DETECTION OF LEUKEMIA USING RANDOM FOREST CLASSIFIER

[1] Ammulya Ambasth, [2] Shreya Phule, [3] Harshala Walgude, Ms. N. A. Mulla

Abstract : One of the very well-known type of blood cell cancer is leukemia. Leukemia begins with bone marrow and leads to death if not treated early. The disease is divided into two main categories (AML and ALL) and each category consists of similar symptoms which differ slightly from each other that may confuse in diagnosing. Our project will be used for detection of leukemia. To detect leukemia many algorithms are used. In this project, detection is done using Random Forest classifier and feature extraction is done using Convolutional Neural Network (CNN).

Index Terms **- Component,formatting,style,styling,insert.**

I. INTRODUCTION

As stated by the statistics of World Health Organization(WHO), Cancer is considered as the 2nd most hazardous disease in the world which is nearly equal to one in six deaths worldwide.

WBCs are essential part of immune system of human beings which serves in fighting specially again infection and other types of diseases.

Leukemia is frequently found blood disease in children and grownups. Leukemia is kind of cancer that starts out and progresses in the blood cells, whereas huge amount of cancer cells begin in human body parts. The metabolic functions of the body are badly affected without blood as it is the crucial content in the body. In human systems, old WBCs get destroyed so that the new WBCs can replace them. In Leukemia, old WBCs do not expire and they remain inhuman blood and so the new WBCs that are generated do not get sufficient space to thrive. Hence, the working of human blood gets disturbed and the formation of WBCs becomes unusual and uncontrolled.

Leukemia which is detected in the blood and bone marrow, is induced by fast production of unusual WBCs.The large amount of unusual WBCs are unable to resist the contagion, and they hinder the capability of bone marrow in the generation of RBCs as well as platelets.

Leukemia is classified as chronic and acute. Chronic leukemia develops slower than the acute leukemia. Acute Leukemia requires instant treatment. There are two broad classifications of leukemia:

- Acute leukemia
- Chronic leukemia

These are as described below:

Acute Leukemia: It is categorized by fast development of the ailment and development of the immature WBCs. This disease is difficult to treat and cure as it becomes severe in less time. Usually in little kids. Acute Leukemia are categories into two types that are ALL and AML.

1) Acute Lymphocytic Leukemia (ALL): It occurs in children from age 1-12 years and adults of age 40 years. Here lymphocytic cell of WBCs gets affected. ALL also known as acute lymphoblastic leukemia. ALL most common in men compare to females.

2) Acute Myeloid Leukemia (AML): It occurs in children of age 1 year and older age patient. Enlargement of the spleen and bone pain these are the prime symptoms of acute myeloid leukemia. In this myeloid line of stem cells are affected.

Chronic Leukemia: For this type, the body does not indicate any of the signs at an early stage. It means at an early stage the unusual WBCs doesn't hamper functioning of the usual WBCs. It develops gradually and damages huge part of WBCs and gets the signs at the final phase which is unable to cure.

1) Chronic Lymphocytic Leukemia (CLL): The older age patients who undergoes from the old age illnesses are the main victims of this disease. It doesn't show any symptoms at the initial level.

2) Chronic Myeloid Leukemia (CML): The age group patients between 35 to 45 years are the victims of this disease. Genetic alterations occur at an initial phase of the myeloid cells.

Image processing is concerned with extraction of useful information from the digital images. It is practiced in several applications such as Medical Imaging, Forensic Studies, Textiles, Military, Film industry, Document processing, Printing Industry. Medical imaging practices the data extracted from the digital image that improves the diagnosis of various types of diseases.

For proper diagnosis of leukemia image preprocessing has to be done which includes contrast enhancement and image segmentation techniques. Finally the extracted features are compared with those stored in the database and finally classified as AML and ALL.

Keyword:

leukemia, classifier, white blood cells, Random Forest.

