International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
OPEN ACCESS
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Machine Learning Based Blood Cell Categorization in Semar Images

Aniruddha Deshmukhe, Samarth Dagade, Nikhil Deshmuksh, Nilesh Jadhav, Prof. Ashish Manwatkar, Nitin Dhawas
Nutan Maharashtra Institute of Engineering and Technology, Talegaon(D), Pune, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 10 Nov 2023
Published: 30 Nov 2023

Publication Issue
Volume 10, Issue 6
November-December-2023
Page Number
255-262

ABSTRACT

The distribution of blood cells in peripheral blood smear (PBS) is important in the diagnosis of blood diseases such as leukemia, anemia, disease, cancer and polycythemia. In blood analysis, hematologists always use a microscope to determine the total number, morphology, and distribution of cells. Hematology analyzers, flow cytometers provide a complete and accurate blood count (CBC) to reveal abnormalities in blood smears. The equipment is expensive, time- consuming [1], requires manual intervention, and is not available in many hospitals. Therefore, a cheap and effective method that can identify different bacteria from a single Periferal Blood Semer Image(PBS) is needed. Automatic classification of samples improves the hematological process, accelerates the diagnostic process and increases the accuracy of the measurement process. The proposed system uses a semi-automatic method to compartmentalize and divide blood cells into white blood cells (WBC) and red blood cells (RBC). The texture of the cells is extracted using Grayscale Co- occurrence Matrix (GLCM) and fed into Logistic Classifiers like Naive Bayes classifier, K-nearest neighbor, decision tree, K-means cluster, random forest, regression. , ANN and SVM.

Keyword: Deep Learning, Machine Learning, Cell Counting, RBC, WBC.

I. INTRODUCTION

Blood is a continuous fluid containing red blood cells, white blood cells, and platelets that circulate throughout the hematopoietic system in the bone marrow.

As shown in Fig.1(a), White blood cells are the largest cells and are very rare. Depending on the five types of white blood cells, they have a nucleus surrounded by cytoplasm and an average diameter of 8-20μm. Normal red blood cells are biconcave disc- like cells with a pale centre. They have no nucleus and are red. It is approximately 7-8μm in diameter and more. A difference in the size, volume, or shape of red blood cells represents a red blood cell abnormality. Platelets are cytoplasmic fragments of megakaryocytes, usually

2–3 μm in diameter, blue in color, and the latter are quite abundant [1]. Blood smears are often used to follow up abnormal CBC results. Smear a small amount of blood on a slide, stain it with Leishman, Giemsa, and Wright-Giemsa to see the cells better, and then look at it under the microscope. PBS can identify the appearance of blood and abnormalities that may be visible on the slide. For example, the appearance of red blood cells that are smaller than normal and lighter in color indicates some types of diabetes. Similarly, the presence of immature white blood cells can add information to other tests to help diagnose cancer, leukemia, or bone disease. PBS depicts different blood cells, shown in fig.1(a).

Fig.1 (a) Formation of blood cells

(b) Blood smear image Fig. 1 Blood cells

Blood cell classification procedures

It takes a long time and is not reliable. Therefore, the purpose of this study is to help doctors to diagnose blood diseases. Therefore, different learning (ML) techniques are used to classify blood cells in multiple blood smear images into white blood cells and red blood cells. The proposed work will be executed using Python by executing machine Learning algorithms.

II. RELATED WORK

The next section gives a brief overview of previous work in the field of automatic blood segmentation and classification. Alam et al. [4] reported an objective detection method to identify and count blood cells using the YOLO algorithm and achieved 96.1% accuracy for red blood cells and 86.89% accuracy for white blood cells from 364 samples.

