ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

# Automated analysis of Histpathology images for cancer diagnosis – A review

**Gopalakrishnan T[1]*, Rajeesh J[2], Palanikumar S[3], and Narasimhan K[4]**

[1]Assistant Professor/Department of ECE, Noorul Islam Centre for Higher Education, Noorul Islam University, Thuckalay, Kumaracoil, Tamilnadu, India.
[2] Professor and HOD/Department of ECE, College of Engineering Thalassery,Kerala, India.
[3]Associate Professor, Dept. of Information Technology, Noorul Islam Centre for Higher Education, Noorul Islam University., Kumaracoil, Thuckalay,Tamilnadu,India.
[4] Assistant Professor, School of EEE,Department of ECE,SASTRA University, Thanjavur, Tamilnadu, India.

**ABSTRACT**

Cancer is the second most widespread disease in India responsible for maximum mortality with about 0.3 million deaths per year. All types of cancers have been reported in Indian population including the cancers of skin, lungs, breast, rectum, stomach, prostate, liver, cervix, esophagus, bladder, blood, mouth Early diagnosis and timely treatment reduces the death rate. Histopathology slide can be digitized and stored in digital form. Computer assisted analysis of histopathology images reduces the mortality rate, burden of pathologist and improves the accuracy of diagnosis. Image analysis algorithms together with machine learning techniques helps to build the automated system to aid the pathologist in taking the correct diagnostic opinion. In this paper brief analysis of algorithms used for preprocessing, segmentation and feature extraction followed by classification is explored. Further complete analysis literature is done and various issues has been identified which has to be addressed and need to be solved for building the robust automated system.

**Keywords:** Histopathology images, Staining, Segmentation, Features, Soft computing techniques, accuracy.

***Corresponding author***

## INTRODUCTION

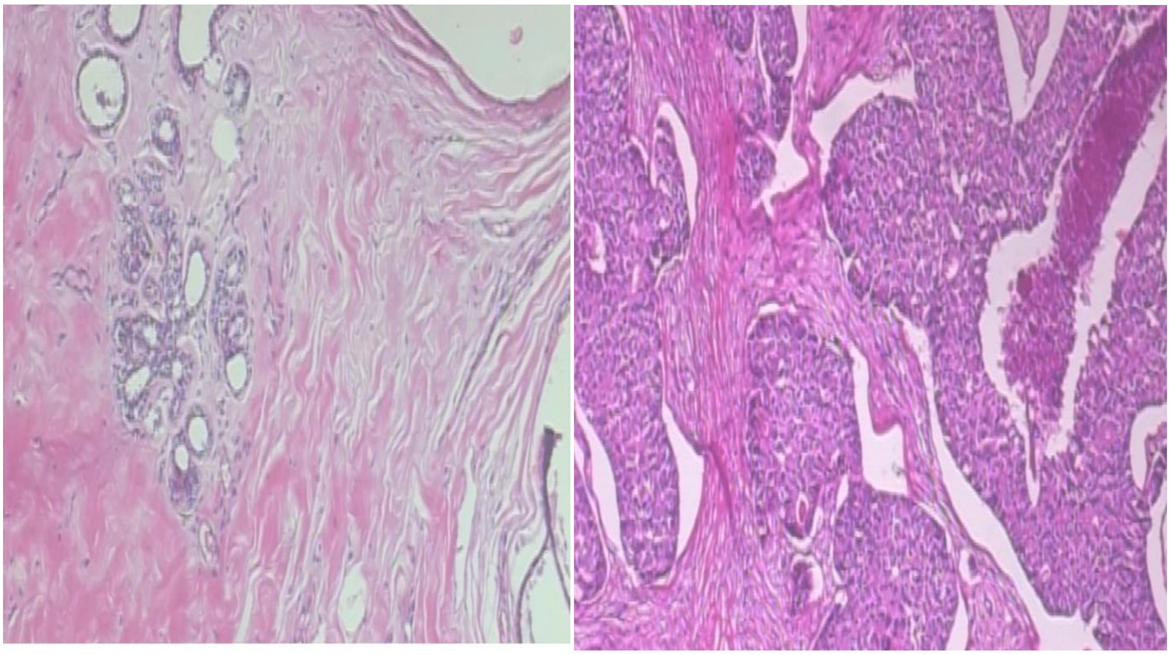
Histology is the scientific study of biological tissues. Histopathologists visually examine the regularities of cell shapes and tissue distributions, decide whether tissue regions are cancerous, and determine the malignancy level. Histopathology images helps for the study of composition of biological tissues using special staining techniques pooled with light and electron microscopy[1]. To visualize totally different elements of the tissue underneath a magnifier, the elements are dyed with one or more stains. The aim of staining is to reveal cellular components; counter-stains are used to provide contrast. Hematoxylin-Eosin (H&E) staining has been employed by pathologists for over 100 years. Hematoxylin stains cell nuclei blue, whereas fluoresceine stains protoplasm and animal tissue pink. Owing to the long history of H&E, well-established strategies and publications, there's a robust belief among several pathologists that H&E can still be the common follow over successive fifty years [2& 3].