LITERATURE SURVEY

1) Author :- Harun et al. 2015[2]
   - Method :-
     - Pre-processing - They have used median filter for pre-processing
     - Segmentation - They applied Fuzzy C-Means, K-Means, Moving K-Means for segmentation.
     - Feature Extraction - Saturation component of HSI colour area
   - Strengths :- Modified K-Means algorithm named Moving K-Means to give enhanced results than the Fuzzy C-Means as well as K-Means

2) Author : - Goutam et al..2015[3]
   - Method :-
     - Pre-processing - They have used median filter for pre-processing.
     - Segmentation - They applied K-Mean Clustering for segmentation
     - Feature - For extraction LDP is used
     - Classifier - For classification SVM Classifier is used
   - Strengths :- High precision results using LDP Operator for the feature extraction
   - Weaknesses :- Only used blood samples containg AML and concentrated on cancer as well as non-
   - cancer cells.

3) Author :- C. Vidhya et al., 2015[6]
   - Method :-
     - Pre-processing - They have used morphological filtering for preprocessing.
     - Segmentation - They have used morphological filtering for pre-processing.
     - Feature Extraction - For extraction histogram of the LDP is used
     - Classification – Support Vector Machine
   - Strengths :- LDP feature extraction method is used

- Weaknesses – Concentrates on cancerous as well as noncancerous classification only

4) Author :-A. Mishra and N. Patel, 2015[7]
   - Method :-
     - Pre-processing - Median filtering, image cleaning, Wiener Filter
     - Segmentation - lymphocytes and myelocytes are grouped using K-means Zack Algorithm
     - Feature Extraction - Geometric :- eccentricity , convexity, perimeter, radius, area, rectangularity, symmetry, compactness, concavity, elongation , solidity Texture:- homogeneity ,entropy, correlation, energy Statistical feature :- gradient matrix ,skewness, mean, variance
     - Detection – Based on Solidity and Roundness
     - Classification - SVM
   - Strengths :- Leucocytes grouping done at the Image cleaning stage and selection of cytoplasm and the nucleus
   - Weaknesses :- Concentrates on cancerous as well as noncancerous classification only. Little Testing data

5) Author :- Y. Li et al. 2016[8]
   - Method :-
     - Pre-processing - Get one contrast-stretched gray image and one H component image from transformed in the HSV colour space. Performes contrast stretching
     - Segmentation - RBC separation (in Component image from HSV Colorspace) and Image background extraction (in gray-scale image) by using Dual threshold values obtained from golden section search method Post processing: mathematical morphological operation and median filtering
   - Strengths :- Proposed a dual threshold method for segmenting WBCs. Two threshold values are obtained using golden section search method for background image mining as well as RBC separation

Mathematical Model

W: input volume size
F: receptive field size of the Conv Layer
S: stride
P: amount of zero padding used
W,H,D: width, height and depth

The Convolutional Layer:

The Conv layer is the core building block of a Convolutional Network that does most of the computational heavy lifting.
To summarize, the Conv Layer:
Accepts a volume of size $W_1 \times H_1 \times D_1$
Requires four hyperparameters:
- Number of filters K,
- their spatial extent F,
- the stride S,
- the amount of zero padding P.

Produces a volume of size $W_2 \times H_2 \times D_2$ where:
- $W_2=(W_1-F+2P)/S+1$
- $H_2=(H_1-F+2P)/S+1$ (i.e. width and height are computed equally by symmetry)
- $D_2=K$

With parameter sharing, it introduces $F \cdot F \cdot D_1$ weights per filter, for a total of $(F \cdot F \cdot D1) \cdot K$ weights and K biases.
In the output volume, the d-th depth slice (of size $W_2 \times H_2$) is the result of performing a valid convolution of the dd-th filter over the input volume with a stride of S, and then offset by dd-th bias.

Pooling Layer:

Its function is to progressively reduce the spatial size of the representation to reduce the number of parameters and computation in the network, and hence to also control overfitting.
More generally, the pooling layer:
Accepts a volume of size $W_1 \times H_1 \times D_1$
Requires two hyper parameters:
- their spatial extent F,

- the stride S,

Produces a volume of size $W_2 \times H_2 \times D_2$ where:

- $W_2=(W_1-F)/S+1$
- $H_2=(H_1-F)/S+1$
- $D_2=D_1$

Introduces zero parameters since it computes a fixed function of the input
Note that it is not common to use zero-padding for Pooling layers
It is worth noting that there are only two commonly seen variations of the max pooling layer found in practice: A pooling layer with F=3, S=2 (also called overlapping pooling), and more commonly F=2, S=2. Pooling sizes with larger receptive fields are too destructive.

