World Journal of Biology Pharmacy and Health Sciences
eISSN: 2582-5542
Cross Ref DOI: 10.30574/wjbphs
Journal homepage: https://wjbphs.com/

(REVIEW ARTICLE)

Segmentation White Blood Cells by Machine Learning Algorithms

Porimal Mollik *

Designation of Computer Science & Engineering, University of Calcutta, 87/1, College Street, Kolkata-700 073, India.

World Journal of Biology Pharmacy and Health Sciences, 2023, 13(02), 009–014

Publication history: Received on 17 December 2022; revised on 31 January 2023; accepted on 02 February 2023

Article DOI: https://doi.org/10.30574/wjbphs.2023.13.2.0159

Abstract

Blood and its elements have a vital position in human life and are the best indicator for deciding many pathological states. Specifically, white blood cells are of great significance for diagnosing hematological disorders. In this analysis, 350 microscopic blood smudge images have experimented with six machine learning algorithms for the sort of white blood cells, and their renditions have resembled. Thirty-five distinct geometric and statistical (consistency) features have been pulled from blood pictures for practicum and test parameters of machine learning algorithms. According to the outcomes, the Multinomial Logistic Regression (MLR) algorithm accomplished better than the other techniques, with an average of 95% test victory. The MLR can be utilized for the automatic classification of white blood cells. It can be used mainly as a source for diagnosing diseases for hematologists and internal medicine experts.

Keywords: White Blood Cell; Blood Cell; Machine Learning Algorithm

1. Introduction

Blood is a form consisting of plasma and blood cells in the heart's and veins' circulatory procedure, which we call the cardiovascular procedure in the body. Microscopic examination of peripheral blood smudge outcomes in hematology is expensive and time-consuming [1], [2], [3]. White blood cells (leukocyte, WBC) are usually misidentified because they are not naturally durable [4], [5]. For this basis, blood tests with classic techniques will likely encounter these issues. At the exact time, due to the statistical tendency and inconsistencies [6] of the judge (hematologist), both the subjective evaluation of outcomes and the sluggish progress of the operation. For these grounds, the growth and use of computer-based methods instead of conventional strategies will significantly contribute to the acceleration of the research process and more precise outcomes [5]. White Blood Cells are segmented into five distinct classes: monocytes, neutrophils, basophils, eosinophils, and lymphocytes [7]. The automation of digital method, specially in healthcare, plays a important role in transforming patient care services [26].

Experimenters are increasingly curious to develop automated medical image analysis algorithms, such as microscopic blood smears. Experimenters function on picture processing, computer concepts, synthetic neural networks, machine learning algorithms, etc., and methods for blood cell research. Some of the analyses in the publications are as shadows, in their research, desired to enhance the work of Turk and Pentland in choosing eigenvectors from monochrome photos. Thus, they used three elements in the color image rather than a monochrome print. The Bayesian classifier organized the Eigen partitions, not the material or geometric properties of the picture. They utilized thickness and color details as parameters in the decision-making procedure. First, they rescanned the intake pictures, segmented and circled them, and finally recognized three vectors describing the intensity and color details. The Fuzzy C-means clustering process automatically divides the nucleus and cytoplasm of leukocytes. Then suitable effects are removed from the heart, the cytoplasm, and the cell. These effects are organized by SVM. [24] in their analysis, offered a method for the type verification of white blood cells in flow cytometry. They processed the manners of parametric datasets in a multidimensional limit utilizing Support Vector Machines (SVM). [8] employed an artificial neural web of two steps to

* Corresponding author: Porimal Mollik

verify white blood cells. The first step involved a pre-classification operation of a backpropagation algorithm (BPNN). In the second stage, they showed a hybrid sample using the support vector machine (SVM) & Puls-Coupled neural network (PCNN) to decrease the detected errors. Thus, they targeted minimizing the unfavorable aspects. [9] suggested Otsu's automated thresholding algorithm for the segmentation of blood cells and the picture enhancement and arithmetic system for leukocyte segmentation. K-NN verifier was applied to verify blast cells from normal lymphocyte cells. They contained a 93% perfection rate according to the test outcomes. [10], in their research, utilized image processing methods such as color transformation, picture fragmentation, edge detection feature extraction, and white blood cell verification. They verified the dengue virus infections of patients with the decision tree techniques. According to the outcomes obtained, they said that 167 cell pictures were successful in leukocyte verification with 92.2% and 264 blood cell pictures with 72.3% perfection in dengue verification. [11] utilized Random Forests to verify leukocyte cells like mono-nuclear and polymorph-nuclear cells apart from blood smear pictures obtained with 40X exaggeration.

