DETECTION OF LEUKEMIA USING SOFT COMPUTING TECHNIQUES

Vishnu Sharma
Reg-no - 11805069
BTech Computer Science Engineering
Lovely Professional University

I. ABSTRACT

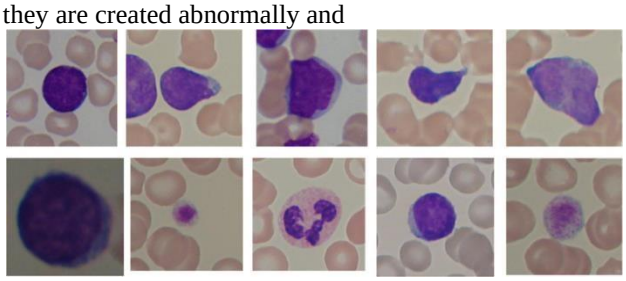
Leukemia is caused by excessive production of leukocytes in the bone marrow of a person, and image-based detection of malignant WBCs is an important step for its detection. Convolutional Neural Networks (CNNs) are present currently for this type of image classification but their computational cost for training and deployment is high. There are different models that are being used for detection of leukemia now.
One of the Important technique is Feature extraction in which multidimensional vector is extracted from an image, where each dimension represents the attribute of the image that carries information that is useful in classification of the image. Different attributes like average color coordinates, area of nucleus and cytoplasm , average color co-ordinates and number of pixels in the nuclear perimeter are used. Classification techniques like K-Nearest Neighbors, Decision Tress, logistic Regression, Random Forest, Support Vector Machine can also be used for detection.

Keywords—Leukemia, feature extraction & clustering, Classification

II. INTRODUCTION

Mainly three types of cells are found in blood: red blood cells, platelets and white blood cells. Red blood cells transport oxygen from the heart to all tissues, and carries away carbon dioxide. They make up to 50% of the overall volume of blood. White Blood Cells are responsible for immunity, as they are the main defense of the body against infections and diseases. Therefore there is rapidly increasing demand for classification of WBCs. WBCs are categorized into two types, according to the appearance of the cytoplasm. The first type is Granulocytes that includes Basophils, Eosinophils and Neutrophils. The second type is Agranulocytes that includes Lymphocytes and Monocytes.

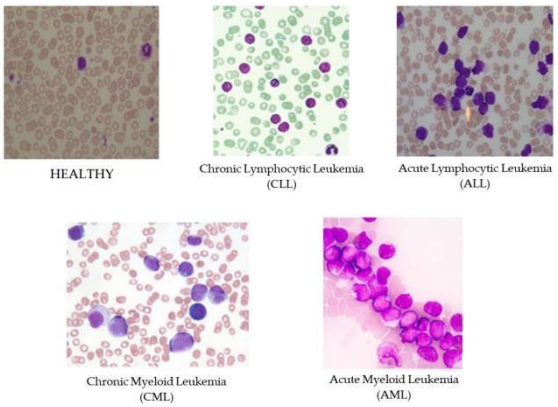
Leukemia, a type of malignant tumor that has affected millions of people. It starts in the lymphatic system where blood cells are produced. Firstly, it begins in the bone marrow and is then distributed in the blood cells of the entire body. Normally, WBCs grow based on body needs, but in case of Leukemia, they are created abnormally and



become inefficient. Although they can often be detected by their dark purple-like appearance, the analysis and further processing become very complicated due to variability in shape and texture.

The category of Leukocytes includes cells that can greatly vary from each other. While they can be distinguished by their shape and size, the problem is that WBCs are surrounded by other blood components like red blood cells and platelets.

Traditionally leukemia was detected manually by experts under microscope. This manual examination of the disease used to be time consuming, costly and totally dependent on expert's knowledge and skills. Since leukemia shows symptoms which resembles to normal minor disease it is very difficult but necessary to be identified it at early stages. Hence manual examination methods are not suitable.



Detection of leukemia using automated system will overcome the drawbacks of the traditional methods. Using a digital image processing it is possible to detect the leukemia image and classify it into its respective types. Different types of image processing

algorithms are being used for detection of leukemia cells in the images of various blood cell samples. Microscopic blood cell images are collected for different types of leukemia and are used as the input to these systems.

