

A Survey of Feature Extraction for Pap-Smear Image Classification

K. Shanthi¹; S. Manimekalai²

¹Research Scholar, PG & Research Department of Computer Science, Theivanai Ammal Collage of Women, Thiruvalluvar University, Tamil Nadu, India.

¹ softshanthi@gmail.com

²Head & Assistant Professor, PG & Research Department of Computer Science, Theivanai Ammal Collage of Women, Tamil Nadu, India.

² mamekaroshini@gmail.com

Abstract

Pap smear is a screening procedure for the cervical cancer. It is used to detect precancerous and cancerous in the cervix. High grade changes can be detected through a Pap smear test and the treatment can prevent the growth of cancer. This paper describes many features of single-cell images that is used to extract relevant and informative data. The features were used for the cervical cancer classification and their limitations are also discussed in detail.

Key-words: Pap-Smear, Cervical Cancer, Feature Extraction.

1. Introduction

The field of medical science has benefited greatly from image processing. It has made it possible to carry out a variety of processes in the field. For long years, researchers have been researching optimizations. The survey's findings on deep and machine learning have offered academics a lot of room to experiment with new ideas in this field [23]. Deep learning and image processing have been used to predict many types of tumors in the past. Cancer identification at an early stage is extremely difficult. The findings are inconclusive when it comes to cancer analysis. The Pap test, often known as a screening test, is used to diagnose cervical cancer. Finding the cancer is the first step in a cancer diagnosis. The next stage is to determine the cancer's kind or classifications. It also shows whether the patient has a high risk of developing cancer in the future. Some outcomes

may not require treatment and can be resolved on their own. These outcomes can usually be treated on their own. The results of a cell study that reveals abnormalities may necessitate further treatment.

The second most frequent malignancy in women is cervical cancer. In most cases, it begins in the uterus [2]. It's an easily detectable cancer that can be diagnosed with a Pap test. It can be present with vaginal bleeding, but it may not be noticeable until the disease has progressed [11].

The prime aim of this research is to categorize the properties of various types of cervical cancer cells. It is also aimed to examine the various strategies for cancer cell classification. The Pap test is a cutting-edge cancer screening method. It consists of a cell examination [3].

2. Related Work

A review of the literature revealed that statistical approaches have been used in various research on the prediction problem. The objective which we want to achieve is to use an existing image processing approach to build a process for identifying the properties of cervical cancer cells. This approach is mostly used to classify diseases based on their characteristics.

Andrew Ware., et al. [2] discussed the cervical cancer cataloguing from Pap-smears with the help of enhanced Fuzzy C-Means (FCM) algorithm and obtained an accuracy of (98.8%), which compares favourably to accuracies obtained in other studies such as deep convolutional network (97 percent), Ensemble method (97 percent). (96 percent). Chandranand presented a new method for feature extraction identification centred on the GLCM technique. They employed the Herlev Dataset and got a 94 percent accuracy [13]. The detailed survey of identifying the Cervical Cancer was proposed by Geetha S., et al for perceiving cervical cancer cells in their early stages [17]. K. Shanthi, S. manimekalai, and others [3] An accurate cervical image segmentation approach was developed utilizing Lagrange dual dictionary learning in a convolutional neural network (94.65 percent). Giada Mailoli and her colleagues [20] suggested an evolutionary approach for identifying characteristics in high-dimensional datasets. Manoj Sharma, et al. [12] used a genetic algorithm and an adaptive boosting technique to predict cervical cancer prognosis and found an accuracy of 80%. (94 percent).

3. Pre-processing

Pre-processing is the step of enhancing an image quality by removing noise and redundancy.

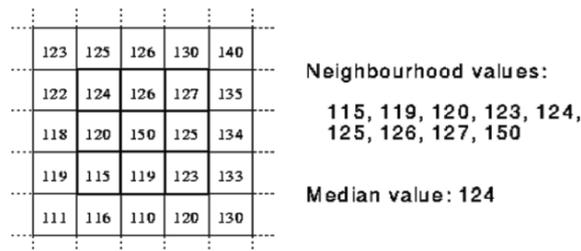
- i. Color Conversion is a process to convert the image from RGB into Grayscale.

- ii. Denoising images used to eradicate the noise occurred in the input image. In the proposed work, median Filter is used to get rid of the noises.

A. Median Filter

The pixel values were ordered according to intensity in median filtering. The median is the midway value, which is the output value. The differences of the mean and the extreme values is not shifted, and that is effectively removed when used with a median filtering scheme.