In ranking white blood cells, the images' color, texture, and geometric properties were utilized as input parameters of artificial intelligence-based algorithms. Some of the research in the literature are as shown; [12] for lymphocytes, monocytes, & neutrophil cells only; histogram equalization, corner extraction, and threshold-based automated segmentation. The geometric properties of the pictures were applied for the verification process, and 100 blood smear pictures were used in the tests. [13] decided to verify and count leukocytes according to 5 categories by utilizing the shape, density, and texture properties of microscopic blood pictures. The wavelet properties obtained by the Dual-Tree Complex Wavelet Transform (DT-CWT) system for the verification process were applied as parameters of the SVM classifier. [14] offered a simple sorting method using color details and morphological features. As the first stage in a two-step classification method, they have broadly modified leukocyte cell nuclei and leukocyte borders.

2. Materials and Methods

In our earlier analysis [15], microscopic blood smear photos were segmented, and the blood cells were separated into three main batches: erythrocytes, platelets, and leukocytes. We utilized the leukocyte pictures to be split into five separate classes by machine learning algorithms. Statistical and geometric characteristics of WBC shots were acquired for the input parameters of machine learning algorithms.

2.1. White Blood Cells

White Blood Cells (WBC) also named leukocytes, are made in the bone marrow. Leukocyte cells form nuclei and cytoplasm. They are separated into five groupings: basophil, eosinophil, lymphocyte, monocyte, and neutrophil. Leukocytes, which guard the body against transmittable diseases and unknown substances, constitute a vital part of the immune system. 4×109 - 11×109 units in one liter of a fit grown-up human. That is, a drop in the blood is about 7000 to 25000. Figure 1 depicts a fit adult's average number of white blood cells. Neutrophils are the many common leukocytes in human blood. The kernels consist of 3-5 lobes. Polymorph nuclear include 99% of the cells, while polymorph nuclear cells account for about 70% of the total leukocyte calculation. Eosinophil’s hold onto lots of eosin dye (a type of acid red dye) when dyed, producing their large granules red. Their lifespan is 1-2 weeks, including 2-3% of all leukocytes. They have an average diameter of 10-12 μm, and their cores are two lobes.

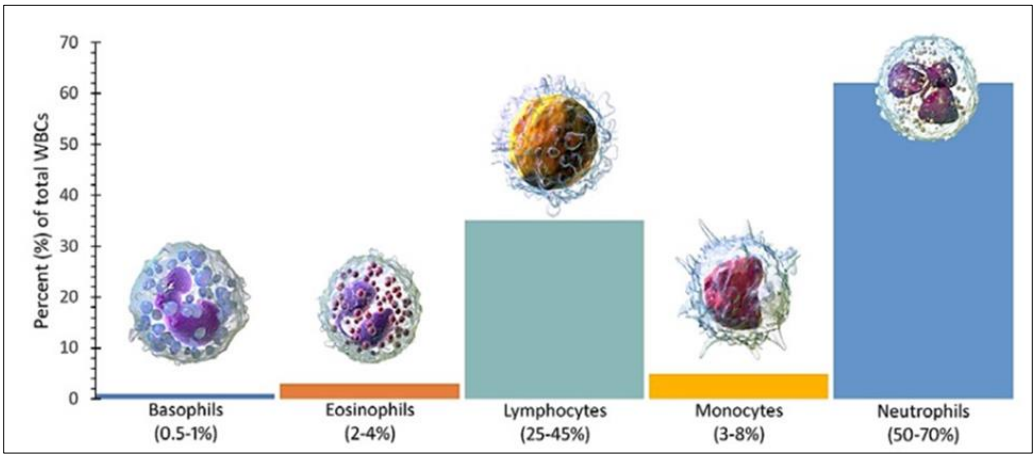

Figure 1 Average values for a normal adult white blood cell count [25].

Another class of leukocytes is named basophil. Its granules hold onto lots of direct dyes and have dark blue-purple color. Its nucleus is rare and consists of two lobes that cannot be resolved. It is also the shortest number of leukocytes. Monocytes are essential peripheral blood cells (15–22 μm). Folds can be noticed in the nucleus, which can be of distinct forms (round, lobular, kidney, bean, or horseshoe). Lymphocytes are cells that can separate and give new lymphocytes. When they meet immunogenic (antigenic) push, morphological change, differentiation, and expansion, it is the most ordinary type of leukocyte in the blood after neutrophils. In this analysis, 350 different WBC photographs were employed as the dataset. Figure 2 displays sample leukocyte cell images utilized in the modification process.

