Classification of Acute Lymphoblastic Leukemia in Blood Cell Images Using Machine Learning

Shravan Choudhary, Ji Ye Chun, Phillip Galonsky, Haoran Zhang

Introduction / Background

Acute lymphoblastic leukemia (ALL) is a type of cancer where the bone marrow produces too many immature lymphocytes. ALL is most common in young children and progresses rapidly, so quick diagnosis is critical for ALL patients to receive timely treatment. Among various diagnosis methods for ALL, the microscopic analysis of blood cells is the most economical and has the advantage of being relatively non-invasive. However, microscopic analysis is time consuming and requires the supervision of a medical expert. Moreover, the results of such analysis are limited by their subjective nature and reliance on the expert's skill. Over the last decade, various machine learning methods have been implemented for the diagnosis of ALL to overcome these shortcomings. For feature extraction, most studies use the combination of shape, texture, and color features as descriptors [1-3]. The most frequently used classifiers are support vector machine (SVM), decision tree, K-nearest neighbors, naive Bayes [1], multilayer perceptron (MLP) [2], and convolutional neural network (CNN) [3].

Methods

We plan to utilize the publicly available C-NMC dataset provided for free online by the The Cancer Imaging Archive at https://doi.org/10.7937/tcia.2019.dc64i46r. The dataset contains 15,114 microscopic images of cells from a total of 118 individual subjects, of which 69 have ALL and 49 are healthy. The images have been processed so that they contain only cells on a black background. The dataset is divided into a training set and a preliminary testing set for which the ground truth for each image has been marked by an expert oncologist. We propose to extract three types of features from the images: morphological, color, and texture features and use support vector machine (SVM) to classify malignant lymphocytes from normal cells.

Potential Results

Academic papers have reported an average accuracy above 90% [1-3], with the best results achieving an accuracy around 97% [2]. Thus, we aim to reach at least 90% accuracy with our approach.

Discussion

The best possible outcome for our project is a highly accurate (> 97%) model that classifies blood cells as either ALL or normal. Although a number of accurate models have already been developed to solve this classification problem, the concept of using machine learning to screen cells for cancer is still relatively new. Our project will hopefully contribute to the growing body of classifiers for ALL. Thus, the value of our project resides in the value of having multiple methods for classifying ALL blood cells: in having distinct models perform the same classification task, some methods may compensate for shortcomings in others, catching cancer where others do not.

Referenced Papers

[1] L. Putzu, G. Caocci, and C. Di Ruberto, "Leucocyte classification for leukaemia detection using image processing techniques," *Artif. Intell. Med.,* vol. 62, pp. 179-191, Nov. 2014.

[2] I. Vincent, K. Kwon, S. Lee, and K. Moon, "Acute lymphoid leukemia classification using two-step neural network classifier," in *21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, 2015, pp. 1-4.

[3] S. Rajpurohit, S. Patil, N. Choudhary, S. Gavasane, and P. Kosamkar, "Identification of acute lymphoblastic leukemia in microscopic blood image using image processing and machine learning algorithms," in *Int. Conf. Advances in Computing, Communications and Informatics*, New York, 2018, pp. 2359-2363.