**MATERIALS:**

Open source ground truth histopathology images are helpful to analyze the proposed algorithm. Breast cancer histopathology images are downloadable from the resource [46]. Total of 2480 Benign images and 5429 malignant images are available for free download. Other histopathological image dataset is available with the weblink mentioned in [47] &[48].

**Preprocessing:**

The five main stages in the preparation of histology slides are fixing Processing, embedding, sectioning and staining. Color normalization has to be performed to overcome the problems of variation in staining and scanning conditions. It is necessary to remove the noise that rises during staining process. Normally preprocessing steps include color normalization, denoising and enhancement. For tissue level feature extraction, to identify focal areas, thresholding process is carried out to remove noise[4 &5]. In cellular level segmentation, noise reduction has to be followed by segmentation to segment nuclei in tissue[6,7 &8]. Figure (1) shows the benign and malignant histopathology images at 40x magnification.



**(a)** **(b)**

**Figure(1) (a) Normal and (b) malignant histopathology image at 40x**

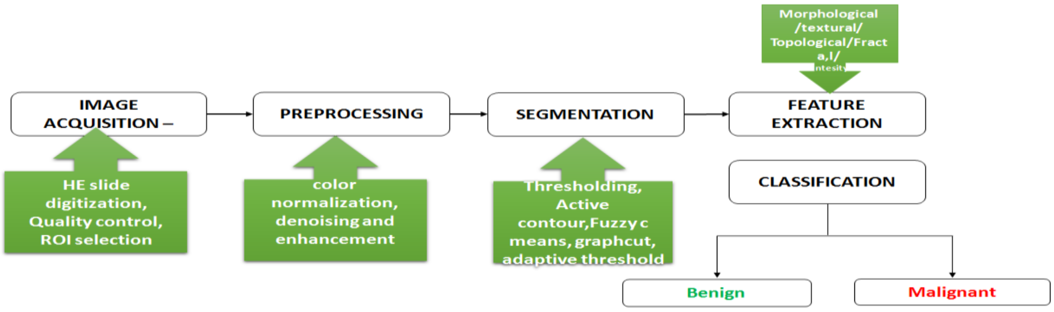



**Figure (2) Block diagram of the automated system for the analysis of Histopathology images.**

**Segmentation:**

In Histopathological image analysis, the main step is to segment the structures lymphocytes, cancer nuclei, and glands. To diagnose the severity of the disease it is necessary to know the size, shape, extent and other morphological appearance. To segment the nuclei normally the following techniques are used namely thresholding, fuzzy c-means clustering, and adaptive thresholding[9-13]. Generally segmentation techniques can be grouped under two heading namely region based approach and boundary based approach. Under Region based approach two predominantly used techniques and thresholding and clustering algorithms. Thresholding does not produce good result when there is a considerable variation in the image set. Thresholding helps to separate cells from background but not overlapping cells.[21 ,22]. Clustering algorithm based on unsupervised learning approach. Pixel based features are extracted namely color and textural then these features are used to form clusters. Then these clusters are classified either as cell or non cell[23 ,24]. Active contour model is widely used for boundary based segmentation. The success of active contour model depends on the initial selection of contour points. The energy function can be used to penalize the discontinuity in the curve[25 &26].