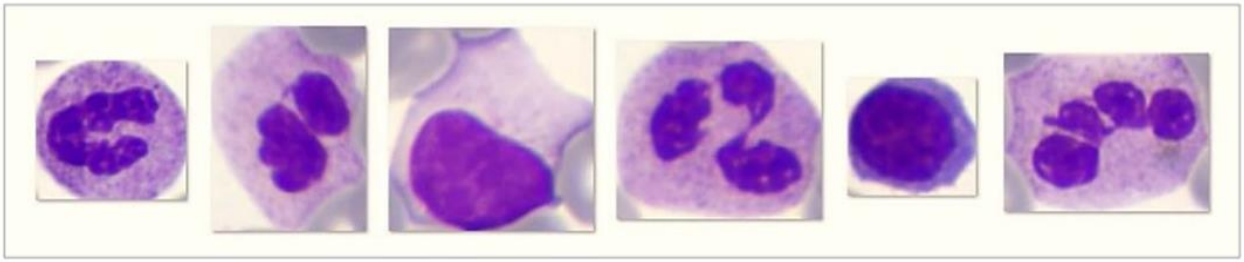

Figure 2 Sample images of leukocyte cells used in the study [25]

Feature Extraction Segment extraction in image processing transforms large quantities of unnecessary data into a reduced display. The method of converting input data into a property dataset is named feature extraction [16]. Feature extraction methods research objects and images to extract the most distinctive segments representing various object classes. Property vectors are employed as input parameters to classifiers assigned to the class to which they are described. The main target of feature extraction is to decrease the original data by scaling specific properties or properties that distinguish a piece of information set from another set. Feature extraction is a necessary procedure in organizing white blood cells [4], and the desired properties affect the execution of the classifiers. The conversion accuracy depends on the number of features and quality properties. An essential part of the studies on microscopic images for modifying leukocyte cells has been based on geometric and tissue-based possessions [17]. The geometric controls used to distinguish cells enclose the shape and size of the nucleus, the form and size of the white blood cell, the number of nucleus lobes, cell circularity, and nucleus rectangularity [18]. Tissue is the exact granule and chromatin-induced belongings in the nucleus. The texture feature [19] contains statistical data such as mean, standard deviation, skewness, kurtosis, and entropy of brightness. This analysis utilized 35 additional statistical and geometric features to classify leukocyte cells. We used the reference polygons of the WBC cells (cytoplasm and nucleus) we segmented in the earlier study.

2.2. Machine Learning Algorithm

Machine Learning Algorithm is the technological paradigm that creates beliefs from the general data using mathematical and statistical procedures and makes projections about the secret with these assumptions. Machine learning is one of the quickest growing areas of computer science, with a broad range of applications. Some scholarly analyses in the past have revealed that after a particular phase, the machines must understand the data. As a result, experimenters conducted their studies to hover various problems by employing multiple symbolic strategies [20]. A considerable number of these procedures can measure, predict, and classify. This unit cites the possessions of machine learning algorithms utilized in this analysis for organizing WBCs.

2.2.1. Judgment Tree Classifier

The Judgment Tree is a machine learning algorithm that can categorize data by constantly dividing the dataset according to a particular criterion. A judgment tree format consists of roots, nodes, branches, and leaves. The tree structure's posterior part and the leaves upper part are called roots. Each piece in the dataset represents nodes. The connection between the nodes is called the branch. It is essential to choose which node to start partitioning in decision trees. If the proper node does not start, the tree's number of nodes and leaves will be very high. Many decision tree learning algorithms are obtainable in the literature. In this study, the C4.5 algorithm was selected.

2.2.2. Random Forest

The Random Forest algorithm was designed by [21]. Instead of creating a single decision tree, this process combines the decisions of multiple multivariate trees, each prepared in distinct training sets. As an impact, it is an algorithm that attains high levels of success in cracking classification issues. In the Random Forest algorithm, determining branching

criteria and specifying a proper pruning strategy, as in the other conclusion tree procedures, is essential. Gain ratio and Gini index are the most typically utilized gain measurement procedures in determining the branching standards. The process of this algorithm is based on two distinct parameters: the number of trees to be formed and the number of models UMAGD, (2019) 11(1), 141-152, [15] utilized for each node. In the sorting process, mainly, the user-defined tree is formed.

