Leukemia Diagnosis in Blood Microscopic Images Using Local Binary Pattern and Supervised Classifier

Dr.R.Suganya¹, B.BrindhaMalar², Kiruthika M³, S.Leelavathi⁴

¹Assistant Professor, ²³⁴ Students, Department of Information Technology,
Sri Krishna College Of Technology, Coimbatore-641042

ABSTRACT
Acute lymphoblastic leukemia (ALL) is a group of hematological neoplastic during childhood that gets characterized by a huge number of lymphoid blasts in the human blood. It makes 80 percentage of childhood leukemia. It occurs between the ages of 3-7 years. Diagnosis of this disease is really hard. It may end up in wrong diagnosis. It has signs and symptoms similar to other disorders. The effective way of diagnosing leukemia is examining the stained blood smear or bone marrow aspirate. Techniques that can be used for identification of particular type of leukemia are fluorescence in situ hybridization, immunophenotyping, analysis of cytogenic and cytochemistry. Since the above methods are not that much affordable and also it takes more time to complete the process, there is a need of an automated detection of leukemia. Important features that can be useful for detecting the leukemia are nucleus shape and texture. In the existing paper, hausdorff dimension along with contour signature is implemented for grouping a lymphoblastic cell nucleus. Neural network classifier can be used for classification.

Keywords: Leukemia, HBC.

I. INTRODUCTION
Blood is a liquid that is very important for humans, because it serves as a means of transporting substances such as oxygen, material metabolism, the body’s defense from germs, regulating acid-base balance, and has many other uses to support life. Circulating blood is an indication of the condition of the human body in a healthy condition or sickness. One of the blood diseases is white blood disease (granulopoetic system) caused by abnormal growth of WBC. The objects captured by the microscope will be processed using image processing. From the experimental results, the original image was obtained from a microscope. The process of diagnosing a disease that is fast and precise is necessary in connection with the accuracy of the data to determine program policies in the prevention and cure of diseases, determine the right and correct treatment, and treatment evaluation. In addition, the development of technology in pattern recognition, which generally aims to recognize an object by extracting important information contained in an image, can help detect an abnormality in the human body through the image produced by the scanner that is used in this case. By researchers as a supporting tool for diagnosing cancer of white blood cells (leukemia). Abnormalities contained in the
blood, can diagnose diseases suffered by humans. Along with technological developments, blood image retrieval with a scanning electron microscope can be done. In the process of recognizing this disease, a system that has acquired a blood image will do the pre-processing process and make the image into median blur. By utilizing the RGB image changes to HSV model the characteristics of each image will be adapted. This project is a combination of two previous studies which will identify two types of leukemia using a method that has been used to identify one type of leukemia with an accuracy rate of 93%. The resulting blood image can be used for disease recognition through an image processing process using certain method to detect a blood disease. In this study will be used method of recognition of blood disease through blood images as input for image processing, and the results of image processing will be analyzed using the if else branching method. With the results of this research, it is expected to be able to support healthcare kiosks where these tools are used to check human health conditions.

II. LITERATURE SURVEY

KNN algorithm for detecting breast cancer was specified by Tobias christiancahoon in 2000. He also illustrates that supervised and unsupervised methods have higher discrepancies in predicting accuracies. Mean and standard deviation can be added to improve accuracy. In 2012, K. Anuradha explained that image quality can be improved by using ultrasonography and optical imaging. She also states that methods used for brain and breast cancer cannot be used for oral cancer.

In 2013, Mussaratyasmin states that common type of cancer is breast cancer. She explains various methods for detecting cancer. Due to its negative approach mammography it is replaced by biopsy MRI scans. Dr. M.Sangeetha has illustrated the innovative techniques towards the recognition of carcinoma using classification and regression technique.

III. METHODOLOGY

This section represents the design and manufacture of systems used for leukemia identification systems based on human blood cells. Input image is original image that was obtained from a microscope. Preprocessing is used to improve image quality, segmentation is used to separate image objects taken with the background. Extraction feature is used to find leukemia features. After getting the features, we can detect leukemia.

Pre-Processing

To improvise the standard of the original image and facilitate the next processing step, pre-processing needs to be done.

Median Filtering
Lab Space Conversion
Lloyd’s Clustering Algorithm
Median Filtering

White blood cancer cells will be repaired using the median filtering method to get a better image. In OpenCV there is a separate library for image improvement using the median filtering method. By using the "cvmedianBlur" function and entering the
matrix value parameter from the pixel of an image, the system will perform the repair process automatically. In this study I used a pixel matrix value of 7 which represents the value of the pixel matrix that will be searched for by its middle value. The average filter pixel value in the results image is obtained by using equation.

\[ F(x, y) = \text{result pixel filter value.} \]
\[ G(x, y) = \text{input image pixel} \]
\[ m, n = \text{average matrix size} \]

Lab Space Conversion

In 1976, the International Commission on Illumination (CIE) defined the CIELAB color space (also known as CIE L*a*b* or sometimes abbreviated as simply "Lab" color space). The colors are expressed by three values: L*, a*, b*. The CIELAB is defined to produce the equal amount of change in the numerical values as of the visually predefined change. It is designed in that way so that the three parameters can be measured, the space is also a 3d real number space that allows infinite amount of possible colors. Practically, the space is usually mapped to a 3d integer space of digital representation. Thus the values of L*, a*, b* are absolute with a defined range.

