A literature survey of automated detection of cervical cancer cell in Pap smear images

N. Lavanya Devi* and P. Thirumurugan

Department of Electronics and Communication Engineering,
P.S.N.A. College of Engineering and Technology,
Dindigul, Tamil Nadu, India
Email: devilavanya@gmail.com
Email: thirujl.murugan@gmail.com
*Corresponding author

Abstract: Cervical cancer is one of the major reasons for gynecologic cancer even though it can be treated if detected in the early stage. Pre-cancer identification of the uterine cervix can be done by the Pap smear test. In the present day scenario, Pap smear images are taken and the patient has to wait for an expert's suggestion. This time consuming conventional method can be replaced by automating the process, which gives a rapid and accurate result. The automated screening of cervical cancer saves human resources and material and gives better accuracy by reducing human errors than an expert's review. The major steps involved in the automated classification are preprocessing the image, segmentation, feature extraction, classification, and analysing the classification result. This paper discusses the various algorithms that were used for segmenting and classifying the abnormal and normal cells based on the features extracted.

Keywords: preprocessing; segmentation; classification; feature extraction; Pap smear test; cervical cancer; Pap smear images; gynecologic cancer; uterine cervix; abnormal cell; normal cell.

Reference to this paper should be made as follows: Lavanya Devi, N. and Thirumurugan, P. (2022) 'A literature survey of automated detection of cervical cancer cell in Pap smear images', *World Review of Science, Technology and Sustainable Development*, Vol. 18, No. 1, pp.74–82.

Biographical notes: N. Lavanya Devi is pursuing her PhD under Anna University Tamil Nadu, India. She received her BE and ME degree from Anna University. Her areas of interest are medical image processing and digital design. She is a member of ISTE and IETE.

P. Thirumurugan received his PhD, ME, and BE degrees from Anna University, Chennai. His area of interest is image processing, VLSI, etc. He has received several grants from government organisations like CSIR, ICMR, etc. He received 'Best Researcher Award-2018' and 'Award of Excellence in Research 2018' from International Association for Science and Technical Education (IASTE), Chennai. He is a life member of ISTE, IASTE.

1 Introduction

The most common type of cancer among women is cervical cancer with the mortality rate as high as 275,000 deaths annually whose cure rate depends on the stage of cancer when detected (Wiebe et al., 2012). Pap test is based on obtaining the uterine cervix cells, and smearing on the glass for checking the presence of human papilloma virus (HPV) which is considered as the cause of cancer.

The abnormal cervical cells with precancerous symptoms are called as dysplastic cells. Dysplastic cells can be categorised into mild dysplastic cells, moderate dysplastic cells, and severe dysplastic cells based on the severity of the abnormality. Mild dysplastic cells will have a bright large nucleus, moderate dysplastic cells will have a dark large nucleus cell and severe dysplastic cells will have a dark large nucleus with dark small cytoplasm (Li et al., 2012). Hence, the cells with an abnormality can be differentiated from the normal one by various features of the nucleus and cytoplasm including nuclear perimeter, area, maximum axle length L, maximum axle width W, etc.

Section 2 presents a review of various segmentation methods employed so far. Section 3 gives a brief overview of a survey on feature extraction and Section 4 elaborates different classification techniques implemented recently. Section 5 briefly describes commonly used evaluation metrics to quantify the efficiency of the proposed algorithm. Finally, the conclusion is included in Section 6.

2 Survey of Segmentation

Segmentation partitions the given image into different parts. If the image has a sufficient difference in the grey level value, the required object can be well differentiated from the background (Veluchamy et al., 2012; Lavanya Devi et al., 2013). In the segmentation process of Pap smear images, the image will be divided into background, nucleus, and cytoplasm. The challenges faced by the segmentation algorithm are:

1 obscure boundary of the cytoplasm due to low-intensity contrast with the background

2 contaminations caused by bloodstains, inflammatory cells, etc. (Wiebe et al., 2012).

As the cells are stained, they are coloured with blue and red and contribute to grey level values whereas the background remains colourless and contributes to the white pixel.

The segmentation can be broadly divided into supervised and unsupervised methods. In the supervised method of segmentation, the user should have prior knowledge about the output to be obtained. In the unsupervised method of segmentation, the user may not have prior knowledge about the output to be obtained. The following algorithms are frequently used for performing segmentation of cervical cells.

- radiating gradient vector flow (RGVF)
- edge-based Laplacian of Gaussian
- multi-scale watershed segmentation
- region growing segmentation
- active contour technique

- clustering-based segmentation
- weakly-supervised learning in CNN.

