Acute Lymphoblastic Leukemia Detection And Classification Of Its Subtypes Using Pretrained Deep Convolutional Neural Networks

B.Ramya, Dr.Uma Rani

Abstract: Leukemia is an incurable disease of white blood cells that have an effect on blood and bone marrow in human body. Acute myeloid leukemia (AML) is one of the familiar types of leukemia amongst adults. The symptoms of leukemia are non-specific and are equivalent to the symptoms of other communal disorders. We employ modified convolutional neural network for automatic detection of acute lymphoblastic leukemia and classification of its subtypes into 4 classes, namely, L1, L2, L3 and Normal. To reduce overtraining, data augmentation technique was used. We also evaluated the data sets with various color models to verify the performance over diverse color images. A sensitivity of 100%, specificity of 98.11%, and accuracy of 99.50% is achieved for Acute Lymphoblastic leukemia detection. Compared to standard methods, the proposed technique was able to accomplish high accuracy without microscopic image segmentation based classification compared to existing classification methods Circumventing Ant Colony Optimization (CACO_CNN), particle swarm optimization (PSO_CNN) and proposed classification modified particle swarm optimization (PSO_CNN) provided high accuracy.

Keywords: Automated Leukemia detection, acute lymphoblastic leukemia, convolutional neural network, Local Binary Pattern, Machine Learning, Modified PSO.

I.INTRODUCTION

Leukemia may be a cancer of the white blood cells or bone marrow. Bone marrow is a soft tissue within the bone. Diagnosing leukemia is predicated on the actual fact that WBC count is accrued with immature blast (lymphoid or myeloid) cells and minimized neutrophils and platelets. The presence of excess variety of blast cells in peripheral blood may be an important symptom of malignant neoplastic disease. The automatic detection of malignant neoplastic disease from blood microscopic image usually avoids the issues of manual testing of blood smear and additionally increases the accuracy. It reduces the process time and thereby increasing the potency. The early identification of acute lymphocytic leukemia symptoms in patients will greatly increase the likelihood of recovery. Nowadays the sickness disease will be known by automatic specific tests like genetic science and Immuno pheno typing and morphological cell classification created by full-fledged operator's perceptive blood/marrow magnifier image. Those strategies are not enclosed into massive screening programs and are applied only if a typical symptom seems in traditional blood analysis. The Immunophenotyping and Cytogenetic diagnostic strategies are presently most well-liked for its great accuracy with relevant to the method of blood cell observation that presents undesirable drawbacks: slowness and it presents a not standardized accuracy since it depends on the operator's capabilities and weariness. Conversely, the morphological analysis simply needs a picture not a blood sample and therefore is appropriate for low value and remote diagnostic systems. In blood smear, variety of red cells is more than white blood cells. Platelets are little particles and aren't clinically important. Blood cells kind within the bone marrow, the soft material within the centre of most bones. Leukocytes or white corpuscle are cells involved in defensive the body against infective organisms and foreign substances.

- *Research Scholar, Department of computer science, Sri Ramakrishna College for Women Coimbatore.*
- *Associate professor, Department of computer science, Sri Ramakrishna College for Women.*
- *Coimbatore.*

As an example, a picture could contain up to one hundred red cells and only one to three white cells. Platelets are little particles and aren't clinically vital. Blood cells kind within the bone marrow, the soft material within the centre of most bones.

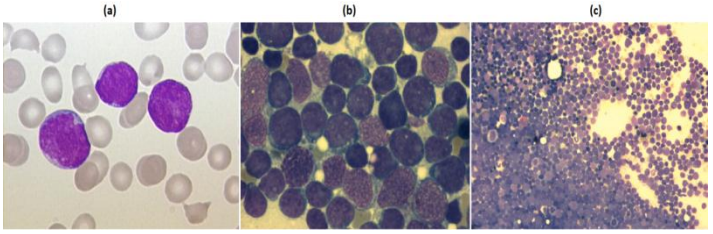


Fig (*a) acute lymphoblastic leukemia (ALL), peripheral blood of a child, (b) bone marrow smear (large magnification), (c) bone marrow smear.*