Alomari et al. [5] reported a reconstruction method with 98.4% accuracy for white blood cells and 95.3% accuracy for red blood cells over 100 images. Cruz and colleagues proposed a method to count red blood cells using spot analysis and terrain shifting of the hue saturation value (HSV). [6] The average accuracy of 10 blood tests is 96%. Wei et al. [7] proposed a method to distinguish WBC and RBC using H and S features. The authors of this study achieved accuracy of 92.9% of 100 Wright-Giemsa dyed images. Lodo et al. [8]

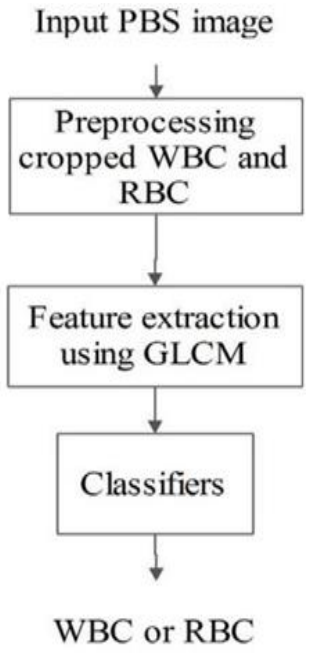
He extended the haemocytometer method by seeding each cell according to its neighbour and the SVM method. This method uses 368 ALL-IDB dataset images and achieves 98% accuracy for red blood cells and 99.2% accuracy for white blood cells. Tran et al. [9] proposed a deep segmentation study, and the accuracy of red blood cell and white blood cell segmentation was 94.93% and 91.11%, respectively. Tobias et al. [10] proposed a faster R- CNN model to classify WBC and RBC variants and achieved 98% accuracy on the BCCD dataset. Although deep neural networks and many machine learning algorithms are used for classification of blood cells. The proposed system uses a simple method that can distinguish white blood cells and red blood cells in many images.

III. Literature Survey

Sr No.	Paper Name	Author	Description
1	Classification of blood cells into white blood cells and red Hood cells from blood smear images using machine learning techniques (2021)	Navya K.T, Keerthana Pusad, Ben Mohan Kumar Singh	Various crucial steps are taken to classify blood cells from images to diagnose blood related disorders such as anemia, an infection, cancer etc.
2	Feature Extraction of Whiter Blood Cells Ung CMYK- Moment Localization and Deep Learning in Acuate Myeloid Leukemia Blood Smear Microscopic Images (2021)	TUSNEEM AHMED M. MOHD ALJURF 4	Artificial intelligence has revoltionalized medical diagnosis, particularly for cancers Acute myeloid leukemia (AML) diagnosis is a tedious protocol that is prone to human and machine errors. In several instances, it is difficult to make an final decision even after carefull examination by an experienced pathologist
3	Lightweight Efficient Net33 Motel Based on Depth wise Separable Convolutions for Enhancing Classification of	AMREEN HATOOL LAND	Acute lymphoblastic leukemia (ALL) is a type of leukemia cancer that arises due to the excessive growth of matter white blood cells (WBCs) in the bone marrow. The ALL rite for children and adults is nearly 80% and 40%, respectively. It affects the production of immature cells, leading to an abnormality of neurological cells and potential fatality
4	Automatic Detection of White Blood Cancer From Bone Mac Microscopic Images Using Convolutional Neural Networks 2020	DEEPIKA KUMARI, NIKITA	Leukocytes, produced in the bone marrow, make up around one percent of all blood cells. Uncontrolled growth of these white blood cells leads to the birth of blood cancer. Out of the three different types of cancers, the proposed study provides a robust mechanism for the classification of Acute Lymphoblastic Leukemia (ALL) and Multiple Myelona (MM) using the SN-AM dataset
5	Blood Cell Classification using Neural Network Models 120231	Jagrit Mitra, Kartik Vijayran, Kartikeya Verma, Anurag Goel	Blood cells classification is a crucial aspect in medical diagnosis. Several machine learning models have been proposed under various researches for classification of blood cells in recent years
6	An Automated Blood Cells Counting& Classification Framework using Mask R-CNN Deep Learning Model (2020)	Najmeddine Dhieb 2, Hakim Ghazrai2, Hichem Beshes), and Yehia Massoud2	In this paper, we proposean automated blood cells counting framework using convolutional neural artwork (CNN), instance segmentation, transfer learning, and mask R-CNN techniques Red and white blood cells are identified, classified, and counted from microscopic blood smear images.

