 2040 by

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

prediction, with approximately 57337 new cases of cervical cancer and 39,608 deaths due to cervical cancer noted only in India i.e. almost 59% increase in incidence cases.

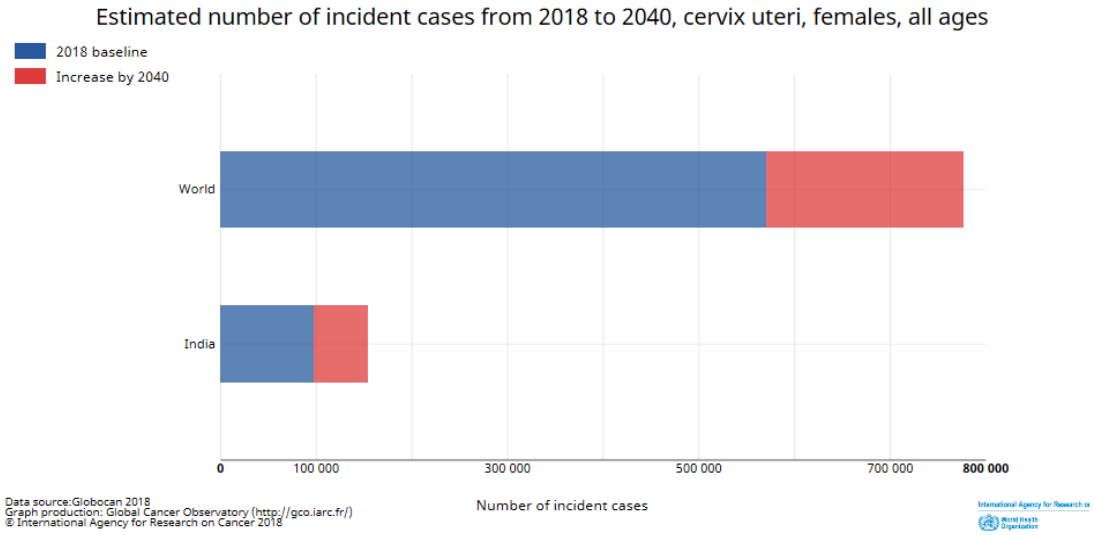

Fig 1.No of cervical cancer Incident predicted from (2018 to 2040)[1]

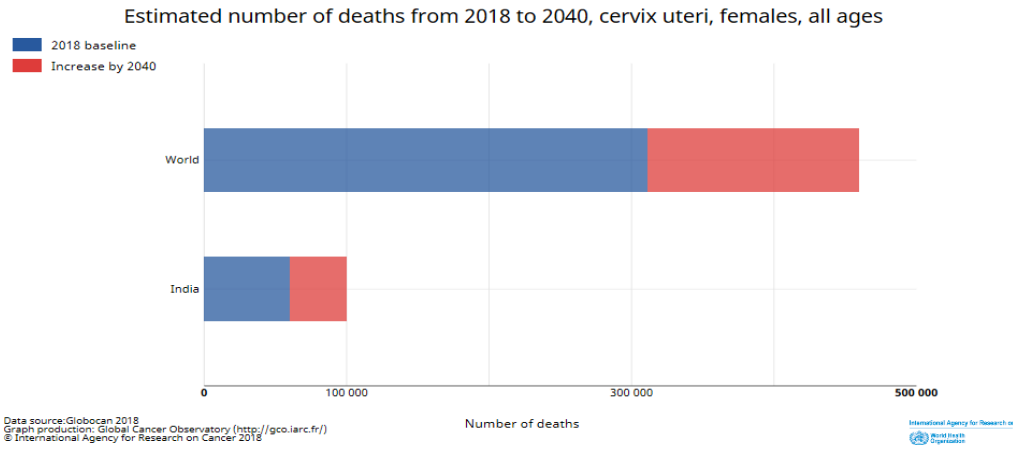

Fig 2. No of cervical cancer deaths predicted from (2018 to 2040) [1]

Figure 1 and 2 highlights that approximately 59% of incidence cases will increase with respect to 2018 in year 2040.Major health issue that is affecting the today's women's health is cancer of cervix, which is more prevalent in under developmentcountries [2]. Due to continuous use of Pap test, in the last 40 years, the cervical cancer death rate has been reduced by more than 50% in USA.[2]Pap test screening techniquehelps in finding changes in the cervix before it turns to cancer and with proper medication at preliminary stage it can be prevented. Pap testhelps in finding cervical cancer early in the most curable stage.Cervical

cancer is the one of major cause of morbidity and mortality in India. Cervical cancer is preventable if it diagnosed at the pre-cancerous stage with adequate and repetitive cytological screening by Pap smears.Discussion of this paper is as follows: Section 1 discuss about the Pap test, Pap smear cells and feature of Pap smear cells. Section 2 discusses about various techniques and methodologies used by several researchers. Section 3 consists of Summary or observation from the papers discussed in section 2. Last section 4 concludes the review.