II.CELLULAR ELEMENTS SEGMENTATION

The segmentation algorithmic program aims to separate each blood corpuscle in its 2 most vital elements: nucleus and cytoplasm. As a result of the colour intensity of a picture element isn't enough to with success segment cells from pictures with colour variations, this algorithmic program uses features of the neighbouring pixels as contextual data to get homogenised regions. With the assistance of a professional, we analyzed the features of samples of bone marrow cell pictures with totally different staining so as to design an algorithmic program which will suitably segment leukocytes and their individual nucleus. Our leukemia digital pictures were obtained from bone marrow samples exploitation the Wright's stain methodology. during this analysis we discovered some colour and texture features that we accustomed distinguish between blood cells such as: 1) red blood cells get orange and rose shades whereas leukocytes reveal purple tonalities in their nucleus, and blue and rose tones in their cytoplasm (for lymphocytes and myelocytes, respectively), 2) the colour intensities non-heritable by the nucleus are darker than those of the cytoplasm, and 3) the texture of the

nucleus and cytoplasm of leukocytes, red blood cells, and also the image background is totally different among them.

III.COLOUR AND TEXTURE SEGMENTATION AND CELL IDENTIFICATION

A segmentation model that employs contextual colour and texture information in order to classify picture elements corresponding to cell elements in bone marrow pictures with heterogeneous staining. First, a channel is represented with the help of a binary Markov Random Field model that consists of a label field and three observation fields. The observation fields are channel intensity, structural texure and stochastic texure fields. Following this, based on Bayes theorem, the segmentation problem is represented as a Maximum a Posterior (MAP) estimation of the label field.