2.2.3. . k-Nearest Neighbors

(k-NN) the k-NN algorithm was offered by [22]. K-NN is one of the essential pattern distinction and classification techniques that categorize objects according to the closest training samples in the quality space. The purpose here is to determine which new model belongs to which class, according to the *kk* value of the nearest neighbor. When choosing the category of a new vector, the closest *kk* samples picked from the training data are specified. Hence, the new vector is assigned to it by glancing at the classes in which the selected models belong. A recent example has distinct methods (Euclid, Manhattan, Minkowski, etc.) for computing distances according to classified models.

3. Experimental Results and discussion

Three hundred fifty distinct WBC images were operated in practical studies. These images were aimlessly selected and transformed into five separate training and trial data ratios, as displayed in Figure 5. Each dataset was chosen randomly in expansion 100 times to obtain more practical results. Thus, 500 data was organized and analyzed in statistical results for each dataset level, as displayed in Figure 6.

When the five distinct datasets given in Figure 5 are discussed, For DS1, 25% (87 of them) of 350 WBC pictures were utilized for training, and 75% (263 of them) were employed for testing. Again, other datasets were designed according to the fixed rates. In Figure 6, all datasets are prepared and tested at various speeds for each machine learning algorithm; worst, best, mean, standard deviation, median, and mode values are computed.

The best result acquired by the Random Forest algorithm was noticed in DS4 with 67.3% training and 33% test result with 91.38% classification victory. When the algorithm is considered according to special train and test speeds, it can be said to deliver more than 80% victory. The best result acquired according to the SVM algorithm was seen in DS4 with 67.4% training and 33% for testing goals with 84.48% classification success. In cases where the training details are less than 50%, it is noticed that there are 1 or 2 low outliers. When the algorithm is considered according to various training and test rates, it can be stated that it provides success in about 80%. The most relevant result received according to the MLR algorithm was marked in DS4 and DS5 datasets with 96.55% classification success. When the algorithm is considered according to various training and test rates, it can be expressed that it delivers success in the 90% -95% range. It is the most effective success rate ever gained. Figure 8 displays the relative result graph of all machine learning algorithms according to their best test success. As can be noticed here, the two most delicate algorithms are MLR and Random Forest, respectively.

4. Conclusion

In this analysis, statistical and geometrical characteristics were extracted from microscopic blood photographs, and a feature vector composed of 35 different parameters was formed. This feature vector is the input parameter for six machine learning algorithms to classify white blood cells. Five types of data sets were prepared in additional training and test ratios, 100 different combinations of each data set were created, and statistical results were analyzed to test the performance of the algorithms. When the classification of leukocyte cells is evaluated, it is seen that the highest success rate in all datasets and all conditions belongs to the MLR algorithm. The lowest success rate belongs to the k-NN algorithm and produces results close to the SVM and Naive Bayes algorithms. Apart from these, the Random Forest algorithm is the most successful method after MLR. This method is more successful than the Decision Tree algorithm because it combines more than one decision tree. As a result, the success rate of 95% obtained by the MLR algorithm is relatively high, and at the same time, it is more stable than other methods. Therefore, the method can be applied easily to automatic classification systems. The algorithm can be made more potent by methods such as Bagging, Boosting, or Bootstrapping to improve the classification success further. Thus, it is thought that global success rates can be brought to better values by negatively reducing the factors simulating the sensation of blood smear images.

References

[1] Maji, P., Mandal, A., Ganguly, M. & Saha, S. An Automated Method for Counting and Characterizing Red Blood Cells Using Mathematical Morphology. IEEE International Conference on Advances in Pattern Recognition, Kolkata. 2015; 1-6.

[2] Li, Q., Wang, Y., Liu, H., He, X., Xu, D., Wang, J. & Guo, F. Leukocyte cells identification and quantitative morphometry based on molecular hyperspectral imaging technology. Computerized Medical Imaging and Graphics. 2014; 38 (3): 171-178.

[3] Krzyzak, A., Fevens, T., Habibzadeh, M. & Jelen, Ł. Application of Pattern Recognition Techniques for the Analysis of Histopathological Images. Advances in Intelligent and Soft Computing, Berlin/Germany. 2011; 623-644.

[4] Rawat, J., Bhadauria, H. S., Singh, A. & Virmani, J. Review of leukocyte classification techniques for microscopic blood images. 2nd International Conference on Computing for Sustainable Global Development, New Delhi. 2015; 1948-1954.