Sajeena and Jereesh (2015) automated the detection of cervical cancer by RGVF segmentation and SVM classification. The proposed preprocessing of the image, feature extraction, segmentation, classification, and cell labelling as the major steps of computer-aided detection of cervical cancer. RGVF snake is used for properly defining the uncertain boundaries which may be affected by stains. Area of the nucleus, nucleus compactness, major axis of the nucleus, minor axis of the nucleus, nucleus aspect ratio, nucleus homogeneity, nucleus to cytoplasm (N/C) ratio, entire cell area, compactness of the entire cell were the features extracted and the system is trained. For classifying the cells into normal and abnormal categories artificial neural network (ANN), support vector machine (SVM), and Euclidean distance (ED)-based classifiers were used. Finally, the accuracy, sensitivity, and specificity of ANN, SVM, and ED classifiers were compared. Among the three classifiers, SVM gave better results for accuracy, sensitivity, and specificity. The drawback of the SVM classifier is that when the number of training samples is larger it requires high memory and training time.

Neghina et al. (2016) proposed an algorithm for the automated detection of the cervical cells in Pap smear images using polar transformation and K-means segmentation. The seed point needed for K-means segmentation is assumed to be present in the nucleus and the polar transformation is built around the seed point. The probability of belongingness of each pixel to each class is calculated by fuzzy logic and the final decision is taken.

Kumar et al. (2011) developed an algorithm for the detection of cervical cytology in Pap smear images. They used E-smear software for image acquisition and digitisation and image preprocessing was done. Edge-based Laplacian of Gaussian technique was used for nuclear segmentation. Of the various features identified, i.e., BDYVAR, nuclei area, integrated optical density (IOD), eccentricity, homogeneity, contrast, mean R, energy packet 4, energy packet 16, Maitomin, energy detail 2 were ranked accordingly. The training and testing are done by the SVM classifier and were represented by the optimal hyperplane located in the middle of both support hyperplanes. They are defined by a small number of training patterns called support vectors. The true classification of the artefacts is 88.1%.

Gençtav et al. (2012) proposed unsupervised segmentation and classification for analysing the Pap smear images. The meaningful regions are automatically selected from the hierarchical segmentation tree. The ranking of the cell was proposed based on the artefact of the cell, first, constructs a binary tree using hierarchical clustering according to the features extracted from the nucleus and cytoplasm regions. Multi-scale watershed segmentation is used to generate a hierarchical partitioning of cell regions. Nucleus regions could be differentiated from other regions by their spectral homogeneity and shape features. The nucleus region remains the same up to a certain number of some levels and may show wide variation at a particular level, which was the indication of overlapping cells. The classification of the cell was based on the features like size, mean intensity, circularity, homogeneity, nucleus brightness, nucleus area, nucleus shortest diameter, nucleus longest diameter, etc.

Boughzala et al. (2016) developed a method for the segmentation of colour images tested with L*a*b, RGB, YCbCr, and HSV colour spaces in which RGB is device independent, L*a*b includes all the perceivable colours, HSV has the ability to separate the intensity of the colour information hue and saturation, YCbCr represents colours in terms of two chrominance components (Cb and Cr) and one luminance component (Y). K means clustering algorithm is applied to all the above-mentioned colour spaces and the results obtained for RGB turned out to be fruitful. RGB produced an average rate of correct segmentation as 64.69% for nucleus segmentation and correct cytoplasm segmentation as 81.79%.

Oprisescu et al. (2015) explained an algorithm for the Pap smear nuclei detection which included the following steps. In the first step, mean shift filtering is used as a preprocessing step which is considered as a promising method for filtering. In the second step, the nuclei edge gets highlighted as a result of filtering and the edge can be detected. In the third step, seed points were identified for region growing segmentation. In the final step, the suitable threshold value is chosen and eccentric filtering was done to avoid too small & too large regions that were not considered as nuclei. The algorithm suggested produced an average sensitivity of 94.72% and an average specificity of 97.09% for nuclei detection. The summary of the different algorithms for cervical cell segmentation is depicted in Table 1.

Sangworasil et al. (2018) proposed a method for automatic screening of cervical cancer cells. The separation of the nucleus region was done by the k-means method. The separation of the cytoplasm was done by relating the edge to the geometrical rotation method. The classification of the cervical cells into the abnormal and normal classes was done by calculating the ratio of the area of the nucleus to the area of cytoplasm. If the calculated ratio was greater 0.25 then it will belong to the abnormal class otherwise it will belong to the normal class.