1.1 Pap test

The purpose and need of the smear or Pap screening, is to analyseor identify cancerous cells from pre-cancerous cells, The Pap test if done on a regular basis can detect cervical cancer in the early stages and can help in saving human life.

In Pap test, a sample from the cervix is taken and transferred to a rectangular glass plate with a cotton swab or a small brush and coloured using the Pap method.Purpose of this method is to diagnose pre-cancerous cells before they get converted to cancerouscells [3].

1.2 Cell of cervical Pap-Smear

The specimen collected form cervix region is from squamous epithelium, metaplastic epithelium, and columnar epithelium [4]. Furthermore squamous epithelium cells found in lower part of cervix and columnar epitheliumis located in upper part of cervix. It is divided into four layers i.e. basal layer, para-basal layer, intermediate layer and superficial layer. On the other hand columnar epithelium has only one layer of cells. Metaplastic epithelium is found in between squamous epithelium and columnar epithelium. The cell characteristics are different layer by layer.

1.2.1 Feature Extraction

A number of significantdetailed cell characteristics depicted in Table 1 are used for analysing the images of cervical cells. These features are divided as:Shape (cytoplasm roundness, nucleus roundness etc.); Size (nucleus area,cell area, cytoplasm area etc.); Topology (nucleus position, Distribution of nucleus etc.); Percentage of empty cells etc.), Ratio (nucleus/cytoplasm ratio; Texture (Multi-nucleus cells etc.) and Colour intensity (cytoplasm brightness, cell, nucleusetc.);[14].

Table 1: few Cell characteristics/features used inclassification of cervical cancer [14]

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

Characteristics	Normal Cell	Degree of Abnormal cell		
		Mild	Moderate	Severe
Nucleus Area (μm^2)	20-50	50 -	50 -	50 -
Nucleus Intensity	Dark	Light	Dark	Dark
Cytoplasm Intensity	Light	Light	Dark	Dark
$\frac{N}{C}$ $Ratio$	1-2 %	10-20 %	20- 50 %	>50 %

Classification between normal cell and abnormal cell can be considered as 2- class classification. Classification of cells among normal, mild, moderate and severe dysplasia considered as 4-class classification whereas classification among different layer/type of cells such as parabasal, intermediate and superficial squamous epithelial cells,columnar epithelial cells, mild, severe and moderate dysplastic cells considered as 7-class classification.

1.3 Requirement to automate Pap test

The necessity of automating the Pap smear investigation is requiredto segment and categorize pap smear images into two different classes i.e. normal and abnormal [3,14].This can help cytologists reduce the time spent indoing slide examination in the pap screening process. Timely and reliable screening programs leads to timely detection of cervical cancer and this can save lives [3]. Automated pap-smear studyincludes the process and techniques to obtain comprehensive information from pap-smears images using computers for medicalinvestigation and medical interventions [14].

2. Literature Survey

Pap smear is a screening test used to detect precancerous stages of cervix cell so that development of cells to cancerous stage can be avoided as it is highly curable if detected on time. For automated pap smear classification segmentation of nucleus from background background alongwith feature selection algorithm. Related literature reviewsfor segmentation, feature selection and classifier algorithm used by researchers are mentioned in this section.

2.1 Literature Review

Ampazis et al.[3] proposed an effective Pap-smear test classifier with the usage of two highly effective and efficient neural network training algorithms of second order, specifically

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

the OLMAM (Optimized Levenberg-Marquardt with Adaptive Momentum)and the LMAM (Levenberg Marquardt with Adaptive Momentum) for creating an efficient pap-smear classifier. Techniques which these algorithms are using is feed forward neural network technique. Using this technique will have some error in order to normalize that error Mean Square Error (MSE) cost function is defined. Dataset of 917 images from DTU(Technical University of Denmark) is used as a test and training data for the study in this paper. The classification results achieved after applying LMAM and OLMAM algorithms is better than the other intelligent techniques.