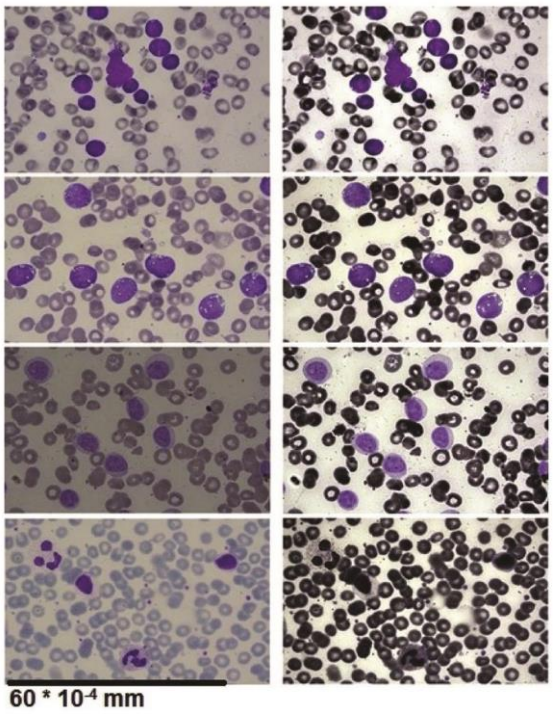


Image processing is composed of various techniques such as image pre-processing, image segmentation, feature extraction and classification. Preprocessing, segmentation and classification are very important steps in image processing. Classification accuracy depends on the correct segmented outputs. This paper proposed a automatic computerized modal for detection of leukemia using soft computing techniques. Different models are designed and created for classification purpose like . K-Nearest Neighbors, Decision Tress, logistic Regression, Random Forest, Support Vector Machine. They are trained on the features extracted from the image of WBCs..

The classifier must achieve lowest missclassification error with lowest standard deviation. Automatic classification and classifying samples accordingly could be a great boon to hematologists. For this leukemic blood with unusual evidence samples are required and have to be manual classified by hematologists for the training data set. Furthermore, the corresponding processing time must not be very high as compared with other tested classifiers. In this research, a methodology will be proposed to detect leukemia by training different models on features of lymphocytes cells. The models formed then will be tested of various samples and the modal with highest accuracy will be used for further research.

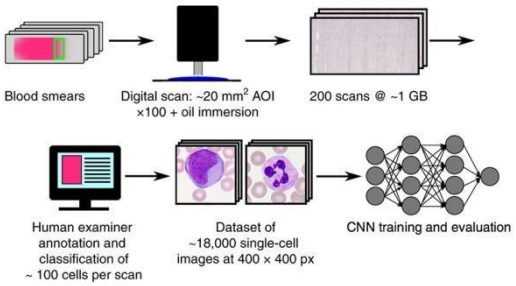
The extraction of index can be done by iterative erosion filtering of the binary nucleus. The goal is to make best modal for identification of different white blood cell categories of a given blood sample using the features extracted from the blood cell image and provide information to physicians or hematologists in the form of diagnostic reference of specific state of leukemia thereby to improve the classification accuracy with the use of soft computing techniques such as K-Nearest Neighbors, Decision Tress, logistic Regression, Random Forest, Support Vector Machine. Swarm Intelligence and other Evolutionary Computing techniques and then evaluating their performances.

III. LITERATURE REVIEW

Different Models have been proposed like

1. Convolution Neural Networks (CNN):

Deep learning using Convolution Neural Networks (CNN) is currently the best choice in medical imaging applications such as detection and classification. Even though CNNs achieve the best results on large data sets, the problem is a lot of data and computational resources are required to train them, but in mostly cases, the dataset is limited and may not be sufficient to train a CNN modal from scratch.



In such a scenario, transfer learning is used in order to leverage the power of CNNs and at the same time reduce the computational costs. In this approach, the CNN is firstly trained on a large and diverse generic image data set and then applied to a specific task. There are several pre-trained neural networks that have won international competitions like VGGNet, Resnet, Nasnet, Mobilenet,
Inception and Xception.

2. Artificial neural network (ANN):

Applying pattern recognition on blood samples for diagnosing leukemia is an extremely difficult task which frequently leads to misclassification errors due to inherent problem of data overlap. Artificial neural network (ANN) algorithm has been proposed for optimizing the classification of multidimensional data. It focuses on acute

leukemia samples. A new programming tool has been trained with ANN architecture that focuses particularly on the classification of normal vs. abnormal blood samples, namely acute lymphocytic leukemia (ALL) and acute myeloid leukemia (AML). In a proposed model 220 blood samples were considered with 60 abnormal samples and 160 normal samples. The algorithm produced highly sensitivity results with accuracy of 96.67% in ALL classification with increased data set size. With this type of accuracy, this programming tool is used to provide diagnostic references for the specific disease states to medical doctors that are considered for this study. The results obtained prove that for this type of flow-cytometry data a neural network classifier can perform remarkably well. More significant fact is that experimental ALL data considered is gradually increased from small to large data sets, the more accurate are the classification results.