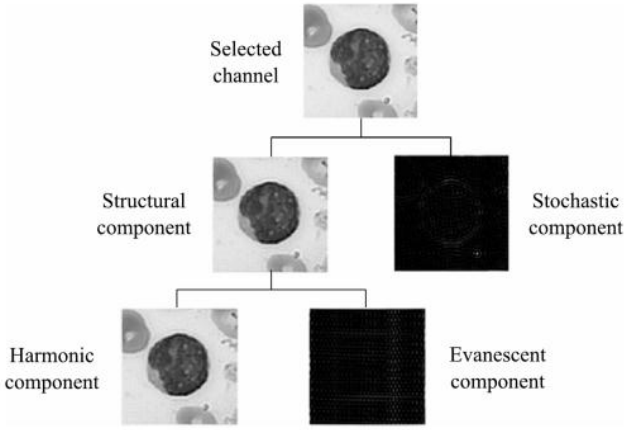


Figure 1: *Wold's decomposition texture model.*

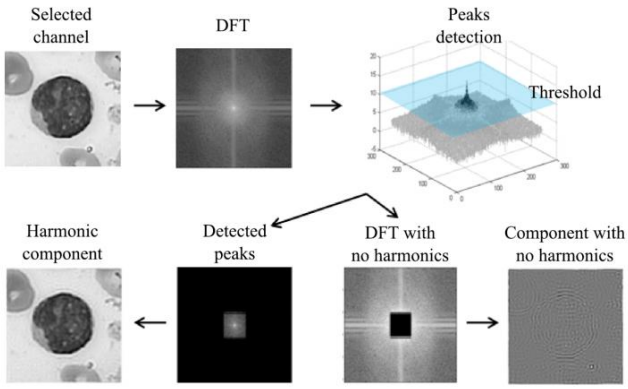


Figure 2: *Harmonic field parameterization.*

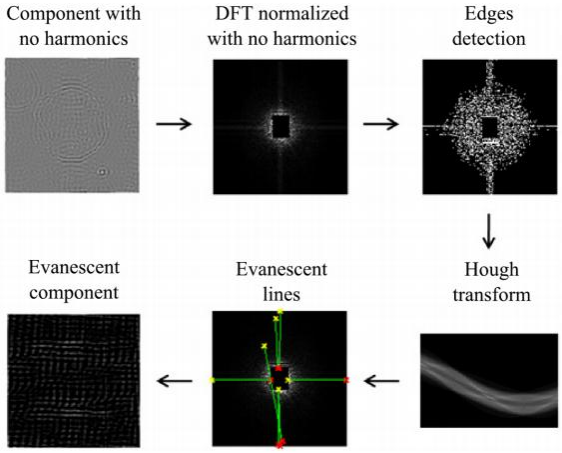


Figure 3*: Generalized evanescent field parameterization.*

IV.PROPOSED SYSTEM

In this work, a network containing 4 layers is proposed. The primary 3 layers are for the purpose of detecting features and therefore the other two layers (Fully connected and Softmax) are for classifying the features. The input image has the dimensions 50x50x3 and the receptive field or the filter size has the dimension 5x5. If the stride is 1 then the filters is moved one pixel at a time. The zero-padding is 2. It'll allow us to regulate the spatial size of the output image. The spatial size of the input volume is precisely preserved and therefore the input and output width and height will be the same. During the research, we found that, altering the dimensions of unique image during the convolution results in decrease of accuracy about 40%. Thus the output image after convolution layer 1 has an equivalent size with the input image. The convolution layer 2 has an equivalent structure with the convolution layer 1. The filter size is 5x5, the stride is 1 and therefore the zero-padding is 2. The amount of feature maps, the channel or the depth in this case is 30. If the amount of feature maps is lower or above 30, then the accuracy will be declined to 50%. By analysis, it is found that the accuracy is also declined to 50% if we remove Convolution layer 2. The Max-Pooling layer 25x25 has Filter size of 2 and stride of 2. The fully connected layer has 2 neural. The principle step in generating the blood cell features is the recognition of various blast cells with highest accuracy. In feature extraction, the input data is transformed into a set of features. The classifier performance is greatly influenced by the feature selection and therefore, the correct selection of features is a very crucial part. The features considered for calculation are Hausdorff dimension, Shape features and Texture features LBP local binary pattern features. Hausdorff dimension is the ratio of the log of a number of squares in the superimposed grid to the number of occupied squares. Among these many techniques, PSO is a very well-known iterative process that works together to find the optimal or closest solution in a multidimensional search space using the concept of searching and exploiting wisely. The learning strategy of this technique is very suitable and reliable for the discovery of structural information in an active area of GO research. This evolutionary computational method was first invented in the mid-1990s by Kennedy and Eberhart (1995) and Kennedy (1997), on the beautiful collaborative proposal of biological populations rooted in the concept of "information sharing and collective intelligence".

4.1 Modified Particle Swarm Optimization

The Original PSO may be a population based stochastic optimization technique developed by Dr. Eberhart and Dr.Kennedy in 1995 during which each and every candidate solutions are called the particles. During this proposed approach, a fitness function is employed in each step of this algorithm. It's initialized with a gaggle of random particles (solutions) then it searches for the optimal solution by updating generations. In each and every iteration every particle is updated by following two "best" values. The first one is that the fitness value, this value is named pbest. Another "best" value that's tracked by the particle swarm optimizer is that the best value, obtained by any particle

within the population. This best value may be a global best and called gbest. When a particle takes a part of the population as its topological neighbours, the simplest value may be a local best and is named lbest. Initially, the particles' velocities are set to zero & their position is randomly set within the boundaries of the search space. Then calculate the speed of the particle with their position using the subsequent equations (1) and (2). Then update the position with updated velocity. v[] = v[] + c1 * rand() * (pbest[] - present[]) + c2 * rand() * (gbest[] - present[]) -----(1) present [] = persent[] + v[] -----(2)

The following explains the algorithm of Modified PSO.

```
P = Particle_Initialization ();
for i = 1 to it_max
        for each particle p in do
                fp = f(p);
                If fp is better than f(pBest)
                pBest = p;
                end
        end
while generation< maxGenerations do
                compute speed();
                update position();
                mutation();
                evaluation(;)
                update particle memory();
                generation++;
end
gBest = best in p in P;
for each particle p in do
        V = v + cl*rand*( pBest - p) + c2*rand*(gBest - p);
        P = p + v;
end
end
```