[5] Pandit, A., Kolhar, S. & Patil, P. Survey on Automatic RBC Detection and Counting. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 2015; 4(1): 128-131.

[6] Sonar, S. C. & Bhagat, K. S. An Efficient Technique for White Blood Cells Nuclei Automatic Segmentation. International Journal of Scientific & Engineering Research. 2015; 6(5): 172-178.

[7] Sharif, M. H. U., Yamaguchi, K. M. & Ahmed, S. U. Blood Cell Segmentation and Classification by Machine Learning. International Journal of Research in Engineering, Science and Management. 2021; 4(12): 44-47.

[8] Rodrigues, P., Ferreira, M. & Monteiro, J. Segmentation and Classification of Leukocytes Using Neural Networks: A Generalization Direction. Studies in Computational Intelligence. 2008; 83: 373-396.

[9] Joshi, M. D., Karode, A. H. & Suralkar, S. R. White Blood Cells Segmentation and Classification to Detect Acute Leukemia. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS). 2013; 2(3): 147-151.

[10] Tantikitti, S., Tumswadi, S. & Premchaiswadi, W. Image processing for detection of dengue virus based on WBC classification and decision tree. 13th International Conference on ICT and Knowledge Engineering, Bangkok. 2015; 84-89.

[11] Saraswat, M. & Arya, K. V. Automated microscopic image analysis for leukocytes identification: A survey. Micron. 2014; 65: 20-33.

[12] Hiremath, P. S., Bannigidad, P. & Geeta, S. Automated Identification and Classification of White Blood Cells (Leukocytes) in Digital Microscopic Images. International Journal of Computer Applications. 201; 2(8): 59-63.

[13] Habibzadeh, M., Krzyzak, A. & Fevens, T. Comparative study of shape, intensity and texture features and support vector machine for white blood cell classification. Journal of Theoretical and Applied Computer Science. 2013; 7(1): 20-35.

[14] Ramesh, N., Dangott, B., Salama, M. E. & Tasdizen, T. Isolation and two-step classification of normal white blood cells in peripheral blood smears. Journal of Pathology Informatics. 2012; 3(13): 1-10.

[15] Elen, A. & Turan, M. K. A New Approach for Fully Automated Segmentation of Peripheral Blood Smears. International Journal of Advanced and Applied Sciences. 2018; 5(1): 81-93.

[16] Krishnan, A. & Sreekumar, K. A Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images. International Journal of Computer Science and Information Technologies. 2014; 5(6): 7877-7879.

[17] Osowski, S., Siroic, R., Markiewicz, T. & Siwek, K. Application of support vector machine and genetic algorithm for improved blood cell recognition. IEEE Transactions on Instrumentation and Measurement. 2009; 58(7): 2159–2168.

[18] Rosin, P. L. Measuring shape: ellipticity, rectangularity, and triangularity. Machine Vision and App. 2003; 14(3): 172-184.

[19] Tuceryan, M. & Jain, A. K. In the Handbook of Pattern Recognition and Computer Vision 2nd Ed. Chen, C. H., Pau, L. F. and Wang, P. S. P., World Scientific Publishing Co. 1998; 207-248.

[20] Sarle, W. S. Neural Networks and Statistical Models. Proceedings of the Nineteenth Annual SAS Users Group International Conference, Texas. 1994; 1-13.

[21] Breiman, L. Random Forests. Machine Learning. 2001; 45(1): 5-32.

[22] Cover, T., & Hart, P. Nearest Neighbor Pattern Classification. IEEE Transactions on Information Theory. 1967; 13(1): 21-27.

[23] Al-Marzok MI, Majeed KR, Ibrahim IK. Evaluation of maxillary anterior teeth and their relation to the golden proportion in Malaysian population. BMC oral health. 2013; 13(9).

[24] Adjouadi, M., Zong, N. & Ayala, M. Multidimensional Pattern Recognition and Classification of White Blood Cells Using Support Vector Machines. Particle & Particle Systems Characterization. 2005; 22(2): 107-118.

[25] Elen, A. & Turan, M. K. Classifying White Blood Cells Using Machine Learning Algorithms. International Journal of Engineering Research and Development. 2019; 11(1): 141-152.

[26] Mohammed, M. A., Mohammed, M. A., & Mohammed, V. A. Impact of artificial intelligence on the automation of digital health system, International Journal of Software Engineering & Applications (IJSEA), 2022; 13(6); 23-29