Figure 1- Median Filter



4. Feature Extraction

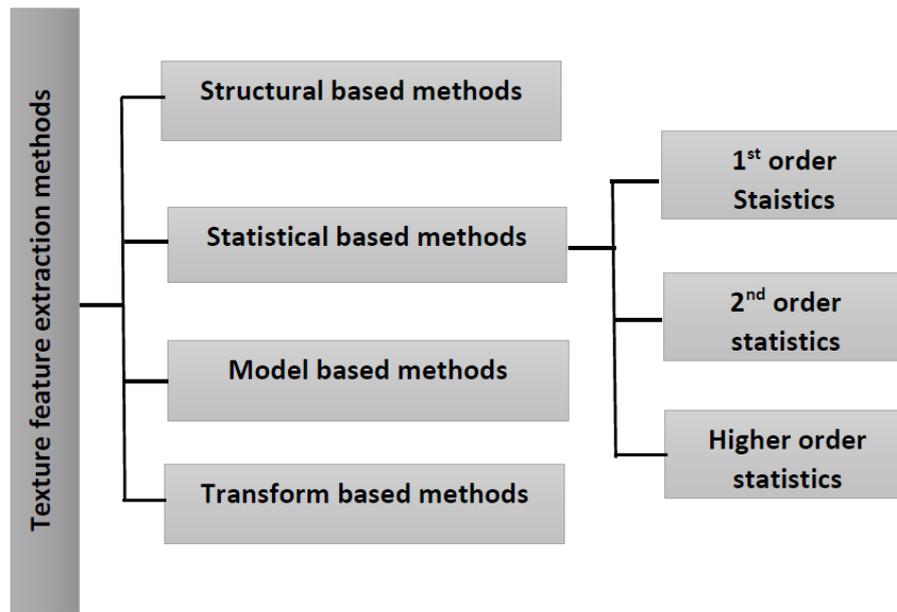
Feature extraction is the step to streamlining the quantity of data that is required to describe a high amount of data. The complexity of the analysis and the total number of variables required for the images classification that are presented were both significant. To minimize the error of false negatives and positive results, various approaches were used. [30].

5. Texture based Features

This method aims to find a way to represent the textures' characteristics in simple and unique form. This method is used to an image classification into regions of interest. The concept of a structural pattern is a simple and effective way to represent a recurring relationship pattern but, the statistical approach are used for estimating intensities in a region. It is used for general and practical applications. The texture features were computed basis on the statistical distributions of the observed intensities. A statistic is a mixture of three or more intensity points. It is divided into first-, second-, and higher-order categories.

Renu Bala., et al [25] discussed a survey on texture feature extraction method and comparing the feature results. Various methods were used to extract the texture features.

Figure 2- Texture Extraction Method



Processing an image consists of three main visual features: texture, color and shape. Texture is a visual pattern that has properties such as fine, contrast, and entropy. lbp, glcm, sgldm and glrlm are the methods were used to extract texture information from various sources.

6. GLCM

Yasha Singh., *et al* [13] discussed Algorithm for screening cervical cancer and obtain 93% accuracy using GLCM for the extraction of features. This methods are used to analyze images where the two pixels were related to a specified spatial region. The Co-occurrence Matrix technique is an integral technique that takes the frequency of a feature and then produces statistical measures from it. Some probabilities of low-level descriptors for combination of textures, colors, and shapes. RGB and GF textures were defined for the extraction of feature and frequency layout of accuracy.

The GLCM is defined by:

$$Pd(i,j) = n_{i,j} = \#\{f(m,n) = i, f(m+dx, n+dy) = j; 1 \leq m \leq M; 1 \leq n \leq N\}$$

- where n_{ij} is the number of occurrences of the pixel values (i,j) lying at distance d in the image.
- The co-occurrence matrix Pd has dimension $n \times n$, where n is the number of gray levels in the image.

For example, if $d = (1, 1)$.

2	1	2	0	1
0	2	1	1	2
0	1	2	2	0
1	2	2	0	1
2	0	1	0	1

$$\begin{array}{c} i \\ | \\ j \end{array} P_d = \begin{array}{ccc|c} 0 & 2 & 2 & 0 \\ 2 & 1 & 2 & 1 \\ 2 & 3 & 2 & 2 \\ \hline & 0 & 1 & 2 \\ & & & j \end{array}$$

There are 16 two of a kind of pixels in the image which fulfil this spatial separation. As, only three gray levels i.e. P [i,j] is a 3×3 matrix. This method is used to extract statistical texture features from second-order data. It is using the relationship between the neighboring pixels [33].