4.2 IMAGE CLASSIFICATION USING CONVOLUTION NEURAL NETWORKS

Convolutional neural networks have been successful in image classification. The strength of CNN is based on its ability to automatically extract high-level features from a multi-level architecture. To train CNN for image classification, one must first build network architecture. The task is to determine the type, number, and order of the layers in the network. The created network seeks to find features that can be used to differentiate classes by assigning a set of 2D images along with their associated class labels. A CNN uses a learning method, which consists of two redundant and alternate passes, named Feed Forward and Backward Pass. A typical CNN's Feed Forward Pass does two main things. The first task is feature extraction through multi-convolutional feature extraction (CFE) layers. For this purpose, an image is sent through multiple CFE layers in a serial manner. A CFE layer consists of three sublayers: a convolutional sublayer, then a nonlinear transformation sublayer, and then a pooling sublayer. Each CFE layer takes properties from the previous layer and builds top-level properties. This process is repeated several times to eventually extract high-level features from the image. These features are input to the layers that are fully integrated in the second task of the feed forward pass, which makes the classification of the input image and receives some error. In the backward pass, the error received from the feed forward pass propagates backwards to adjust the weights in the convolutional sublayers, and, therefore, they can better characterize the classification problem. The same error is also used to find the optimal weights for fully connected layers.

V.RESULT

Images utilized in this investigation were gotten from ALL-Image Database (IDB) informational collection which is an open informational collection accessible on the web. This informational collection was partitioned into 2 renditions. Intense lymphoblastic leukemia-IDB 1 comprised of 108 pictures where 59 pictures were from solid patients and 49 pictures were from patients influenced with leukemia.

Processing Time analysis

To implement our proposed algorithm, we used MATLAB software (R2016a) on a laptop, Intel Core i3 (2.0 GHZ) and 4GB of memory. The resolution of images in our database was 512 x 512. To efficiently evaluate our proposed algorithm, we analyzed each step of our algorithm based on processing time. Table 1 represents the average processing time (PT) of each module in the proposed algorithm.

Module Time	Processing time	
	PSO (Seconds)	Modified PSO (Seconds)
Reprocessing	1.458	1.132
Segmentation	8.548	7.569
Feature Extraction	0.964	0.578

***Table 1**: Processing time*

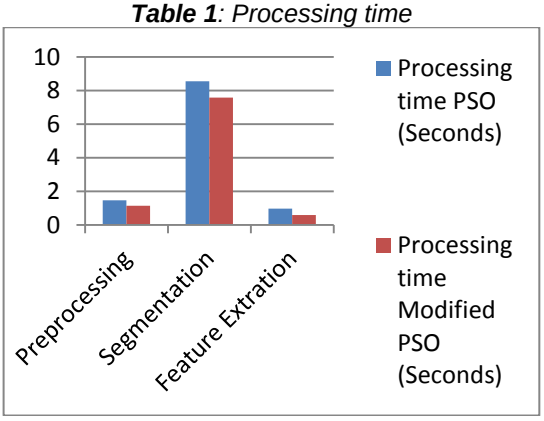


***Figure 4**: Comparison of Processing Time*

Performance Analysis

Accuracy, Specificity and sensitivity are calculated to analyse the performance of the system. These parameters are calculated using true positive, true negative, false positive and false negative values. The parameters is calculated by

Accuracy = (T P + T N) / (T P + T N + F P + F N) ---------- (1)

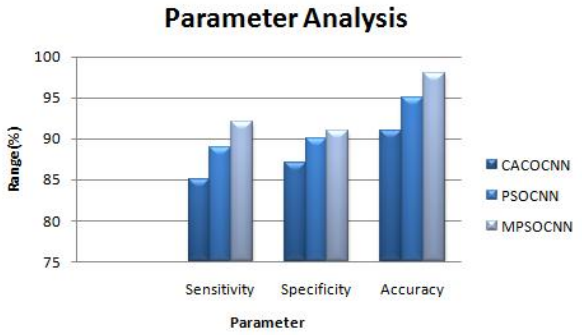
Specificity = TN / (T N + F P) ---------- (2)

Sensitivity = TP / (T P + F N) ---------- (3)

Where TP is the number of real positives detected by the system, FP is the number of real negatives labelled positive, TN is the number of real positives, and FN is the number of real positives missing and the system is labelled negative. Table 2 represents the classification performance of normal and abnormal for 100 specimens.