7	Image Classification of Abnormal Red Blood Cells Using Decision Tree Algorithm (2020)	Vigria Mari E. Batitis &,Merwin han G. Caballes, Abigail A Ciudad, Miccaela D. Diaz, Russel D. Rores, Eng: Roselito E Tolentino	The study aims to increase the number of abnormal rod blood cells that can be detected using image processing. The study used Decision-Tree Algorithm in a machine learning algorithm in classifying.
8	Detection of RBC, WBC Platelets Count in Blood Sample by using Deep Leaming (2023)	Allaparthi HemaSri, Mopuru Devi, Sreenidhi, Valluru Verkata Krishna, Gumidi Vasanti, VMural Moltan Satish	The CNN is med for the image classification. One of the best techniques for achieving the best accuracy in the least amount of time for the blood cells dataset is VGG-16 technique. The proposed system is a combination of the CNN and VGG-16 methods
9	Classification of white blood	.Setu Sai Sowmya Kumarı, K.ChandmSekhara Chuen, M.Loclavathi, K.Pavan Kumar, Mr.N.Rajeswan	White Blood Cells also known as leukocytes plays an important role in the human body by increasing the immunity by fighting against infectious diseases. The classification of White Blood Calls, plays an important role in detection of a disease in an individual.

IV. METHODS AND MATERIAL



Proposed System Architecture

This section describes the model shown in Figure 2.

Fig. 2. Block diagram of the conceptual model

A. Image Acquisition

Many PBS images are acquired online from Kasturba Medical College (KMC), Department of Hematology, Manipal etc. [11th].

B. Pre-processing

Improving image quality before importing images to improve contrast and illumination, white blood cells and red blood cells in PBS images are manually cropped and preprocessed using local adaptive histograms. A Weiner filter [3,3] with a window filter is used to remove noise and artifacts. Gray world color normalization technique is used to adjust the color difference by calculating the average value of each channel of the image. [12].

C. Feature Extraction

Feature extraction is a method of collecting unique objects from patterns that helps distinguish groups from input patterns. Extracting texture features from manually segmented images using GLCM is a statistical method that gives the variation of the corresponding pixel. Using the Graycomatrix and gray coprops functions in MATLAB, texture properties such as contrast, strength, correlation and symmetry shown in the equation can be extracted from the region of interest in the image[13]. Texture equation:

(1) Pij = element i, j of the normalized GLCM.

N = Number of grayscales in the image. μ = mean of GLCM.

σ2 = standard deviation of all reference pixels..

Features extracted from GLCM are transferred to ML algorithms such as Naive Bayes Classifier, K- Means

Decision Tree>(DT), Logistic Regression Random Forest, K. nearest neighbour (KNN), Artificial Neural Network. and support vector machine. Machine learning methods for blood cell classification will be briefly described in the following sections [14][19].

• Naive Bayes Classifier

Naive Bayes Classifier method is based on Bayes theorem. Provide training information for the purpose of the operation and provide new information determined by the attribute value a1,a2.. The Bayesian method classifies new data by assigning the best value to the target, given the attribute value representing the event given in the equation.

• Decision trees

Decision tree learning uses yes-pay-else rules to provide classification by ordering examples from the root to some to the leaf of the tree Each node represents a property and each branch corresponds to one of the possible values. Then repeat the process for the subtree rooted at the new node data value measures the distribution of behavior relative to the presented target, for example by calculating entropy.

• Random Forest

Random Forest is an ensemble learning method used to create better predictive models of the distribution of profits and returns. Create decision trees based on criteria selected from the training program. The final ranking of test items is determined by votes selected from different decision trees.

• K Nearest Neighbours

KNN is an example-based supervised learning method where nearest neighbors are grouped and identified based on Euclidean distance

• Logistic Regression

Logistic regression is used in binary classification problems. It uses the logistic regression equation to measure the relationship between one or more variables and the response variables in the predicted outcome.

• Artificial Neural Network (ANN)

ANN is an efficient computer network that adapts its response to external input. This is a random pattern caused by blood vessels. Artificial neural networks consist of many nodes, called neurons, connected by connections that transmit signals.

Network training and connections through which training can be done are severely affected. The activation function is then applied to the input of the neuron to obtain the output. The activation function can be a sigmoid, ReLU, linear or Gaussian function.

• Convolutional Neural Network (CNN) :

Convolutional Neural Network is a type of deep learning model which designed for process and analyse visual data such as images. They are used in image recognition and classification tasks. CNN use convolution layers to automatically and adaptively learn hierarchical features from input images.