A GLCM is a hierarchy with the same rows and columns as the image. G_{d, I j} is the second-order co-occurrence matrix, where d is the distance between image values I and j and I and j is the direction. GLCM characteristics are estimated in four different directions with two distances with degrees of: 0 d, 45, 90, and 135. The total number of incidences of pair I and j, in the image is the value of G I j). The following equation, which uses distance values (dr, dc) for row and column, defines the co-occurrence matrix G. Contrast, cluster shade, energy, and other characteristics are included.

Features	Formula
contrast	$\sum_{i,j=0}^{G-1} (i-j)^2 P(i,j)$
Cluster shade	$\sum_{i,j=0}^{G-1} (i+j - \sigma_i - \sigma_j)^3 P(i,j)$
Energy	$\sum_{i,j=0}^{G-1} P(i,j)^2$
Sum of square variance	$\sum_{i,j=0}^{G-1} P(i,j) (i - \mu)^2$

7. Deep Learning Features

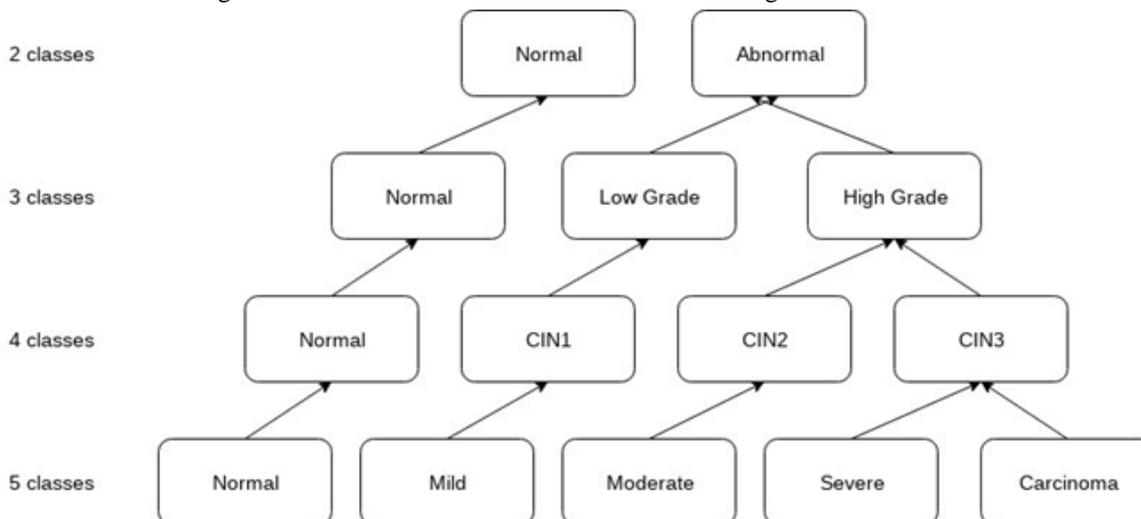
Faraz Faruqi, et al. [10] discovered the efficient utilisation of a Deep CNN for Malevolence Recognition and Cataloguing in Minuscular Uterine Cervix Cell Images, and found that it improved class-wise accuracy.

To perform well, deep neural networks require a vast amount of data. If the data is insufficient, the network can be supplemented using a variety of approaches. this strategy aids the network's efficiency to grasp features from a dataset of images. Despite the fact that several data

instances are formed from the same image, each has its own pixel distribution, resulting in the development of training data instances from various images. The nucleus' sizes and intensities are the essential elements that are used to classify the photos.

Catarina Barata, et al. [37] A convolutional neural network is an image recognition system. Its primary function is to recognize faces and objects. This feature allows you to compute the computed feature relationship without having to specify a lot of parameters [31]. Convolutional neural networks are the set of algorithms that are divided into three parts: convolutional, pooling, and FC. A CNN is defined as a neural network that processes its output in a specific area. CNN's basic building block is the Convolutional layer. It's an example of supervised learning algorithms that necessitate a lot of data for training. Its structural design can be described as a set of feed onward layers with convolution filters implemented. By summarizing the responses, these layers diminish or increase the map feature. Various sub sample layers and convolution sequences make up the CNN architecture. The CNN uses completely connected layers after the final sub-sample layer to convert 2D feature maps to 1-D vector [23].

Figure 3- Different Taxonomies of Cervical Cell Images Classification



Convolution neural networks can predict the correct features of images on the base features. The CNN model takes advantage of the various data to extract important features from cervical images. They have considered various sets of images for the study. The authors performed experiments on 3 sets of images with different taxonomies. They compared the output of the different sets of images and concluded that they achieves better results than those derived from the previous experiments.