Parameter	CACOC NN	PSO CNN	MPSOC NN
Sensitivity	85	89	92
Specificity	87	90	91
Accuracy	91	95	98

Table show parameter analysis of existing CACOCNN, PSOCNN and proposed method of MPSOCNN



This figure 5 show classification parameter of sensitivity, specificity, and accuracy compared with CACOCNN, PSO_CNN existing method and MPSO_CNN classification. Proposed method provided efficient sensitivity, specificity, accuracy.

VI.CONCLUSION

The fundamental point of this paper is core division pursued by highlight extraction to distinguish Leukemia. Shape Features of cores, for example, area, border, and so forth are considered for better accuracy of detection. Proposed Modified Particle Swarm Optimization characterization strategy utilizing CNN as one of the developing modules for the acute leukemia type clinical choice emotionally supportive network. In future study, we will consider about how to improve our engineering to acquire a superior outcome and attempt to classify four kinds of Leukemia.

VII.REFERENCES

[1] K.Srikar, M.Akhil, V.Krishna reddy, “Execution of Cloud Scheduling Algorithms”,
[2] International Innovative Research Journal of Engineering and Technology, Vol 02, No 04, pp.108-111, 2017.
2. Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, “Automatic Tuberculosis Screening Using Chest Radiographs” IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL.33, NO.2, FEBRUARY2014.
3. Laurens Hogeweg*, Clara I. Sánchez, Pragnya Maduskar, Rick Philipsen, Alistair Story, “ Automatic Detection of Tuberculosis in Chest Radiographs Using a Combination of Textural, Focal and Shape Abnormality Analysis” IEEE TRANSACTIONS ON MEDICAL IMAGING,VOL.34,NO.12, DECEMBER 2015.
4. Elmar Rendon-Gonzalez and Volodymyr Ponomaryov “Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM” International conferences on 2016 IEEE.
5. Al-Ajlan, A.; El-Zaart, A.: Image segmentation using minimum cross-entropy thresholding. In: IEEE International Conference on Systems, Man and Cybernetics, 2009. SMC 2009, pp. 1776–1781.IEEE (2009)
6. Tao, W.; Jin, H.; Liu, L.: Object segmentation using ant colony optimization algorithm and fuzzy entropy. Pattern Recogn. Lett.28 (7), 788–796 (2007)
7. Horng, M.H.: Multilevel thresholding selection based on the artificial bee colony algorithm for image segmentation. Expert Syst.Appl. 38(11), 13785–13791 (2011)
8. Naidu, M.; Kumar, P.R.; Chiranjeevi, K.: Shannon and fuzzy entropy based evolutionary image thresholding for image segmentation. Alex. Eng. J. (2017)
9. Pan C, Lu H, Cao F. Segmentation of blood and bone marrow cell images via learning by sampling, Emerging Intelligent Computing Technology and Applications, Springer Berlin Heidelberg. 2009; 336–345.
[3] 10. Pan C, Park DS, Yoon S, Yang JC. Leukocyte image segmentation using simulated visual attention. Expert Systems with Applications. 2012; 39(8): 7479–7494.
[4] 11. Wang W, Hao S. Cell cluster image segmentation on form analysis. IEEE Third International Conference on Natural Computation. 2007; 4: 833–836.
[5] 12. Hengena H, Spoor SL, Pandi MC. Analysis of blood and bone marrow smears using digital image processing techniques. Medical Imaging, International Society for Optics and Photonics. 2002; 624–635.
[6] 13. Mohapatra S, Patra D. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Computing and Applications. 2014; 24(7–8): 1887–1904.
[7] 14. Lim HN, Mashor MY, Hassan R. White blood cell segmentation for acute leukemia bone marrow images. IEEE International Conference on Biomedical Engineering (ICoBE) 2012; 357–361.