COMPUTER AIDED ANALYSIS OF HISTOPATHOLOGICAL IMAGES FOR FEATURES BASED CANCER DETECTION

ABSTRACT

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

CANDIDATE'S DECLARATION

I, **Anuranjeeta**, certify that the work embodied in this Ph. D. thesis is my own bonafide work carried out by me under the supervision of **Dr. Shiru Sharma** and the co-supervision of **Prof. K.K. Shukla** for a period of **6 years** and **4 months from March 2011 to June 2017** at **School of Biomedical Engineering, Indian Institute of Technology (Banaras Hindu University), Varanasi.** The matter embodied in this Ph.D. thesis has not been submitted for the award of any other degree/diploma.

I declare that I have faithfully acknowledged, given credit to and referred to the research workers wherever their works have been cited in the text and the body of the thesis. I further certify that I have not willfully lifted up some other's work, para, text, data, results, etc. reported in the journals, books, magazines, reports, dissertations, theses, etc., or available at web-sites and included them in this Ph. D. thesis and cited as my own work.

Date:
Place: Varanasi **(Anuranjeeta)**

CERTIFICATE FROM THE SUPERVISOR/CO-SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Supervisor Signature	Co-Supervisor Signature
Dr. Shiru Sharma	Prof. K. K. Shukla
(Assistant Professor)	(Professor)

(Signature of the Coordinator of the School with seal)
School of Biomedical Engineering
IIT (B.H.U.), VARANASI.

Abstract

The present thesis is focused on the development of a computer aided diagnosis (CAD) system to investigate the benign and malignant cells using the haematoxylin and eosin stained histopathological images of cancer. Specific tools were used for the identification of cancer: (1) pre-processing, (2) segmentation, (3) feature extraction and (4) classification of histopathological images. It aims to reduce subjectivity in the diagnostic process and assist in making diagnostic decisions more accurate and consistent. Cancer has turned out to be the leading causes of mortality in the world there by creating the need for more awareness globally. Among the developing countries, India suffers leading cause of death from cancer.

Nowadays, CAD system has become a part of routine clinical detection methods for cancer diagnosis using histopathological images at various screening centre and hospitals. Henceforth, it has research in histopathological imaging and diagnostics. This innovation has reduced the workload in pathological laboratories and increased the quality and quantity of the diagnosis. Usually, all the diagnostic features are confined within the tissue sample slides. The state of art facilities is utilized for determining those atypical features and their degree of invasion towards other parts of the cellular environment.

Generally, pathological laboratory experts interpret biopsy samples through microscopes for detection and classification of atypical features. It is a manual and tedious job. Furthermore, the interpretations from the different pathologists differ based upon their observations, training and analysis expertise. The initiation of digital pathology has provided automation in image analysis techniques. As compared to the individual screening of histopathology slides, it is costly, tiresome, time consuming and contains sufficient probability of human errors. Due to the limitation of manual observation and the challenges faced by the pathologist to provide both accurate and uniform evaluation for the enormous number of histopathological images generated in widespread screening within a short duration, there is need of CAD based system for accurate cancer diagnosis and classification. The automated analysis provides a quantitative result of each particular cell in the image to demarcate between benign and malignant categories with higher sensitivity and accuracy.

Pathologists face difficulty in cell image detection as uneven dye causes overlapping, blurring, artifacts, low contrast as well as weak boundary detection. For enhancement and segmentation, Dynamic Stochastic Resonance (DSR) approach has the advantage of identifying target objects of noisy and low contrast images. Based on the comparative study, the DSR proposed methodology performs better than the other conventional methods, as it resulted in high Correlation, Normal Probability Rand (NPR), and low Variation of Information (VI), Global Consistency Error (GCE).

Further, the segmented cells are used for features extraction. For this, morphological, intensity and texture based features have been investigated. After a morphological operation, this feature reveals a better difference in area, perimeter, major axis length, circularity and maximum intensity values between benign and malignant cells. Significant differences are found in size, shape and architecture of cells in benign and malignant.

Finally, classification is performed on breast cancer single and group cells dataset using supervised classifiers Artificial Neural Network (ANN), k-nearest neighbour classifier (k-NN) and Support Vector Machine (SVM) generated features which are used to classify cells or region into benign and malignant categories. It is found that the classification accuracy achieved by the single cells dataset is better than the group cells. Furthermore, it is established that ANN provides a better result for both datasets than the other two k-NN and SVM.

The process of handling MATLAB script for each and every image separately is problematic and hazardous. A Graphical User Interface (GUI), named as CELL CHECK (v1.0) has been performed using MATLAB–GUI with ANN classifier to make the process easy, fast, user friendly and robust identification of benign and malignant cells with a good level of accuracy. It can be used for supporting medical activities in cancer diagnosis.

The proposed method of the CAD system for the classification of benign and malignant cells provides better accuracy than the other existing methods. The various performance metrics accuracy, sensitivity, and specificity, has been calculated. The statistical measure regarding sensitivity and specificity has been obtained by calculating the area under the Receiver Operating Characteristic (ROC) curve.

It is believed that performance indicators of classifiers are as important and crucial as the classifier algorithms themselves since these parameters allow us to truly measure the success of our approach.

Comparative study of different classifiers for histopathological image of breast cancer classification and detection using morphological features based on all cells present in the image Ranking of the feature was done by applying Relief-F algorithms. The efficacy of the feature and classification methods based on the feature of benign and malignant cells have been examined and presented

Comparative evaluations of morphological features of dysplastic changes of cells such as enlarge size, irregular shape, hyperchromatic nuclei, etc. and its correlation with biochemical parameters such as Malondialdehyde (MDA) level is also presented. MDA levels in the serum of cancer patients are higher range which shows the increase in toxicity causing damage to the tissue. MDA can be used as a prognostic marker for cancer patients. Thus, a combinatorial approach including image based tissue analysis as well as the biochemical examination is required to be established for cancer detection.

The automated identification of cancerous cells by histopathological images helps in alleviating the above mentioned issues and provides better results. The important features based approaches are used for the identification of disease. Therefore, improvement in CAD performance increases the diagnostic options and the cure is more likely to be selective. Furthermore, to help the pathologist in screening a large number of images, use of a CAD system may be helpful in exact prognosis free from human error based analysis. CAD system for cancer detection and diagnosis can also be used as the second opinion by the pathologist.

Henceforth, the present thesis work would increase the accuracy of diagnosis by the design of CAD tool for analysis of images so that early, accurate and faster screening of cancer can be achieved. With the proposed solutions for different problems in CAD system are presented in this thesis, we believe that it will provide pathologists useful tools to assisting them with the cancer diagnosis task.

Keywords: Cancer, Histopathological, Enhancement, Segmentation, Features, Morphological, Classification.