**Feature extraction:**

Different types of features are used for building automated system from histopathology images namely morphological features, textural features, intensity based features, topology features and fractal features.

**Morphological features:**

It helps to find the size and shape of the cell. The size is determined by the area , perimeter and radius of the cell. Shape is found by using compactness, symmetry, concavity, length of major and minor axes, roundness and smoothness[8,10].

**Texture features:**

Three methods that are widely used for textural feature extraction are gray level cooccurance matrix, local binary pattern and run length matrices[14,15]

**Intensity based features:**

Normally histogram of an image (color or gray) is used to extract features. Average, mean deviation, skewness and kurtosis of the pixel in each of the RGB plane is extracted as features[16 & 17]

**Fractal based features:**

Fractal geometry provide regularity and complexity of objects by analyzing self similarity. Fractal dimension and lacularity is used for cell analysis [18,19].

**Topology features :**

Structure of tissue is analyzed with the help of distribution of cell. Two familiar methods are voronoi diagram and Delaunay triangulation[20]

**Feature selection:**

Feature selection step is used to remove the redundant features and hence reduce the size of feature subspace. Three major categories of feature selection algorithms are filter, wrapper and embedded approach. Common measure in filter method includes mutual information, pointwise mutual information, Pearson product-moment correlation coefficient. Wrapper method uses predictive model to score feature subsets. Embedded method perform feature selection *as* part of the model construction process[27,28 &29]. Sequential forward selection and sequential backward selection method is widely used in histopathology image analysis.

**Classification:**

It is the crucial stage of the automated system, which helps to assign the class for the input histopathology as either benign or malignant. Among the machine learning algorithm the most popular methods, Artificial neural network, K-nearest neighbor classifier(K-NN), Support vector machine, Bayesian, Decision tree and Random Forest classifier. Support Vector generates a hyper plane to differentiate two sets of data[42]. K – nearest neighbor algorithm is a non parametric method used for classification or regression. An object is classified by a majority vote of its neighbours[41]. Decision tree uses predictive model which makes observation about an item to conclude about the item target value. Random forest classifier uses number of decision tree to get better classification rate[43]. Bayes classifiers are family of classifier used for classification based on posterior probability calculation[44].

**Evaluation of the classification system:**

In general classification system has two stages training and testing. One of the easiest ways to do, with the available data form two disjoint sets one for training and other for testing the system. More data used for training leads to better system design and more data used for testing leads to better appraisal of the system. K-fold cross validation and leave one out scheme used for the better evaluation of the classification system. Further sensitivity, specificity , accuracy used for the evaluation of the classification system.

True positive (TP) - cancer histopathology image classified as cancer
True negative(TN) – Normal histopathology image classified as normal
False positive(FP)- Normal histopathology image classified as cancer
False Negative(FN) – cancer histopathology image classified as normal

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \text{ Specificity} = \frac{TN}{TN+FP}, \text{ Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

**Table 1: Analysis of Literature review of Histopathology image**

| Author name & Reference Number | Type of cancer | Feature type | Classification efficiency |
|---|---|---|---|
| Jafari-Khouzani and Soltanian-Zadeh [35] | prostate cancer | Features derived from nuclear structures | 97% Accuracy |
| Weyn *et al.* [36] | Lung Cancer | Features from nuclear structure | 87.1 -96.8% Accuracy |
| van de Wouwer [37] | Breast tissue section | Nuclear feature | 67.1% Accuracy |
| Tabesh [38] | prostate | Nuclear feature | 81% Accuracy |
| Demir [39] | H&E stained brain tissue | Nuclear feature | 95.5 – 97.1% |
| Keenan [40] | H&E stained cervical tissue | Nuclear feature | 62.3%–76.5% |

## CONCLUSION

In this review, we explored the steps in developing a automated system for the diagnosis of malignant cancer from histopathology images. Starting from image acquisition, preprocessing, feature extraction, feature ranking, classification and evaluation of classification system is analyzed in detail. In each step all the methods explored in literature is discussed. Further the major observation the amount of data that has to be handled with histopathology image is large compared radiological images. Ensemble based learning system give better recognition rate and boundary based nuclear feature helps for better classification rate.