Table 1 Summary of different algorithm for cervical cell segmentation

Authors	*Method*	*Drawback*
Dagher and El Tom (2007)	Hybrid watershed balloon snake method	Reducing watershed over segmentation problem.
Krishnan and Sujatha (2010)	Contour based approach using magnitude and directional values to extract boundary	Multiple object detection or overlapping region identification not possible.
Yang-Mao |et al. (2008)	An edge enhancement nucleus and cytoplast contour detection	Not suitable for overlapping of the cells.
Byju et al. (2013)	Edge based approach using customised Laplacian of Gaussian filter to segment cell nuclei	Performs well for stain varied images. Performance can still be improved using parallel processing.
Shah (2012)	Two phase segmentation approach merging an unsupervised clustering and shape classification model	Suitable for the Pap smear images at lower magnification.

Iwai and Tanaka (2017) suggested a method of an automatic system for screening cervical cancer to support the cytopathologists. The authors divided the cells into red and blue cells because the colour of the cell will be different for superficial, basal, and intermediate squamous cells. The nucleus will be segmented from the cells by superpixel segmentation. R component is used for red cells and 0.3 is defined as the threshold. The images are converted into HSV space and the threshold is defined as 0.6 for the hue component for the blue cells. Parameter roundness is used to define the circularity. The colour density of the nucleus and its enlargement is used as the features.

Zhao et al. (2019) developed an automated segmentation using the deformable multipath ensemble model for the cervical nucleus in Pap smear images. Feature information is used more efficiently by a dense block of the U-shaped convolution network. To deal with irregular shape and size deformable convolution is used. The main architecture of ensemble modelling consists of the expansive path and a contractive path. The contractive path consists of a dense block for future extraction, a concatenation block to receive the information from the feature extraction, and a transition down the block. In the expansive path, decoding stages for the corresponding blocks are formed.

3 Survey of feature extraction

Feature extraction is the process of extracting the various characteristics which will help in the accurate classification of cervical cells into affected and non-affected parts of the cell. The various features extracted by the researcher for the classification are

- nucleus feature
- texture feature
- cellularity feature
- acellular feature (light area)
- coarseness feature
- contrast feature
- pseudo colour feature extraction method.

Guo et al. (2016) considered cellularity features, acellular (light area) features, nuclei features, combination features, texture features, advanced layer-by-layer triangle features. Nuclei ratio, cytoplasm ratio, acellular ratio were considered as cellularity features; intensity ratio, ratio R, ratio G, ratio B, ratio light area, light area to background area ratio, luminance ratio were considered as acellular features; background to nuclei area ratio, average nucleus area, were considered as nuclei features; ratio acellular number to nuclei number, ratio acellular area to nuclei area were considered as combination features; contrast energy, correlation, and uniformity along with the statistical parameters generated from grey-level co-occurrence matrix (GLCM) were considered as texture features; and finally triangles in the top layer, triangles in mid-layer, triangles in the bottom layer were considered as advanced layer-by-layer triangle features. SVM and LDA classifiers were used.

Bhan et al. (2016) proposed a method for accurate segmentation by dividing the whole image into small non-overlapping blocks and then the texture features were extracted from the GLCM. Independent level sets segmentation was used in this paper. Features like autocorrelation, cluster prominence, the sum of squares, sum variance showed a great variation in the area of overlapping. The obtained dice coefficient used for the evaluation of the efficiency of the proposed algorithm was 0.76.

Chen et al. (2014) suggested a method for the semi-automatic segmentation of nucleus and cytoplasmic contours and classification of Pap smear images. The nucleus cytoplasmic contours (NCC) is detected by a NCC detector. The image analysis system was provided with the micrometre image, various variables, including nuclear perimeter, area, axle parameters like width and length, N/C ratio, the maximum and average length of the axle to perimeter, average and maximum length from the centre of gravity to perimeter for calibrating the size and shape irregularity of a cell. The features like coarseness and contrast were also considered for analysing the texture of the nucleus. Wrapper method based on recursive feature elimination (RFE), recursive feature addition (RFA) was used to eliminate the unwanted features, to add an important feature respectively which creates a large margin between the classes. Genetic algorithm along with SVM was used for classification which provided better results within an acceptable time.

4 Survey of classification

Image classification is the process of assigning pixels in the image to desired categories. Supervised classification requires expert knowledge whereas unsupervised classification does not require any prior knowledge. The following are the repeatedly used algorithms

- naïve Bayes
- SVMs
- random forest tree
- convolutional neural network
- Gabor filters.