3. Fuzzy logic:

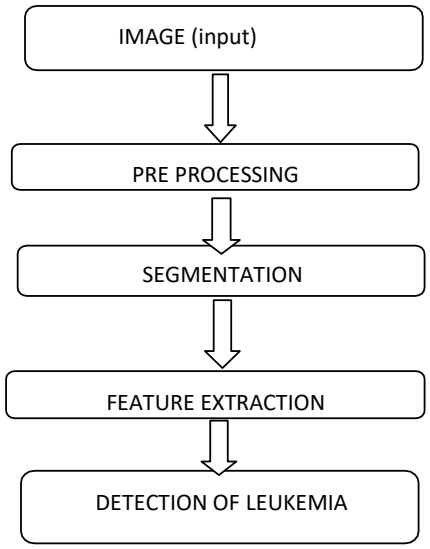
A fuzzy feature modal has been proposed for white blood cell differential counting to diagnose types of acute leukemia. The accuracy of diagnosis is found to be higher than that by numerical features after dealing with uncertainty of white blood cell features and inflexibility of diagnosing. In the proposed model on this study on acute leukemia diagnosis 120 acute leukemia images and fuzzy decision tree method with the accuracy rate of diagnosis is 84% using fuzzy features and 76.6% using numerical features has been used. Given the importance of accurate diagnosis of acute leukemia in patients, this proposal is essential and planned to be introduced in an Indonesian hospital. Recently, a new domain of research called neuro-fuzzy systems has attracted a lot of attention due to its efficacy. This may lead to the path for further improvement in accuracy of classification of leukemia.

IV. PROPOSED METHODOLOGY

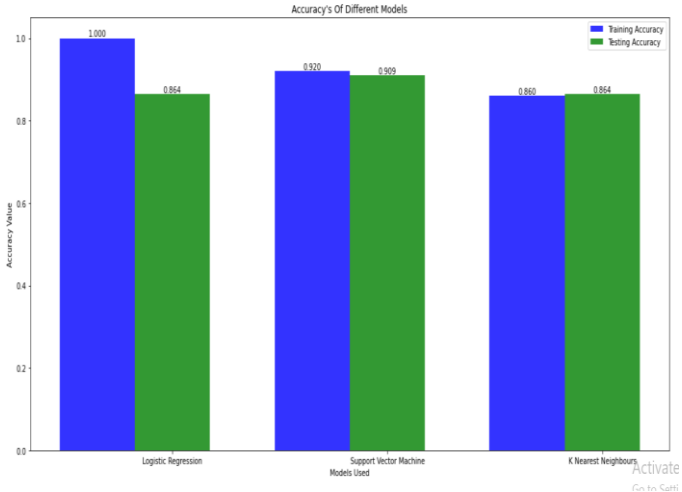
Research work begins with the segmentation of white blood cells from blood plasma using new advanced computing methods and segmentation algorithms. New and old (already used) and some other existing segmentation algorithms will be compared for their accuracies, to determine the optimal algorithm finally to be adopted for the next stage of work.

First the data goes through Feature Extraction then the extracted Features are put through different standardization techniques. To increase the accuracy Principal Component Analysis (PCA) is applied.

After the segmentation of leukocytes soft computing classifiers are used to classify using the extracted features such as area, perimeter, convex area, solidity, major axis length, orientation filled area, ratio between cell and nucleus area, mean gray level, rectangularity and circularity can be used for the classification of leukocytes. In the proposed modal 7129 features have been used of 72 samples.



The classifier used must achieve lowest missclassification error with lowest standard deviation. For this 5 different classifier have been deployed like K-Nearest Neighbors, Decision Tress, logistic Regression, Random Forest, Support Vector Machine. Then the classifier giving the best result has been choosen for further work



. V. RESULTS AND DISCUSSION

Total of 3 models have been trained on the dataset that are K-Nearest Neighbors, logistic Regression, Support Vector Machine.

Logistic Regression is providing accuracy of 86%. Support Vector Machine is providing accuracy of 90%. K Nearest Neighbor Classifier is providing accuracy of 90%. But K Nearest Neighbor algorithm has a drawback that it needs user to pre assign the k value. More work has been done to overcome this drawback. In order to overcome this drawback we have trained modal on 20 values of K and taken out the K which gives out the best accuracy. Larger dataset can be used and applied to the system in order to test the results.

This models specifications are

- Model is preprocessed with PCA n_components as 5
- In K Knearest Neighbor K has been taken as 4
- In Logistic Regression C has been taken as 5
- In Support Vector Machine degree has been taken as 15

VI. CONCLUSION

In this paper we have proposed a technique for detection and classification of Leukemia into its types. Dataset of 72 patients with 7129 features has been used. K-Nearest Neighbors, logistic Regression, Support Vector Machine are used for classification. K value has been selected as 4. It gives 90% of segmentation accuracy for the used dataset. Various statistical features are extracted for classification purpose of samples.
Different segmentation and classification algorithms are used. Logistic Regression is providing accuracy of 86%. Support Vector Machine is providing accuracy of 90%.

VII. REFRENCES

I. https://airccj.org/CSCP/vol3/csit3206.pdf
II. https://www.medicalnewstoday.com/articles/14 2595
III. https://www.nature.com/articles/s41598-020-59215-9
IV. https://acadpubl.eu/hub/2018-118-21/articles/21e/1.pdf
V. https://www.ncbi.nlm.nih.gov/pubmed/20013155
VI. https://link.springer.com/article/10.1007/s12555-012-0393-6
VII. Farag.A,"Computer based Acute leukemia classification " , Proceedings of the 46th IEEE International symposiums on Circuits and Systems, December 2003.
VIII. Fabio Scotti, "Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope",Proceedings of CIMSA 2005 – IEEE