## REFERENCES


[1] http://www.ivyroses.com/HumanBody/Histology/What-is-Histology.php.

[2] Fox H. Is H&E morphology coming to an end? British Medical Journal. 2000;53:38.

[3] N. Stathonikos, M. Veta, A. Huisman, and P. J. van Diest, “Going fully digital: Perspective of a Dutch academic pathology lab,” *J. Pathol. Inform.*, vol. 4, Jun. 2013.

[4] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, A. Murray, Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa, IEEE T. Inf. Technol. B. 2(1998) 197-203.

[5] P.W. Hamilton, P.H. Bartels, D. Thompson, N.H. Anderson, R. Montironi, Automated location of dysplastic fields in colorectal histology using image texture analysis, J. Pathol. 182(1997) 68-75.

[6] W.N. Street, W.H. Wolberg, O.L. Mangasarian, Nuclear feature extraction for breast tumor diagnosis, IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology 1905(1993) 861-870, San Jose, CA.

[7] C. Gunduz, B. Yener, S.H. Gultekin, The cell graphs of cancer, Bioinformatics 20(2004) i145-i151.

[8] J. Gil, H. Wu, B.Y. Wang, Image analysis and morphometry in the diagnosis of breast cancer, Microsc. Res. Techniq. 59(2002) 109-118.

[9] S. Naik, S. Doyle, A. Madabhushi, J. Tomaszeweski, and M. Feldman, “Automated gland segmentation and Gleason grading of prostate histology by integrating low-, high-level and domain specific information,” presented at theWorkshop on Microscopic Image AnalysisWith Applications in Biology, Piscataway, NJ, 2007.

[10] P. S. Karvelis, D. I. Fotiadis, I. Georgiou, and M. Syrrou, “A watershed based segmentation method for multispectral chromosome images classification,” in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2006, vol. 1, pp. 3009–12.

[11] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, “Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer,” *BMC Med. Imag.*, vol. 6, p. 14, 2006.

[12] D. C. Fernandez, R. Bhargava, S. M. Hewitt, and I. W. Levin, “Infrared spectroscopic imaging for histopathologic recognition,” *Nature Biotechnol.*, vol. 23, pp. 469–74, Apr. 2005.

[13] L. E. Boucheron, Z. Bi, N. R. Harvey, B. Manjunath, and D. L. Rimm, “Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery,” *BMC Cell Biol.*, vol. 8, Suppl. 1, p.S8, 2007.

[14] R.M. Haralick, Statistical and structural approaches to texture, Proc. of IEEE. 67(1979) 786-804.

[15] M.M. Galloway, Texture analysis using gray level run lengths, Computer Graphics and Image Processing 4(1975) 172-179.

[16] F. Schnorrenberg, C.S. Pattichis, C.N. Schizas, K. Kyriacou, M. Vassiliou, Computer-aided classification of breast cancer nuclei, Technol. Health Care 4(1996) 147-161.

[17] Z.H. Zhou, Y. Jiang, Y.B. Yang, S.F. Chen, Lung cancer cell identification based on artificial neural network ensembles, Artif. Intell. Med. 24(2002) 25-36.

[18] S.C. Cross, Fractals in Pathology, J. Pathol. 182(1997) 1-8.

[19] Heymans, J. Fissette, P. Vico, S. Blacher, D. Masset, F. Brouers, Is fractal geometry useful in medicine and biomedical sciences? Med. Hypotheses. 54(2000) 360-366.

[20] B. Weyn, G. Van de Wouwer, S. Kumar-Singh, A. Van Daele, P. Scheunders, E. Van Marck, W. Jacob, Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis, Cytometry 35(1999) 23-29.

[21] J.-P. Thiran, B. Macq, Morphological feature extraction for the classification of digital images of cancerous tissues, IEEE T. Bio-Med. Eng. 43(1996) 1011-1020.