System Architecture:

Image Acquisition

WBCs images are obtained with the aid of digital microscope. There are inbuilt cameras in the digital microscope which tends to capture the WBCs digital images.

Pictures of the blood stains of leukemic patients and pictures of the blood stains of non-leukemic patients; have been gotten from online databases.

Image Pre-Processing

Noise is obtained in the microscopic images because of manual intervention and excessive stains. Here shadows of nuclei is mainly presented by noise. The region of interest is a nucleus of WBCs, so the images are processed to eliminate undesirable noises and regain important one. Some of the earlier studies have shown that image improvement techniques like contrast enhancement will develop medical image quality. Images are enhanced using this process.

The images acquired were in CMYK form, therefore pre-processing and changing to RGB form was carried out.

Image Segmentation Algorithm

The segmentation process is used for simplification and to represent the WBC image more significantly and easier to study. The Convolution Neural Network algorithm has been used in this case.

Feature Extraction

To overcome the main problem faced because of number of variables that need huge amount of space and computation during data analysis, feature extraction is necessary. The features that were extracted are shape based features and statistical features.

To choose if cell is normal or not, feature extraction of WBCs is done. Leukemia detection is carried out by considering following features:

Statistical: standard deviation, variance, mean, gradient matrix, and skewness of histogram of image matrix of cell are obtained.

Textural: Correlation, cell homogeneity, energy, contrast, factor and entropyare included.

Geometrical: perimeter, area of cell, symmetry, radius, concavity, and eccentricity are included.

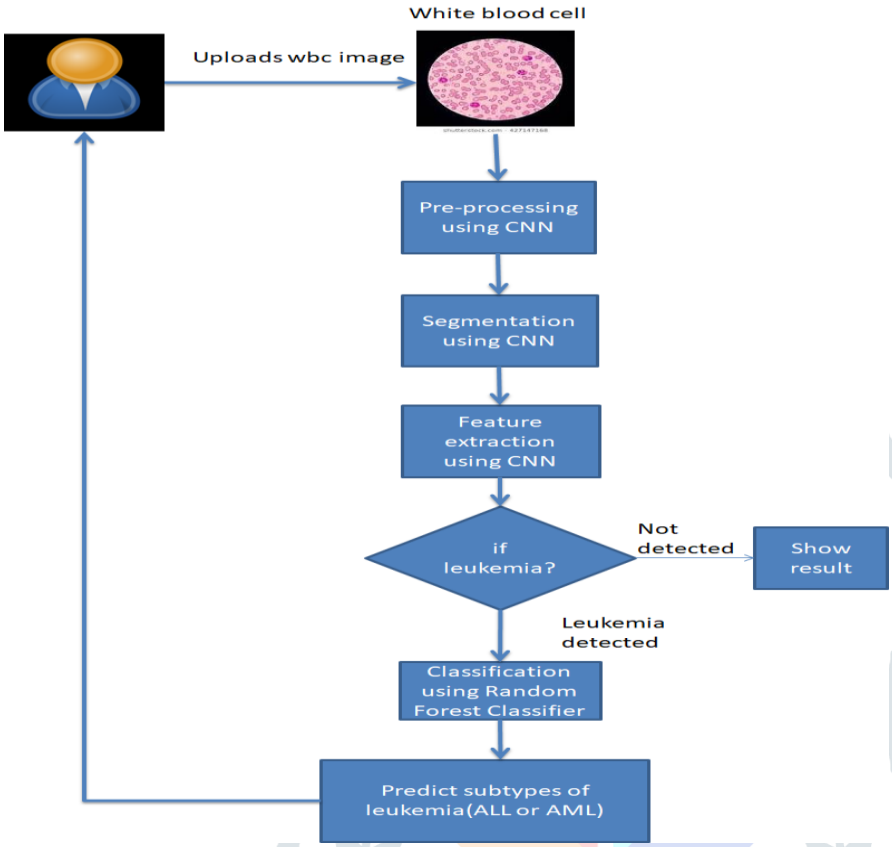
Classification

The classification model chosen for this phase is the Random Forest Classifier, which is a machine learning technique.