Zhang and Lu (2017) used convolutional neural network as the classifier which comprises convolutional operation which performs a 2D convolution, nonlinearity and pooling operation which down samples the feature maps, followed by fully connected layers which convert the feature map to feature vector as the major steps. Even though this classification algorithm gives high accuracy it requires a large time for classification and also this method misclassifies a few severe dysplasias (4.1%) and carcinoma (2%) cells as normal. The demerit of this method is that it is hard to detect advanced stages of abnormality.

Kurniawati et al. (2016) compared the performance of naïve Bayes, SVMs, and random forest tree classifiers by considering accuracy, recall, precision, and ROC curve as the performance metric. Based on the performance evaluation, random forest tree turned out to be the best classifier among the other two classifiers.

Faturrahman et al. (2017) proposed deep belief network algorithm for the classification of the cervical cell with multi feature fusion among local binary pattern (LBP), GLCM, and shape features. This algorithm classified abnormal and normal cells with 97.35% accuracy. A summary of different features and algorithms used for classification is depicted in Table 2.

Table 2 Summary of different features and algorithms used for classification

Authors	*Proposed algorithm*	*Feature considered*	*Disadvantages*	*Accuracy*
Gençtav et al. (2012)	Non-parametric hierarchical segmentation. Uses spectral, shape and gradient information for classification.	12 features were extracted for ranking cervical cells including both spectral and shape	Cells with double nucleus not considered.	96%
Chen et al. (2014)	SVM wrapper method based on recursive feature addition (SVM-RFA)	13 morphometric variables from four cellular features and textural features.	Method expected to diminish the number of grey level zone. Required more statistical tool.	96.12% and 98.61% accuracy using four cluster and two cluster classifier.
Plissiti and Nikou (2013)	Fuzzy C-means with kernel principal component analysis (K-PCA) Gaussian kernel	20 features concerning about intensity and shape characteristics.	Nonlinear dimensionality reduction improves performance. Unsupervised classifier based on nuclei features cannot suitable for dimensionless cases.	90.42% accuracy considering all features. 90.58% accuracy considering only nuclei features.

5 Evaluation

The performance of the proposed algorithms can be measured by the precision, sensitivity, accuracy, etc.

$$\text{Specificity} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{3}$$

where

- true positive (TP) – pixels correctly classified as abnormal cells
- true negative (TN) – pixels correctly classified as normal cells
- false negative (FN) – pixels wrongly classified as normal cells
- false positive (FP) – pixels wrongly classified as abnormal cells.

6 Conclusions

Automated detection of the cervical cell will reduce the false-negative cases which will reduce the mortality rate in developed countries where most of the deaths are caused due to the ignorance about the situation. In this paper, a detailed literature review on different segmentation algorithm, feature extraction, and various classifiers which contributed to the accurate automated detection of tainted cervical cells in Pap smear images is done. Finally, the validation of the proposed algorithm is also done.

References

Bhan, A., Vyas, G. and Mishra, S. (2016) 'Supervised segmentation of overlapping cervical pap smear images', *International Conference on Signal Processing and Communication (ICSC)*, IEEE, pp.225–228.

Boughzala, O., Guesmi, L., Ben Abdallah, A. and Bedoui, M.H. (2016) 'Automatic segmentation of cervical cells in Pap smear images', *24th Conference on Computer Graphics, Visualization and Computer Vision*, pp.343–349.

Byju, N.B., Sujathan, V.K., Malm, P. and Kumar, R.R. (2013) 'A fast and reliable approach to cell nuclei segmentation in PAP stained cervical smears', *CSI Trans. ICT*, Vol. 1, No. 4, pp.309–315.

Chen, Y-F., Huang, P-C., Lin, K-C., Lin, H-H., Wang, L-E., Cheng, C-C., Chen, T-P., Chan, Y-K. and Chiang, J.Y. (2014) 'Semi-automatic segmentation and classification of pap smear cells', *IEEE Journal of Biomedical and Health Informatics*, Vol. 18, No. 1, pp.94–108.

Dagher, I. and El Tom, K. (2007) 'WaterBalloons: a hybrid watershed balloon snake segmentation', *IEEE Int. Conf. Neural Networks – Conf. Proc*., Vol. 26, pp.1–6.

Faturrahman, M., Wasito, I., Mufidah, R. and Ghaisani, F.D. (2017) 'Multi feature fusion using deep belief network for automatic pap-smear cell image classification', *IEEE International Conference on Computer, Control, Informatics and its Applications*.