8. Conclusion

The paper discussed various features of Pap-Smear Analysis. This section briefly explained the various pre-processing techniques used in Pap-Smear analysis. It also provided a comprehensive explanation on their importance in Pap-smear analysis. This carried paper work presents a review of the various features of the research study and provides guidelines for future research work.

References

- Pandian R, Lalitha Kumari S, Ravi Kumar DNS, “Analysis of feature extraction techniques using lung cancer image in computed tomography”, *Biomedical Research*, 79-81, 2020.
- Wasswa Williama, Andrew Wareb, Annabella Habinka Basaza-Ejiric, Johnes Obungolocha, “Cervical cancer classification from Pap-Smear using an enhanced fuzzy C-means Algorithm”, *Elesvier*, 23-33, 2019.
- K. Shanthi and S. Manimekalai, “An efficient cervical image segmentation method using lagrange dual dictionary learning in convolutional neural network”, *Annals of R.S.C.B*, 25, 1944-1957, 2021.
- Md Mamunur Rahaman, Chen Li, Xiangchen Wu, Yudong Yao, “A survey for Cervical cytopathology images analysis using deep learning”, *IEEE*, 8, 2020.
- Wan Azani Mustafa, Afiqah Halim, Khairul Shakir Ab Rahman, “A Narriative Review: Classification of Pap-Smear cell image for cervical cancer diagnosis”, *Tech Science Press*, 22, 2020.
- Xiaofu Huang, Ming Chen, Peizhong Liu, Yongzhao Du, “Texture feature-Based classification on Transrectal Ultrasound image for Prostatic cancer detection”, *Computational and Mathematical Methods in Medicine*, 2020.
- Min Ji, Lanfa Liu, Runlin Du, Manfred F. Buchroithner, “A Comparative Study of texture and Convolutional neural network feature for detecting collapsed building after earth quakes using Pre and post event satellite imagery”, *MDPI*, Vol. 11 6, 2019.
- Xiaofu Huang, Ming Chen, Peizhong Liu, Yongzhao Du, “Texture Feature-Based Classification on Transrectal Ultrasound Image for Prostatic Cancer Detection”, *Computational and Mathematical methods in Medicine*, 2020.
- Catarina Barata, M. Emre Celebi, Jorge S. Marques, “A Survey of Feature Extraction in Demoscopy Image Analysis of Skin Cancer”, *IEEE*, 14, 2016.
- Y. Jia, E. Shelhamer, J. Donahue, and S. Karaye Caffe, “Convolutional architecture for fast feature embedding”, *In Proc. of MM*, 675, 2016.
- Shanthi PB, Faraz Faruqi, Hareesha KS, Ranjini Kudva, “Deep Convolution Neural Network for Malignancy Detection and Classification in Microscopic Uterine Cervix Cell Images”, *Research Article*, 2016.
- H. Lin, Y. Hu, S. Chen, J. Yao, and L. Zhang, “Fine-grained classification of cervical cells using morphological and appearance based convolutional neural networks”, *IEEE Access*, 7, 71541-71549, 2019.