Jantzen and Dounias[5] in the study used Pap-Smear image data which contains 917 images of pap-smear cells including 242 normal images and 675 abnormal images, further 20 features are used for the classification and classified the dataset in roughly 7 classes unevenly.The set of 3 classescontainnormal cells, while the set of 4 classescontain abnormal cells. The basic minimum requirement is to classify Pap cells into normaland abnormal cells, which is classified as a 2-class problem. By decreasing the 7 class problem into 2 class problem the accuracy can be increased and better classification of cancerous and non-cancerous cells.Paper is using feature selection algorithm with advanced classifiers techniques such as Ant colony optimization which is bringing down the error in classification of cells to just 1-2% from 6% error rate earlier.

Marinakiset al. [4]proposed making a pre diagnosis of the cell before they get converted to cancerous one. This paper is taking 2 data sets one of 917 images and second one of 500 images. Each cell image is read for 20 different features which are predefined by the team which is divided into 7 different classes and they are trying to convert the 7 class problem into 2 class problem. For knowing the best possible performing feature from the subset selection problem, an effective algorithm is needed which can classify the problem into 2-class problem, intelligent algorithm like genetic algorithm(GA) is proposed to classify the features into 2- different classes. GA which is based on the Darwin's theoryof survival of fittestto reproduce for next generation. Genetic algorithm is combined with the nearest neighbour based classification technique. Combining Results from GA and nearest neighbour shows that classification accuracy largely outperforms other previously used intelligent methods.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

Genctavet al. in paper [6]proposed the identification of precancerous cells based onshape,colour,characteristics of the nuclei and cytoplasm. Automating this process is difficult due to complex cell structure. In this paper Genctav A. et al. is proposing an unsupervised approach for classification and segmentation of malign and non-malign cells.[6] In segmentation, region of interest (RoI) of cell area is found in which cell is separated from rest of the noise in image,amulti-scale hierarchical segmentation algorithm is used to divide the cell regions based on circularity and homogeneity, a2-point classifier is used to confirm the distinction betweencytoplasm and nuclei within the cell regions. Cervical cancer is curable if early precursor lesions and cancerous changes caused by HPV are found in early stages. Procedure discussed in paper by Genctav et al. proposed to build a tree using classification algorithms and then organizing the cells in a sequential order using an algorithm called the optimal leaf ordering algorithm that is used to maximize the similarity of neighbouring leaves without any training example constraint or adjusting parameters.Performance evaluation is done using two data sets which are used to demonstrate the effectiveness and efficiency of the proposed approach in images having overlapping cells, inconsistent stainingand poor contrast. Curing the cervical cancer is very much related to detecting the stage of cancer at the time of diagnosis.

Riana et al. [7] in their paper discussed the benefits of segmenting and classifying the cells in terms of nuclei and cytoplasm for segmenting the cells from the background and taking the cells in the front which are inflammatory. With the help of inflammatory cells detection of cell nuclei will become easy and accurate and hence the classification of cancerous and noncancerouscells is more accurate and reduce the interference of the cells resulting in misclassification. They are using a database of 222 cytological images of pap smears which is having 418 cells. In dataset after segmentation and preprocessing the image is passed using morphological analysis and then the cell nuclei is found and further it can be used to differentiate cancerous from noncancerous cell.

Tsai et al. [8] in their paper developed a nucleus contour (CNC) and cytoplast detector algorithm to extractcytoplast and nucleus from a cervical papsmear image.In this paper, the author is proposing a2-group enhancer to mark a clear-cut differentiator linefor the pixels lying between two objects, and the maximal colour difference (MCD) method is used to draw

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

the nucleus contour to that of cytoplast. Various edge detection techniques like canny edge detection etc. are used to detect the cytoplast and Nucleus contour

Lumini et al.[9] and chen et al. [10] proposed the use of SVM along with other selection strategy techniques so that the Support Vector Machine (SVM) can be trained in a more intelligent way. Both the paper are focussing applying SVM either on filter type approaches which are independent of SVM and other is Wrapper based techniques in which modifications of SVM can be used in better feature selection of the image.