• Support Vector Machines (SVM)

SVM performs classification by finding the best hyperplane (highest edge hyperplane) in the height space that separates different events in the same category. SVM uses kernel functions such as Gaussian radial valued functions, polynomials, and sigmoid kernels to find ways to classify objects based on a set of labels. However, the selection of kernel functions must be done by trial and error.

V. RESULT AND ANALYSIS

This section compares the results obtained when dividing WBC and RBC. Around 200 white blood cells and red blood cells were distributed from PBS images obtained from KMC, Manipal and internet. Figure 3 shows that white blood cells and red blood cells are separated for further processing. Preprocessing of fragmented images improves the image. 3. Cut blood cells (a) white blood cells (b) red blood cells Good picture. Blood cells were

obtained from low-contrast, illuminated, and multistained PBS images. As shown in Figure 4, the robustness is normalized using the local adaptive

histogram and grayscale global color normalization method. The resulting image is then converted to grayscale. Feature extraction. Blood cells are separated into WBC and RBC using GLCM method to extract the characteristics and feed them into different products. Fig. 4. First blood cells from the dataset, select 75% of the data for training and 25% for testing samples. Based on the recorded data, doctors make various classifications of blood results and give their accurate meaning. Compare the performance from different products; such as positive interest (TPR), negative interest (FPR), negative interest, negative interest, high rate and high rate. The results obtained from the classification are calculated by the following formula:

Total = TP + TN + FP + FN

Accuracy = (TP + TN) / Total

Misclassification rate = (FP + FN) / Total br >Total positive rate (sensitivity) = TP / (TP + FN)

False positive rate = FP/(TN+FP)

Negative value (true) = TN / (TN + FP), FPR 0%, Accuracy 100% and Prevalence
45%.

Generally speaking, for a good classification model, the range and FPR should be less than 100%. The K-means algorithm produces data points where K (number of groups) = 2 as shown in Figure 5 There are too many cells in the wrong group. Multilayer Perceptron (MLP) is a simple neural network that uses graphs. K-means the distribution of blood cells. Distribution of blood using a regression method to identify blood cells. There is an input layer, a hidden layer with "relu" and an output layer with a "sigmoid" activation function. Loss calculation is done using binary cross entropy; for example gradient descent method "man" as optimization method and accuracy as index. The ANN output at different times is shown in the snapshot [Figure 6] with 100% accuracy and small error.6. ANN output for blood cell classification. As can be seen from the output, logistic regression classifier provides more than 97% of random results and SVM and KNN with radial root kernel function gives 95.7%. For the above random results, DT gives more than 95.7% of random results and more than 93% of random results. Forest provided 93% of the PBS images, and the distribution of white blood cells and red blood cells was 94%.

VI. CONCLUSION

An automatic classification model is proposed to classify blood cells in PBS images into white blood cells and red blood cells using various machine learning algorithms. Grayscale color normalization is used to adjust the color and light changes of objects according to use. The training classification
model shows that the logistic regression classifier
best fits the data with 97% accuracy and 100% accuracy with 94% TPR

In this study, we can continue to use object detection and classification CNN models to avoid big data and the first step of blood cell subtype classification.

VII. REFERENCES

[1]. H. Mohan, Textbook of pathology. Jaypee Brothers, Medical Publishers Pvt. Limited, 2018.

[2]. M. Jagannathan-Bogdan and L. I. Zon, “Hematopoiesis,” Development, vol. 140, no. 12, pp. 2463–2467, 2013.

[3]. K. W. Jones, “Evaluation of cell morphology and introduction to platelet and white blood cell morphology,” Clinical Hematology and Fundamentals of Hemostasis, pp. 93–116, 2009.

[4]. M. M. Alam and M. T. Islam, “Machine learning approach of automatic identification and counting of blood cells,” Healthcare technology letters, vol. 6, no. 4, pp. 103–108, 2019.

[5]. Y. M. Alomari, S. Abdullah, S. N. Huda, R. Zaharatul Azma, and K. Omar, “Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection

algorithm," Computational and mathematical methods in medicine, vol. 2014, 2014.