Gençtav, A., Aksoy, S. and Önder, S. (2012) 'Unsupervised segmentation and classification of cervical cell images', *Pattern Recognit*., Vol. 45, No. 12, pp.4151–4168.

Guo, P., Banerjee, K., Stanley, R.J., Long, R., Antani, S., Thoma, G., Zuna, R., Frazier, S.R., Moss, R.H. and Stoecker, W.V. (2016) 'Nuclei-based features for uterine cervical cancer histology image analysis with fusion-based classification', *IEEE Journal of Biomedical and Health Informatics*, November, Vol. 20, No. 6, pp.1595–1607.

Iwai, A. and Tanaka, T. (2017) 'Automatic diagnosis supporting system for cervical cancer using image processing', *Proceedings of the SICE Annual Conference*.

Krishnan, N. and Sujatha, S.N.N. (2010) 'Segmentation of cervical cancer images using active contour models', *2010 IEEE Int. Conf. Comput. Intell. Comput. Res*., pp.1–8.

Kumar, R.R., Kumar, V.A., Kumar, P.N.S., Sudhamony, S. and Ravindrakumar, R. (2011) 'Detection and removal of artifacts in cervical cytology images using support vector machine', *International Symposium on IT in Medicine and Education*, IEEE, pp.717–721.

Kurniawati, Y.E., Permanasari, A.E. and Fauziati, S. (2016) 'Comparative study on data mining classification methods for cervical cancer prediction using pap smear results', *IEEE 1st International Conference on Biomedical Engineering (IBIOMED)*, Yogyakarta, Indonesia.

Lavanya Devi, N., Priya, S.P. and Krishanthana, K. (2013) 'Performance analysis of face matching and retrieval in forensic applications', *International Journal of Advanced Electrical and Electronics Engineering*, Vol. 2, No. 2, pp.100–104.

Li, K., Lu, Z., Liu, W. and Yin, J. (2012) 'Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF snake', *Elsevier Pattern Recognition*, Vol. 45, No. 4, pp.1255–1264.

Neghina, M., Rasche, C., Ciuc, M., Sultana, A. and Tiganesteanu, C. (2016) 'Automatic detection of cervical cells in pap-smear images using polar transform and k-means segmentation', *Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE.

Oprisescu, S., Rădulescu, T., Sultana, A., Rasche, C. and Ciuc, M. (2015) 'Automatic pap smear nuclei detection using mean-shift and region growing', *International Symposium on Signals, Circuits and Systems (ISSCS)*, IEEE.

Plissiti, M.E. and Nikou, C. (2013) 'Cervical cell classification based exclusively on nucleus features', *International Conference Image Analysis and Recognition*, pp.483–490.

Sajeena, T.A. and Jereesh, A.S. (2015) 'Automated cervical cancer detection through RGVF segmentation and SVM classification', *IEEE Intl. Conference on Computing and Network Communications (CoCoNet'15)*, pp.663–669.

Sangworasil, M., Sukkasem, C., Thongchoom, R. and Janyasupab, M. (2018) 'Automated screening of cervical cancer cell images', *Biomedical Engineering International Conference*.

Shah, S. (2012) 'Automatic cell image segmentation using a shape-classification model', *IAPR Conf. Mach. Vis. Appl*., pp.428–432.

Veluchamy, M., Perumal, K. and Ponuchamy, T. (2012) 'Feature extraction and classification of blood cells using artificial neural network', *American Journal of Applied Sciences*, Vol. 9, No. 5, pp. 615–619.

Wiebe, E., Denny, L. and Thomas, G. (2012) 'Cancer of the cervix uteri', *International Journal of Gynecology & Obstetrics*, Vol. 143, No. Suppl. 2, pp.100–109, Published by Elsevier Ireland Ltd.

Yang-Mao, S.F., Chan, Y.K. and Chu, Y.P. (2008) 'Edge enhancement nucleus and cytoplast contour detector of cervical smear images', *IEEE Trans. Syst. Man, Cybern. Part B Cybern*., Vol. 38, No. 2, pp.353–366.

Zhang, L. and Lu, L. (2017) 'DeepPap: deep convolutional networks for cervical cell classification', *IEEE Journal of Biomedical and Health Informatics*, Vol. 21, No. 6, DOI 10.1109/JBHI.2017.2705583.

Zhao, J., Li, Q., Li, X., Li, H. and Zhang, L. (2019) 'Automated segmentation of cervical nuclei in pap smear images using deformable multi-path ensemble model', *IEEE 16th International Symposium on Biomedical Imaging (ISBI2019)*, pp.1514–1548.