A forest with several decision trees is created by Random Forest Algorithm which is governed as classification algorithm. Class with is in demand or popular receives an unit vector from each tree, then combining these outcomes cause the final sort outcomes. Maximum accuracy outcomes are achieved by having maximum number of trees. A random forest algorithm is one of the strongest algorithms which is

practiced widely in various applications. Random Forest is having many qualities. The traditional machine learning algorithms usually give low classifier accuracy, and easy got over-fitting.

Random Forest acquires great classification accuracy, tolerate outliers and noise well and never gets over fitting. It operates efficiently on large data sets. It provides an experimental method for detecting variable interactions. Random Forest has been one of the most popular research methods in data mining area and information to the biological domain.



System overview

In this project, CNN classifier(Convolution Neural Network) algorithm is used to to analyzing visual imagery.

Conclusion:

In this paper, we propose a system which will detect different types of Leukemia diseases. Classification of the image is done using Random Forest Classifier. Prior detection of sub-types will help doctors to treat patients as early as possible.

ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on '**Detection Of Leukemia Using Random Forest Classifier:**'

I would like to take this opportunity to thank my internal guide Ms. N. A. Mulla for giving me all the help and guidance I needed I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

I am also grateful to our Head of Computer Engineering Department, for her indispensable support and suggestions.

Name of Students

[1] Ammulya Ambasth [2,] Shreya Phule [3] Harshala Walgude

REFERENCES

[1] M Mohamed, Hend, Rowan Omar, NermeenSaeed, Ali Essam, Nada Ayman, TaraggyMohiy, and Ashraf AbdelRaouf. "Automated detection of white blood cells cancer diseases." In *Deep and Representation Learning (IWDRL), 2018 First International Workshop on*, pp. 48-54. IEEE, 2018.

[2] N.H. Harun, Nasir, A.S. Abdul, Mashor, M.Y. Mashor, R. Hassan, "Unsupervised Segmentation Technique for Acute Leukemia Cells Using Clustering Algorithms." International Journal of Computer,Electrical, Automation, Control and Information Engineering Vol:9, No:1, 2015.

[3] D. Goutam, S. Sailaja, "Classification of Acute Myelogenous Leukemia in Blood Microscopic Images Using Supervised Classifier." International Conference on Engineering and Technology. Maret. 2015.

[4] M. Amin, S. Kermani, A. Talebi, et al. "Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using KMeans Clustering and Support Vector Machine Classifier." Journal of Medical Signals and Sensors, 5(1), pp 49-58, Jan-Mar, 2015.

[5] M. Amin, S. Nasser., S. Kermani, A. Talebi, "Enhanced Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images based on Feature Reduction using Principle Component Analysis". Frontiers in Biomedical Technologies, Volume 2, Issue 3, November 2015.

[6] C. Vidhya, P.S Kumar, K.Keerthika, C.Nagalakshmi, B.M devi, "Classification of Acute Lymphoblastic Leukemia in Blood Microscopic Images using SVM", Intl. Conf. on Engineering Trends and Science &Humanities (ICETSH), pp 232-236, 2015.

[7] N. Patel, A. Mishra, "Automated Leukaemia Detection using Microscopic Images", 2nd Intl. Symposium on Computer Vision and the Internet. Procedia Computer Science 58 pp 635 – 642, 2015.

[8] Yan Li, Rui Zhu, Lei Mi, Yihui Cao, Di Yao, "Segmentation of White Blood Cell from Acute Lymphoblastic Leukemia Images Using DualThreshold Method".Journal of Computational and Mathematical Methods in Medicine. Volume 21 April 2016.

[9] A.H. Kandil, O.A. Hassan, "Automatic Segmentation of Acute Leukemia Cells", International Journal of Computer Applications Volume 133 – No. 10, January 2016.