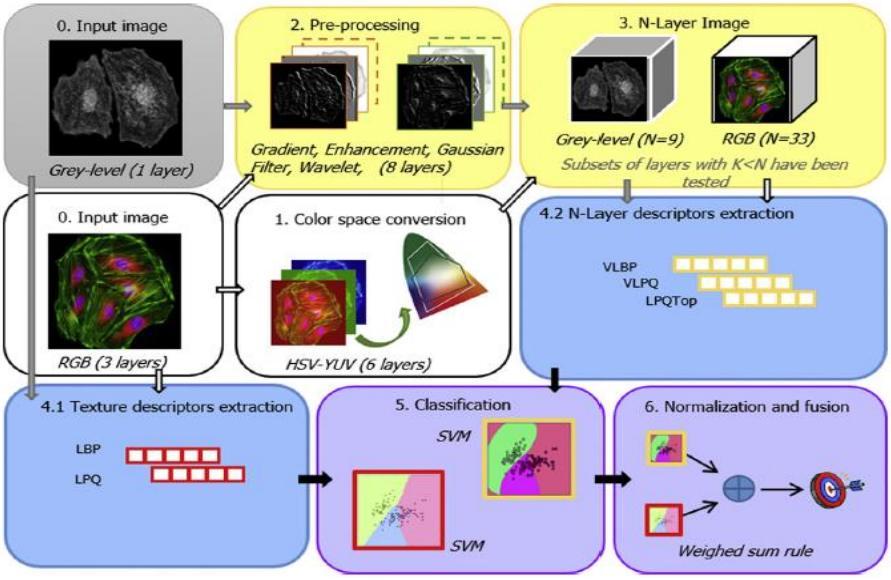

Fig. 3 Schema of the approach [9]

Sukumar and Gnanamurthy [11] in their paper proposed the computer aided or software aided automaticdiagnosis and detection method for cervical cancer using HPV or Pap smear images. The anticipated methodology includes preprocessing, feature extraction, segmentation of the core region and classification of the Pap test image. Morphological operations are used to segment the nucleus cell region. The Grey, wavelet and GLCM level characteristics (Grey level co-occurrence matrix) are extracted from the normal and dysplastic cell nucleus. These extracted characteristics are trained and classified using the ANFIS (Neuroadaptive Diffuse Inference System) classifier.The method proposed in the paper has achieved 99.65% specificity, 92.68% sensitivity and 98.74% accuracy for dysplasticcell segmentation.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

Plissiti et al. in their paper proposed an automated technique for boundary recognition and determination of cells nuclei in the conventional Pap smear image. Nuclei area is determined using the morphological based image reconstruction process and nuclei boundaries are segmented using watershed transform. For the elimination of false positive results, the characteristics that characterize the shape, texture and intensity of the image are extracted from the regions of the candidate nucleus and a classification step is performed to determine the true nuclei.Dataset of 90 pap smear images is taken with 10248 cells and both supervised (SVM) and unsupervised (K-means spectral clustering) clustering techniques are implied which helped in finding the cell nuclei more accurately and thus better classification of cells can be done in 2 class problem.

Winer et al. [13] stated that women who do not attend or delay in the Pap screening test have an increased risk of cervical cancer as early detection with Pap test helps in timely diagnosis and prevents Pap smear to convert to cancerous cell.Clinical trials in countries with organized screening programs have shown that sending self-sampling kits to women having high-risk human papillomavirus (HPV) infection, helps women who have not been examined increases participation, and helps in early detection of the disease.[13] with the help of self-screening kits more awareness and randomized controlled trial set are executed because of which in past 40 years cases of cervical cancer has been reduced by 50% in US [1].

3. Techniques Used/ Observation from various papers

Below Mentioned table summarizes the methods used by various authors for pre-processing, segmentation and classification of pap smear or cervical cells.

Table 2: Discussion of techniques and dataset used in various papers

Dataset Used	Techniques of segmentation, number of Features, problem	Classification	Metrics/ Result	Ref. No & years
500 cases contains 200 normal and 300 abnormal cell images collected	Nucleus and cytoplasm extraction , 20 features used	LMAM (Levenberg Marquardt with Adaptive Momentum) and OLMAM	Accuracy = 98.86 %	[3] , 2004

from Herlev University Hospital, Denmark	and 2 class problem	(Optimized Levenberg - Marquardt with Adaptive Momentum)		
917 Samples contain 242 normal and 675 abnormal	Nucleus and cytoplasm extraction , 20 features, 7 class problem and 2 class problem	Minimum distance classifier	Error rates for the 7-class problem are much larger than for the 2-class problem.	[5], 2006
917 and 500 samples	Nucleus and cytoplasm extraction , 20 features, 7 class problem and 2 class problem	K–nearest neighbor(KNN)with all features activated and infusing GA	Accuracy = 97.2 % for 2 class problem and Accuracy = 92.8 % for 7 class problem	[4], 2009
917 cell images collected from Herlev University Hospital, Denmark. Hacettepe data set, 82 pap smear image, collected from Hacettepe University Hospital	Nucleus and cytoplasm extraction, 14 Feature, 7 class problem	unsupervised clustering approaches	Precision= 88%	[6], 2012
222 images collected from Clinical Pathology Laboratory, Indonesia.	Segmentation with global threshold, 3 feature	gray level thresholding and distance rule	simplifies the nuclei detection process	[7], 2015
26 cervical smear image collected	Couture Detection	K-mean algorithm	Proposed algorithm used	[8], 2008

from Taichung Hospital, Taiwan			to segment nucleus and cytoplasm from background.	
500 images collected from Herlev University Hospital, Denmark	Grey level, wavelet and GLCM (Grey level co-occurrence matrix) features	ANFIS (Adaptive neuro fuzzy inference system) classifier.	Sensitivity= 92 .68%, accuracy = 98.74%	[11] , 2016

