COMPUTER AIDED ANALYSIS OF HISTOPATHOLOGICAL IMAGES FOR FEATURES BASED CANCER DETECTION

THESIS SUBMITTED FOR THE DEGREE OF

Doctor of Philosophy

in

Biomedical Engineering

By

Anuranjeeta

Supervisor

Dr. Shiru Sharma

Co-supervisor

Prof. K.K. Shukla

SCHOOL OF BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
(BANARAS HINDU UNIVERSITY)
VARANASI- 221005
INDIA

Enrolment No. 330370

2017

Chapter 8: Conclusion and future work

8.1 Conclusion:

Worldwide, the occurrence of breast cancer is prominent and dominant in female subjects. Owing, to this deadly medical ailment and enormous diagnosis of breast cancer from histopathological slides in pathology laboratories, certain computer aided tools are essential for developing smoother, faster and effective strategy in its accurate diagnosis and for reducing the workload over the lab personals. Based upon this essential necessity, we have proposed the CAD based system in this thesis work. The proposed CAD system identifies and detects the benign and malignant breast cancer cells depending upon its unique feature characteristics. This automated approach will aid pathologists for faster, efficient and precise diagnosis of breast cancer cells.

The executive summary of entire thesis work based on the implementation of CAD system for detection of breast cancer cells from histopathology slides has been embodied here as follows:

Chapter 1 In this chapter a brief introduction about the present scenario regarding the examination of abnormalities in the tissue structure by the pathologists and challenges in this area for the detection of cancer has been provided. The essential need for automation of the procedure through a CAD system is addressed. In perspective of addressing the requirements for this necessity, the present chapter briefly represents the problem statement and motivation, aims and objectives, research contributions of the present thesis work. At the end of the chapter, the thesis organization section concisely highlights all the chapters present in the thesis.

Chapter 2 This chapter concluded that histopathological modalities have their own advantage, disadvantage, and properties. The problems present in histopathology images are artifact, low contrast, weak boundary and noise present due to uneven staining of the dyes. Furthermore, different preprocessing, segmentation and classification techniques are having their own advantage and disadvantage. Finally, a brief survey of different automated image analysis techniques for histopathological images has been concluded.

Chapter 3 concluded that the proposed approach Dynamic Stochastic Resonance (DSR) based technique utilizes the Discrete Cosine Transform (DCT) coefficient of image for successful enhancement and segmentation of breast cancer cells of histopathological images. The same algorithm has also been applied for cervix, prostate and ovarian cancer. This method helps in detection of cancerous cells accurately. Proposed DSR method is well capable of preserving the fine edge structure in the image as well as luminance and contrast of the images. Further, the segmentation results shows that, the proposed methodology performs better than the other conventional methods, as it obtained high Correlation, NPR, and low VI, GCE. DSR based segmentation obtained better results with 0.776 average Correlation, 0.979 average Normalized Probabilistic Rand (NPR) index, 0.011 average Global Consistency Error (GCE) and 0.185 average Variation of Information (VI). This method has improved the interpretability or perception of information in images to make a better analysis.

Chapter 4 concluded that morphological, intensity and texture features are able to extract the features from the segmented images from histopathology images of breast cancer for differentiation of benign and malignant cells. For experimentation purposes, UCSB image data set of 58 images (26 malignant and 32 benign) has been used. Total 30 features including (morphology -9, intensity-4 and texture-18) based were extracted from the segmented cells of the images for further classification purpose. Region prop and Gray Level Co-occurrence Matrix (GLCM) properties are used here to compute morphological, intensity and texture features. The quantitative analysis of morphology-based features is recognized well for the classification of cancerous and noncancerous cells. From, the overall results and observations, it can be concluded that the morphology, intensity, and texture based features provide a better difference of quantitative measure of area, perimeter, major axis length, minor axis length circularity and max intensity values between benign and malignant cells. It is observed that malignant cells have the considerably greater magnitude for computed features as compared to benign. Significant variation has been found in numerical values of features in the case of benign and malignant cells. These images will provide assistance to the pathologists to identify malignant cells. The results reported here have the potential for classification of normal and cancerous cells.

Chapter 5 utilizes the classification processes using a clinically significant set of morphology, intensity, and texture based features as used in chapter 4 of this thesis. Three classifiers such as Artificial Neural Network (ANN), k-Nearest Neighbour (k-NN) and Support Vector Machine (SVM) were trained and further used for the classification of cells into benign and malignant categories based on UCSB breast cancer dataset. Henceforth, from experimental results, it was observed that ANN classifier is performing better for all categories of test cases present in the selected test data. These categories of test data are single cells and group cells dataset. Among all categories of test cases, further, it was observed that the proposed method is performing better for single cell images type sample test cases. For group cell, dataset and single cell dataset the accuracy using ANN are 83.78% and 95.68% respectively. Considering other parameters, classifying with ANN for both the dataset provide an almost better quantitative result for sensitivity, specificity, BCR - measure, and MCC when compared to the other classifiers. Further, for classifying the dataset of a group of cells, SVM classifies better than k-NN, however, for the dataset of a single cell, the classification of k-NN is better than SVM. Six morphological features have been considered to plot ROC curves such as area, convex area, circularity, max intensity, minimum intensity and perimeter for each of the three classifiers, ANN, k-NN, and SVM. It is shown by the curves that the features like area, convex area, perimeter and maximum intensity were providing the best Area under Curve (AUC) values; however, the AUC values for circularity and minimum intensity were not so significant.