Lloyd’s Clustering Algorithm

Finding the solution is unfortunately NP hard. The procedure alternates between two operations. (1) Once a group of centroids is out there, the clusters are updated to contain the points closest in distance to every centroid. (2) Given a group of clusters, the centroids are recalculated because the means of all points belonging to a cluster. The two-step procedure continues until the assignments of clusters and centroids not change. The two-step procedure continues until the assignments of clusters and centroids not change. As already mentioned, the convergence is guaranteed but the answer could be an area minimum. Practically the algorithm has been processed for multiple numbers of times and has been averaged. For the starting set of centroids, several methods are often employed, as an example random assignation.

IV DETECTION OF IMMUNE MAGNETIC BEADS AND CELLS AND COUNTING PROCESS

In order to detect cells in the microscopic images, computer vision algorithms are employed and automated cell detection and quantification methods for 20× and 40× images are proposed. The images contain immune magnetic beads and cells in various shapes and sizes. Cells in the images might appear as either isolated single cell, covered by beads or smaller fragments. When a cell or cell fragment isn't bound by any beads, its boundary and characteristic inner texture are often observed. However, when it's partially covered by one or more beads the bound beads might obstruct some a part of the cell boundary.
Another case is full coverage of cells by beads. During this case, a cluster of beads appears within the image.

Convolutional Layer: To explore many filters on the input image convolution layer is responsible. In this CNN we have used 32 feature maps with the size of 3x3. Sliding is used to apply convolution filters to the image. The filter values were determined randomly. We used 2 convolution layers to avoid overfitting.

Max-Pooling Layer: To decrease the dimension of the image max-pooling layer is being used as it focuses on the important features of the image. In our network, we used a max-pooling layer with the size of 2 × 2. We doubled the number of this layer, as well.

Flatten Layer: This layer is used to reduce the 2d image into 1d array as it can be used as a node for connected networks.

Fully Connected Network: This part was a naive connected full forward network that consisted of one input layer (the flattened layer in our case), a hidden layer, and an output layer. Here, the hidden layer is consisting of 128 nodes that have 10 percentages of Dropout then Batch-Normalization to minimize overfitting. Due to a simple calculation, the ReLU activation function had been implemented. In the output layer, we setup two types of optimizers SGD (stochastic gradient descent) and ADAM optimizer one type at a time. We added 5 output nodes (each node represents each leukemia type and HEALTHY samples) and all of them were controlled by a SoftMax activation function.
Experiment 1: The first experiment examines the capacity of a neural convolution network. This is purely a method of grading binary images that we do not have to use methods of image transformation. The ALL-IDB dataset consisted of the original ALL subtype and HEALTHY samples with 144 training samples for each class and 35 test samples for each class. We also applied a batch-normalization which was responsible for minimizing the amount of performance fluctuations.

Experiment 2: In this experiment, we considered making a binary classification with the same 2 classes in Experiment #1, where ALL and HEALTHY image transformation techniques were used. As a result, we used 980 samples as training data and 245 samples as test for both.

RESULTS TABLE

Accuracy and loss results for each experiment

<table>
<thead>
<tr>
<th>EXPERIMENT NAME</th>
<th>ACCURACY</th>
<th>LOSS</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary classification ALL and HEALTHY</td>
<td>99.65%</td>
<td>81.74%</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>99.92%</td>
<td>89.35%</td>
<td>0.002</td>
</tr>
</tbody>
</table>

VII EVALUATION PROCESS

To evaluate our model, we selected 2 main evaluation metrics that were used to measure the performance of any neural network. First, accuracy was the number of correctly classified image samples among the other samples. There are two types of accuracy for neural network: Training accuracy has been used to detect the
Performance of the model during training, and Validation accuracy training is used to detect the performance in classifying the unseen model. Secondly, the loss is the metric that focuses on calculating the error of prediction, which is used to adjust the weights of neural network nodes. The metric is calculated during the training and validation process. The dataset is divided into two parts, 70 percentages is for training and 30 percentage is used for validation, before applying any image transformation techniques. Then we applied the image transformations for both parts to increase the number of data samples. Large training dataset is used to work with deep convolution network. Because small training dataset may lead to over fitting problem. We have used data augmentation to solve this problem. The larger the training set the more accurate the CNN capacity is. Although the data augmentation may be used, it may increase the noise in the original dataset to overcome this, batch normalization and fine tuning of dropout can be used. But it may occur due to the similarity in the structure of blood cells. Multi classification percentage is dropped to 81.74v percentage. This clearly explains the drop in performance and also states that adding data augmentation leads to increase the noise in original data set.

VIII OUTPUT GRAPH

IX CONCLUSION

Leukemia is one of the terrific types of cancers that can occur to human. Blood smear analysis is commonly used to detect this kind of cancer. Here, we have used a different methodology that can detect four types of leukemia subtypes by using a microscopic blood image using convolution neural network. We have used data augmentation methods to overcome the over fitting problem to deal with the limited number of sample issues. We have obtained a result that shows, using CNN we have gained an outcome that can overcome other machine learning algorithms. That is 88 percentage of accuracy has been gained in one among the types of leukemia (ALL and Healthy samples) and 81 percentage of accuracy has been gained in the overall leukemia subtypes. It is costly throughout the classification of the medical image. But it is stable enough to stand amongst the overall classification process. The future of this system is using hybrid deep learning algorithm and recurrent neural network and also adding extra samples to the dataset.

REFERENCES