4. Conclusion

KNN algorithm provides anexceptional classifier method for Pap smear images. Combining Genetic algorithm and KNN or feed forward networks can further improve the performance of the system. Furthermore application of multi-level segmentation and classification techniques can further improve in distinguishing the images into 2-class problem from multi class problem.Majority of the proceduresand algorithmic steps have been validated by training with test datasets provided from the online cervical image datasets from DTU (Technical University of Denmark) and some other online datasets available as well. After training through datasets the results have been validated through proper testing of the new images, which are resulting in correct identification of malign and non-malign cervical cells.

5. References

[1] International association for research on cancer, "A tool that predicts the future cancer incidence and mortality burden worldwide from the current estimates in 2018 up until 2040," WHO, 2018. [Online]. Available: http://gco.iarc.fr/tomorrow/home. [Accessed 24 11 2018].

[2] American Cancer Society, "Cervical Cancer," [Online]. Available: https://www.cancer.org/cancer/cervical-cancer.html. [Accessed 24 11 2018].

[3] N. Ampazis, G. Dounias and J. Jantzen, "Pap-Smear Classification Using Efficient Second Order Neural Network Training Algorithm," *Methods and applications of Artificial Intelligence,* pp. 230-245, 2004.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

[4] 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 Biolog yand Medicine,* pp. 69-78, 2009.

[5] J. Jantzen and G. Dounias, "Analysis of Pap-smear Image Data," *Nature-Inspired Smart Information Systems,* 2006.

[6] A. Genctav, S. Aksoy and S. Onder, "Unsupervised segmentation and classification of cervical cell images," *Pattern Recognition,* vol. 45, no. 12, pp. 4151-4168, 2012.

[7] D. Riana, M. E. Plissiti, C. Nikou, D. H. Widyantoro, T. L. R. Mengko and O. Kalsoem, "Inflammatory Cell Extraction and Nuclei Detection in Pap Smear Images," *International Journal of E-Health and Medical Communications,* vol. 6, no. 2, pp. 27-43, 2015.

[8] M.-H. Tsai, Y.-K. Chan, Z.-Z. Lin, S.-F. Yang-Mao and P.-C. Huang, "Nucleus and cytoplast contour detector of cervical smear image," *Pattern Recognition Letters,* vol. 29, pp. 1441-1453, 2008.

[9] A. Lumini, L. Nanni and S. Brahnam, "Multilayer descriptors for medical image classification," *Computers in Biology and Medicine,* vol. 72, pp. 239-247, 2016.

[10] Y.-W. Chen and C.-J. Lin, "Combining SVMs with Various Feature," 2005.

[11] P. Sukumar and R. K. Gnanamurthy, "Computer Aided Detection of Cervical Cancer Using Pap Smear Images Based on Adaptive Neuro Fuzzy Inference System Classifier," *Journal of Medical Imaging and Health Informatics,* vol. 6, pp. 1-8, 2016.

[12] M. E. Plissiti, C. Nikou and A. Charchanti, "COMBINING SHAPE, TEXTURE AND INTENSITY FEATURES FOR CELL NUCLEI EXTRACTION IN PAP SMEAR IMAGES," *Pattern Recogniotion Letters,* vol. 32, no. 6, pp. 838-853, 2011.

[13] R. L. Winer, J. A. Tiroc, D. L. Miglioretti, C. Thayere, T. Beatty, J. Lin, H. Gao, K. Kimbel and D. S. Buist, "Rationale and design of the HOME trial: A pragmatic randomized controlled trial of home-based human papillomavirus (HPV) self-sampling for increasing cervical cancer screening uptake and effectiveness in a U.S. healthcare system," *Contemporary Clinical Trials,* 2017.

[14] N. A. Mat-Isa, M. Y. Mashor and N. H. Othman, "An automated cervical pre-cancerous diagnostic System," *Artificial intelligence in medicine* , vol. 42, pp. 1-11, 2008.