The Graphical User Interface (GUI) has been developed using MATLAB–GUI with ANN classifier for easy, fast and robust identification of benign and malignant cells with a good level of accuracy. It can be used for supporting medical activities in recognition of cancer.

Chapter 6 concludes the comparative study of different classifiers for detection and classification of benign and malignant cells of breast histopathological image based on all cells in the image were done. Morphological features are extracted from the segmented images based on all cells present in the image. A dataset of 70 histopathological images of benign and malignant tissues are selected. CLAHE is used for enhancement of histopathology images. MMT, SIOX, RATS, and TWS have been used for the segmentation proposed from Fiji open access free software for image analysis. TWS performs better in comparison to other methods. In other segmentation techniques cells are overlapping, however in TWS, there is no overlapping has been visualized. TWS based proposed approach is associated with a high value of PRI and low value of GCE and VOI in comparison to other methods. Proposed approach may be recommended for segmentation of images for cancer detection. Eight popular classifiers like MLP, LMT, Random forest, Rotation forest, SMO, Naïve Bayes, J-Rip, and PART have been used. Among all these classification methods Rotation forest classifier performs better differentiation between benign and malignant cells with the accuracy of 85.7% and with maximum BCR value 0.806 for 70 images. MLP, Random forest, SMO, LMT, Naïve Bayes, J-rip provide average results for all images. Further, PART classifier provides very low sensitivity and down accuracy rate as compared to rotation forest classifier for this data set. Ranking of the maximal relevance factor is derived for obtaining important features by applying Relief-F algorithms. It gives appropriate results by taking important features instead of taking a large no of features. It has been shown that selected important features can lead to improved classification accuracy and are able to better differentiate benign cells from malignant cells.

Chapter 7 concludes the comparison results of intensity evaluation, texture and morphological features of dysplastic changes in cell pattern such as enlarge size, irregular shape, and hyperchromatic nuclei based features of malignant cells as compared to benign cells in histopathological images of ovarian and breast cancer cells and its correlation with the several biochemical parameters. For determination of MDA level in serum of ovarian and breast cancer patients, LPO assay has been performed.

The study comprised of 55 controls, 75 ovarian and 75 breast cancer patients. The results reveal that MDA levels, CA-125, WBC are increased, while RBC, Haemoglobin (Hb), and platelets count are decreased in cancer patients as compared to normal. For comparing morphological changes in cells, 5 normal and 5 cancerous tissues have been evaluated for ovarian and breast cancers. Further, morphological operation of these features reveals that a better difference in area, perimeter, major axis circularity and max intensity values between normal and cancer cells. It is found that there is significant difference in size, shape and hyperchromatic nuclei in benign and malignant cells.

Further, it presents an overview of the biochemical test such as RBC count, WBC count, Haemoglobin level, platelets count and CA-125, and Malondialdehyde (MDA) of normal, breast and ovarian cancerous patients. MDA level in serum of cancer patients are found to be raised which show an increase in toxicity so thereby causing damage to the tissue. MDA can be used as a prognostic marker for cancer patients. Increased level of MDA in cancer leads to loss of PUFA in plasma membrane integrity of the cell was observed in the histopathological study. Most of the ovarian cancer is derived from epithelial cells, and ovarian cancer of epithelial origin has been asserted to produce abnormally high MDA. A combinatorial approach including image based tissue analysis as well as the biochemical examination is required to be established for cancer detection.

8.2 Scope for future work:

The research in the design and development of a CAD system involves a broad area of research in the field of image processing, computer vision, and pattern classification. Following research directions may be followed in future for the design of an efficient CAD system for cancer detection.

In the present work, we have worked with features based on morphology, intensity and texture features, however, still more features would be included in future in the algorithm for efficient differentiation between benign and malignant cells such as spectral and wavelets based features. Further, exploration with other features might aid in the deeper analysis of cancer cell morphology.

Research may be carried out to include other classifiers to classify the images in order to check the various accuracy levels. Expanding the size of the dataset will be utilized for classification model required for practical purposes. In future, overlapping cells may be classified by enhancing the intensity values.

Efforts may be done to design a CAD system for cancer detection which uses inputs from various imaging modalities such as different types of cancerous histopathology images. The concepts discussed in this thesis may also be extended to other applications such as cancer detection from cervix, prostate, colon cancer etc. Developed methods can also be used for electron microscopy images for the differentiation of benign and malignant cells of cellular structure more accurately.

Henceforth, the proposed method in this thesis may aid researchers and medical practitioners for cancer detection by using these state-of-the-art methods presented in this thesis.