[6]. D. Cruz, C. Jennifer, L. C. Castor, C. M. T. Mendoza, B. A. Jay, L. S. C. Jane, P. T. B. Brian et al., "Determination of blood components (WBCs, RBCs, and platelets) count in microscopic images using image processing and analysis," in 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE, 2017, pp. 1–7.

[7]. X. Wei, Y. Cao, G. Fu, and Y. Wang, "A counting method for complex overlapping erythrocytes-based microscopic imaging," Journal of Innovative Optical Health Sciences, vol. 8, no. 06, p. 1550033, 2015.

[8]. A. Loddo, L. Putzu, C. Di Ruberto, and G. Fenu, "A computer-aided system for differential count from peripheral blood cell images," in 2016 12th International Conference on Signal-Image Technology & InternetBased Systems (SITIS). IEEE, 2016, pp. 112–118.

[9]. T. Tran, O.-H. Kwon, K.-R. Kwon, S.-H. Lee, and K.-W. Kang, "Blood cell images segmentation using deep learning semantic segmentation," in 2018 IEEE International Conference on Electronics and Communication Engineering (ICECE). IEEE, 2018, pp. 13–16.

[10]. R. R. Tobias, L. C. De Jesus, M. E. Mital, S. Lauguico, M. Guillermo, R. R. Vicerra, A. Bandala, and E. Dadios, "Faster r-cnn model with momentum optimizer for rbc and wbc variants classification," in 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech). IEEE, 2020, pp. 235– 239.

[11]. M. Kashefpur, R. Kafieh, S. Jorjandi, H. Golmohammadi, Z. Khodabande, M. Abbasi, N. Teifuri, A. A. Fakharzadeh, M. Kashefpoor, and H. Rabbani, "Isfahan misp dataset," Journal of medical signals and sensors, vol. 7, no. 1, p. 43, 2017.

[12]. "Image enhancement techniques." [Online]. Available: https://in.mathworks.com (accessed May, 2020)

[13]. "Texture using GLCM." [Online]. Available: https://in.mathworks.com/help/images/ref/gray com atrix.html (accessed June 21, 2021)

[14]. T. M. Mitchell, "Artificial neural networks," Machine learning, vol. 45, pp. 81–127, 1997.

[15]. E. Alpaydin, Introduction to machine learning. MIT press, 2009.

[16]. J. Bell, Machine learning: hands-on for developers and technical professionals. John Wiley & Sons, 2014.

[17]. D. Kriesel, "A brief introduction on neural networks," 2007.

[18]. "Support Vector Machine." [Online]. Available: https://www.sciencedirect.com/topics/computer - science/support-vectormachines (accessed March 7, 2019)

[19]. C. Cortes and V. Vapnik, "Support vector machine," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[20]. Kokane, Chandrakant D., and Sachin D. Babar. "Supervised word sense disambiguation with recurrent neural network model." Int. J. Eng. Adv. Technol.(IJEAT) 9.2 (2019).

[21]. Kokane, Chandrakant D., Sachin D. Babar, and Parikshit N. Mahalle. "Word Sense Disambiguation for Large Documents Using Neural Network Model." 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2021.

[22]. Kokane, Chandrakant, et al. "Word Sense Disambiguation: A Supervised Semantic Similarity based Complex Network Approach." International Journal of Intelligent Systems and Applications in Engineering 10.1s (2022): 90-94.

[23]. Kokane, Chandrakant D., et al. "Machine Learning Approach for Intelligent Transport System in IOV-Based Vehicular Network Traffic for Smart Cities." International Journal

of Intelligent Systems and Applications in Engineering 11.11s (2023): 06-16.

[24]. Kokane, Chandrakant D., et al. "Word Sense Disambiguation: Adaptive Word Embedding with Adaptive-Lexical Resource." International Conference on Data Analytics and Insights. Singapore: Springer Nature Singapore, 2023.

Cite this article as :

Aniruddha Deshmukhe, Samarth Dagade, Nikhil Deshmuksh, Nilesh Jadhav, Prof. Ashish Manwatkar, Nitin Dhawas, "", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 6, pp. 255-262, November-December 2023.
Journal URL : https://ijsrst.com/IJSRST52310587