SYSTEM FOR CLASSIFICATION AND STAGING OF ACUTE LYMPHOCYTIC LEUKEMIA

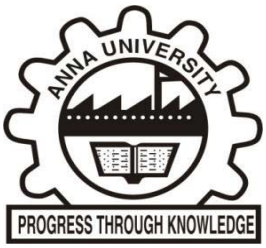
A THESIS

Submitted by

SRIRAM S

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY



FACULTY OF INFORMATION AND COMMUNICATION ENGINEERING

ANNA UNIVERSITY

CHENNAI 600 025

APRIL 2018

ABSTRACT

The Inspection of microscopic Blood cell Images provide important qualitative and quantitative information regarding the presence of Hematic Pathologies.

Neoplastic proliferations of white blood cells are most important group of Leukocyte disorders termed as Leukemia. Proliferation of white blood cells which do not mature leads to accumulation of useless cells which take up more and more bone marrow space at the expense of normal haematopoietic elements termed as Acute Lymphocytic Leukemia (ALL).

The most common diagnostic group of childhood cancers worldwide and in India are the Leukemias. Out of the childhood Leukemia, ALL constitute about 75-80% throughout the world. There are nearly 25,000 children diagnosed with cancer in India every year and around 9000 of these have Leukemia. Even with these conservative estimates, there would be 90,000 children with Leukemia in a decade in India. Acute Leukemia can progress quickly and if not treated would probably be fatal within a few months.

Analysis of microscopic Blood and bone marrow Images by human experts tends to serve as an important diagnostic procedure with respect to ALL. Although it is prone to disadvantages like subjective interpretations, less accurate diagnosis and slowness it is advantageous than recent and accurate methods in terms of simplicity and cost effectiveness necessary for developing nations like India.

So an effort has been taken in this thesis to propose a system for automated detection and classification of ALL by using Image Processing and Fuzzy Based Techniques.

First, as part of this system, a golden standard reference is created for nucleus segmentation of lymphocytes applying manual segmentation by a human hematopathologist. Two automated Image segmentation methods namely the K-means clustering and Fuzzy C-means clustering method were employed for Lymphocyte nuclei segmentation and their results were compared with golden standard reference output by using suitable metrics such as Jacard coefficient, Tanimoto's index, Nucleus Segmentation error and Computation time to choose the best possible segmentation scheme for selected database. This proves to be a very crucial step as the accuracy of this step influences the total accuracy of the entire system. The experimental results show that the Fuzzy C-means clustering method outperformed the conventional K-Means clustering segmentation method with respect to Nucleus segmentation of lymphocytes in more number of tested samples.

As the second part of the work, a set of features are extracted from the segmented nucleus. This main set of features extracted may be classified as Global Co-occurrence matrix features, Gabor features and Tamura Features.

In the third part of the proposed system, instead of feeding the entire set of features extracted from the segmented nucleus to the selected neural classifiers, a Java based data mining tool named WEKA is used to select features that contribute to a greater extent in classification of a lymphocyte nucleus into a normal or an abnormal one. Two different algorithms in WEKA namely the information gain method and the symmetric uncertainity algorithm were used to select the extracted features from the full set of extracted features. Also in this part of the system, different combinations of the features along with three different neural network classifiers namely : the Naïve Bayesian classifier, the Back propagation classifier and the support vector machine classifier were checked for

efficiency in classification of lymphocytes with metrics such as accuracy,training time and testing time.It is experimentally found that the features selected by symmetric uncertainity algorithm of WEKA when fed to the support vector machine classifier gave greater accuracy of about 95% with less training and testing time than compared to other classifiers and feature selection algorithms used.

In the last part of the proposed system, a Mamdani Fuzzy inference system was designed by framing the associated inference rules used so that it may classify the abnormal lymphocyte nuclei into Level 1 ,Level 2 or Level 3 as prescribed by the French-American-British (FAB) classification systems for ALL.

Thus,a complete system has been developed so that it can automatically confirm the presence of ALL from an image of the Peripheral blood smear and also predict the stage of ALL according to FAB so that the system can be used as a computer aided decision support system for aiding the physicians to take appropriate clinical decisions while treating patients suffering from ALL. The system has used images from the ALL-IDB2 dataset and is yet to be tested on real time images which can be seen as a perspective for future research.