- Y. Promworn, S. Pattanasak, C. Pintavirooj, and W. Piyawattanametha, "Comparisons of pap smear classification with deep learning models," in *Proc. IEEE 14th Int. Conf. Nano/Micro Engineered Mol. Syst. (NEMS)*, 282-285, 2019.
- Y. Song, E.L. Tan, X. Jiang, J.Z. Cheng, D. Ni, S. Chen, B. Lei, and T. Wang, "Accurate cervical cell segmentation from overlapping clumps in pap smear images," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 288-300, 2017.
- K.H.S. Allehaibi, L.E. Nugroho, L. Lazuardi, A.S. Prabuwno, and T. Mantoro, "Segmentation and classification of cervical cells using deep learning," *IEEE Access*, 7, 116925-116941, 2019.
- F. Xing, Y. Xie, H. Su, F. Liu, and L. Yang, "Deep learning in microscopy image analysis: A survey," *IEEE transactions on neural networks and learning systems*, 29(10), 4550-4568, 2017.
- F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review," *IEEE reviews in biomedical engineering*, 9, 234-263, 2016.
- Y. LeCun, K. Kavukcuoglu, and C. Faret, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems. IEEE*, 253-256, 2010.
- Rahmadwati, Golshah Naghdy, Montse Ross, Catherine Todd and E. Norachmawati, "Classification Cervical Cancer using Histology Images", *Proceedings of Second International Conference on Computer Engineering and Applications*, 1, 515-519, 2010.
- J.A. Noble and D. Boukerroui, "Ultrasound Image Segmentation: a Survey", *IEEE Transactions on Medical Imaging*, 25(8), 987-1010, 2006.
- M.E. Plissiti, C. Nikou and A. Charchanti, "Automated Detection of Cell Nuclei in Pap Smear Images using Morphological Reconstruction and Clustering", *IEEE Transactions on Information Technology in Biomedicine*, 15(2), 233-241, 2011.
- C. Szegedy, W. Liu, Y. Jia. "Going deeper with convolutions", *IEEE CVPR 2015*, 1-9, 2015.
- A. S´aez, J. S´anchez-Monedero, P.A. Guti´errez. "Machine learning methods for binary and multiclass classification of melanoma thickness from dermoscopic images," *IEEE transactions on medical imaging*, 35, 1036-1045, 2016.
- Z. Yu, D. Ni, S. Chen, and et al., "Hybrid dermoscopy image classification framework based on deep convolutional neural network and fisher vector," in *IEEE ISBI*, 301-304, 2017.
- R. Gupta, A. Sarwar, V. Sharma, "Screening of Cervical Cancer by Artificial Intelligence based Analysis of Digitized Papanicolaou-Smear Images", *International Journal of Contemporary Medical Research*, 4(5), 2017.
- L. Zhang, L. Lu, I. Nogue, R.M. Summers, S. Liu. *DeepPap: Deep Convolutional Networks for Cervical Cell Classification*, 2018.
- Shikha Agrawal, Jitendra Agrawal, "Neural Network Techniques for Cancer Prediction: A Survey", *Procedia Computer Science*, 769 - 774, 2015.
- D. Lavanya, Dr.K. Usha Rani," Analysis of feature selection with classification: Breast cancer datasets", *Indian Journal of Computer Science and Engineering*, 2011.
- Shang, L., Yang, Q., Wang, J., Li, S., Lei, "Detection of rail surface defects based on CNN image recognition and classification", *Proc. Int. Conf. Adv. Communication. Tech*, 11-14, 45-51, 2018.

- S.M. Han, H.J. Lee, and J.Y. Choi, “Computer-aided prostate cancer detection using texture features and clinical features in ultrasound image,” *Journal of Digital Imaging*, 21(S1), 121–133, 2008.
- S.C. Tai, Z.S. Chen, and W.T. Tsai, “An automatic mass detection system in mammograms based on complex texture features,” *IEEE Journal of Biomedical and Health Informatics*, 18(2), 618–627, 2014.
- H. Liu, Y.S. Zhang, Y.H. Zhang, and H.E. Zi-Fen, “Texture feature extraction of flame image based on gray difference statistics”, *Control Engineering of China*, vol. 20, no. 2, pp. 213– 218, 2013.
- G. Castellano, L. Bonilha, L.M. Li, and F. Cendes, “Texture analysis of medical images,” *Clinical Radiology*, 59(12), 1061–1069, 2004.
- Chougrad, H., Zouaki, H. and Alheyane, O. “Convolutional neural networks for breast cancer screening: transfer learning with exponential decay”, *31st conference on Neural Information Processing Systems*, 2017.
- Yu, L.; Chen, H.; Dou, Q.; Qin, J, Heng, P.A, “Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos”, *IEEE J. Biomed. Health Inform*, 21, 65–75, 2017.
- Tian, Z. Liu, L. Zhang, Z. Fei, B. PSNet, “Prostate segmentation on MRI based on a convolutional neural network”, *J. Med. Imaging*, Vol. 8, 2018.
- Dong, C. Loy, C.C, He, K., Tang, X., “Learning a deep convolutional network for image super-resolution”, *In Proceedings of the European Conference on Computer Vision*, 769–776, 2014.
- Rehman, M., Khan, S.H, Rizvi, S.D, Abbas, Z, Zafar, A, “Classification of skin lesion by interference of segmentation and convolutional neural network”, *In Proceedings of the 2nd International Conference on Engineering Innovation*, 5, 81–85, 2018.
- Kallenberg, M, Petersen, K, Nielsen, M, Ng, A.Y, Diao, P Igel, C, Vachon, C.M.; Holland, K, “Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring”, *IEEE Trans. Med. Imaging*, Vol. 35, pp. 1322–1331, 2016.
- Dhungel, N, Carneiro, G, Bradley, A.P, “Deep structured learning for mass segmentation from mammograms”, *In Proceedings of the 2015 IEEE International Conference on Image Processing*, 2950–2954, 2015.
- Albayark, A, Bilgin, G, “Mitosis detection using convolutional neural network based features”, *In Proceedings of the IEEE Seventeenth International Symposium on Computational Intelligence and Informatics*, Budapest, 335–340, 2016.