[22] R.F. Walker, P.T. Jackway, B. Lovell, Classification of cervical cell nuclei using morphological segmentation and textural feature extraction, Proc of the 2nd Australian and New Zealand Conference on Intelligent Information Systems (1994) 297-301.
[23] J.A. Hartigan, M.A. Wong, A k-means clustering algorithm, Appl. Stat. 28(1979) 100-108.
[24] P. Spyridonos, P. Ravazoula, D. Cavouras, K. Berberidis, G. Nikiforidis, Computer-based grading of haematoxylin-eosin stained tissuesections of urinary bladder carcinomas, Med. Inform. Internet Med. 26(2001) 179-90.
[25] K.-M. Lee, W.N. Street, A fast and robust approach for automated segmentation of breast cancer nuclei, Proc of the Second IASTED International Conference on Computer Graphics and Imaging (1999) 42-47, Palm Springs, CA.
[26] H.-S. Wu, J. Barba, An efficient semi-automatic algorithm for cell contour extraction, J. Microsc. 179(1995) 270-276.
[27] A. J. Einstein, H.S. Wu, M. Sanchez, J. Gil, Fractal characterization of chromatin appearance for diagnosis in breast cytology, J. Pathol. 185(1998) 366-381
[28] R. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, New Jersey, 1961.
[29] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.
[30] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, A. Murray, Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa, IEEE T. Inf. Technol. B. 2(1998) 197-203.
[31] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46(2002) 389-422.
[32] P.W. Hamilton, P.H. Bartels, D. Thompson, N.H. Anderson, R. Montironi, Automated location of dysplastic fields in colorectal histology using image texture analysis, J. Pathol. 182(1997) 68-75.
[33] F. Schnorrenberg, C.S. Pattichis, C.N. Schizas, K. Kyriacou, M. Vassiliou, Computer-aided classification of breast cancer nuclei, Technol. Health Care 4(1996) 147-161.
[34] B. Weyn, G. Van de Wouwer, S. et al. Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis, Cytometry 35(1999) 23-29.
[35] K. Jafari-Khouzani and H. Soltanian-Zadeh, “Multiwavelet grading of pathological images of prostate,” *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 697–704, Jun. 2003.
[36] B. Weyn *et al.*, “Automated breast tumor diagnosis and grading based on wavelet chromatin texture description,” *Cytometry*, vol. 33, pp. 32–40, 1998.
[37] G. Van deWouwer, B. Weyn, P. Scheunders, W. Jacob, E. Van Marck, and D. VAN, “Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei,” *J. Microscopy*, vol. 197, p.25, 2000.
[38] A. Tabesh, M. Teverovskiy, H. Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, “Multi feature prostate cancer diagnosis and Gleason grading of histological images,” *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1366–1378, Oct. 2007.
[39] C. Demir and B. Yener, “Automated cancer diagnosis based on histopathological images: A systematic survey,” Rensselaer Polytechnic Institute, Troy, NY, 2006.
[40] S. Keenan, J. Diamond, W. G. McCluggage, H. Bharucha, D. Thompson, P. Bartels, and P. Hamilton, “An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN),” *J Pathol.*, vol. 192, no. 3, pp. 351–362, 2000.
[41] Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". *The American Statistician*. **46** (3): 175–185.
[42] 1.Fabio A. Spanhol, Luiz S. Oliveira, Caroline Petitjean , and Laurent Heutte , “ A Dataset for Breast Cancer Histopathological Image Classification” ,IEEE Transactions on Biomedical Engineering,Vol .63 No. 7 ,pp.1455-1462, 2016.
[43] Cortes, C.; Vapnik, V. (1995). "Support-vector networks". *Machine Learning*. **20** (3): 273-297.
[44] Rokach, Lior; Maimon, O. (2008). *Data mining with decision trees: theory and applications*. World Scientific Pub Co Inc.
[45] Rennie, J.; Shih, L.; Teevan, J.; Karger, D. (2003). *Tackling the poor assumptions of Naive Bayes classifiers*.
[46] http://web.inf.ufpr.br/vri/breast-cancer-database.
[47] http://www.informed.unal.edu.co/histologyDS/.
[48] http://www.histology-world.com/photoalbum/thumbnails.php?album=52.