engineering proceedings

Proceeding Paper

Automated and Enhanced Leucocyte Detection and Classification for Leukemia Detection using Multi-Class SVM Classifier

Pranav More [1,2,*] and Rekha Sugandhi [3]

[1] MIT School of Engineering, MIT ADT University Pune, Maharashtra, India; e-mail@pranavmore2530@gmail.com
[2] School of Technology Management & Engineering, SVKM's NMIMS University, Navi Mumbai, Maharashtra, India; e-mail@pranavmore2530@gmail.com
[3] MIT School of Engineering, MIT ADT University Pune, Maharashtra, India; e-mail@rekha.sugandhi@mituniversity.edu.in
* Correspondence: e-mail@pranavmore2530@gmail.com

Citation: More, P.; Sugandhi, R. Automated and Enhanced Leucocyte Detection and Classification for Leukemia Detection using Multi-Class SVM Classifier. **2023**, *5*, x. https://doi.org/10.3390/xxxxx
Published: 18 May

Abstract: In this day and age, surrounded by innumerable technology, the use of various autonomous systems to recognize various ailments has tremendously benefited the medical industry. An important medical practice is the visual evaluation and counting of white blood cells in microscopic peripheral blood smears. Invaluable details regarding the patient's health may be revealed, such as the discovery of Acute Lymphatic Leukaemia or other serious disorders. This study provides a paradigm for detecting acute lymphoblastic leukemia from a white blood cell's microscopic vision. Microscopic images must go through a thorough pre-processing phase before being classified. In this study, WBCs are separated from blood smear images using morphological techniques, and the segmented region is then searched for a set of textural, geometrical, and statistical properties. Four different machine learning techniques are used to examine the performance of these algorithms: Random Forest (RF), Support Vector Machine (SVM), Naive Bayes classifier (NB), and K Nearest neighbor (KNN). The SVM is effective in classifying and identifying the acute lymphoblastic cell that produces leukemia malignancy, as can be observed after careful comparison. A single classifier is virtually completely useless given the variety of blood smear pictures. As a result, we thought about using an EMC-SVM to classify leukocytes. The suggested method successfully distinguishes White blood cells from sample blood smear images, and it accurately categorizes each segmented cell into the relevant group.

Keywords: Lymphoblastic; Leukemia; Segmentation; Feature Extraction; PCA; Multiclass classifier; Machine learning; SVM

1. Introduction

A crucial task carried out by doctors is the diagnosis, which involves assessing a dataset to determine whether a disease is present or not. These data, which can include indications, symptoms, photographs, and exams among other things, are crucial for identifying disorders. An incorrect diagnosis brought on by an ineffective examination may result in the patient experiencing side effects since potentially inappropriate medications may be prescribed for the treatment of a certain ailment. There are low-cost computing systems that analyze and interpret the data, offering diagnostic aid to specialists at this important stage.

The three types of blood cells that make up the majority of the blood are red blood cells, platelets, and white blood cells. All tissues receive oxygen from the heart through red blood cells, which also expel carbon dioxide. Up to 50% of the entire blood volume is made up of these. White blood cells (WBCs) play important roles in the immune system as well as being the body's first line of defence against infections and diseases. Therefore, correctly classifying WBCs is crucial and is becoming increasingly required. WBCs can be split into two categories based on the appearance of their cytoplasm.

We can also identify platelets and RBCs and count the number of cells, determine their sizes, and determine the typical cell percentages in human blood by processing microscopic blood smear photographs. The five subcategories of leukocytes are monocytes, lymphocytes, basophils, eosinophils, and neutrophils. Multi-class classification is seen to be the best approach for diagnosing and accurately detecting leukocytes and their underlying sub-class, as it can be utilized to quickly identify each category.

Different picture characteristics, such as edges, geometric, statistical, and statistical features, as well as the histogram of gradients (HOG), are used to categorize images. Preprocessing, which includes noise reduction, contrast control, and image sharpening, is the initial step in the classification of images. Various methods are employed to improve microscopic pictures. The improved image is then further processed to separate the WBCs using various segmentation methods.

In several disciplines, including medical diagnostics systems, important patterns for prediction tasks have been extracted using machine learning-based networks. If the illness is identified early and treated in the interim, death rates can be reduced. It makes things desperate and takes a long time for the hematologist to physically find the sickness. In order to get over these issues, computer-aided techniques for leukemia detection are very effective, quick, and accurate. The computer-aided approach still has a number of issues, obstacles, and research gaps, such as the accuracy of leukemia cancer and its types (acute myelogenous leukemia, acute lymphoblastic leukemia, and multiple myeloma), as well as their further segmentation and categorization. The goal of this study is to determine leukemia's subtype and detect the disease.

2. Literature Review

This section provides a synopsis of the most cutting-edge methods currently available for classifying and segmenting leukocytes. The increasing advancement of methodologies has allowed us to investigate the relevance of leukocyte segmentation and categorization, which play an active part in medical haematology to diagnose various hepatic disorders.

In their work [21], a number of mathematical operations and techniques were developed to reduce noise and enhance image clarity. Then, procedures for gamma correction and contrast enhancement are applied. The accuracy of an account is extensively segmented. Leukocyte localization and f-area extraction were the first two steps in their two-step technique for picture segmentation. Each has three further steps. Localization, thresholding, three-phase filtration, identifying neighboring cells, and cell extraction are examples of sub-steps. At the same time, the cytoplasm, nucleus, and localization of the nucleus areas were extracted.

They utilized thresholding and mathematical morphology in order to section off the nucleus of the cell. Morphology is a mathematical technique that separates white blood cells (WBCs), red blood cells (RBCs), and platelets from one another. This technique employs addition and subtraction to blood smear images. The image was divided into background and foreground using threshold segmentation, and the best threshold value for WBC segmentation was then chosen. For the classification of leukocytes, geometric characteristics were retrieved and an SVM classifier was applied [20].

Several strategies have been presented to address the issue of blood cells that are overlapping. These algorithms divide cells either by eroding and growing regions that keep the shape or by combining concave spots with dividing lines [10, 14]. As part of this study, we also propose an algorithm for cell separation that uses information about the blood cell's shape to construct a conical curve that separates the overlapping sections [12].

3. Proposed Framework

The suggested architecture involves a few steps from an input blood sample image to the final results. The input photos are initially gathered. These photographs often come in a range of sizes and resolutions. They are useful for every type of experiment and analysis due to their homogeneous size. For the distinct region of interest, these photos are subsequently shrunk and color filtered. It can be difficult to find attributes that are appropriate for the goal-mining algorithm. The Canny edge detector and HOG feature descriptor are used to accomplish feature identification and description of the images after rigorous inspection of the images and taking into account our purpose. Principle Component Analysis is used to decrease the dimension of the feature. Random Forest, Naive Bayes Classifier, Support Vector Machine, and Logistic Regression are used for classification. Figure 1 describes the proposed methodology step by step way. In the preprocessing phase, various operations such as resizing, noise removal, and contrast adjustment are carried out and then features are extracted. After that dimension reduction is applied for PCA and finally classification is done using classifiers. At the final stage, multiclass classifier is used for classification.

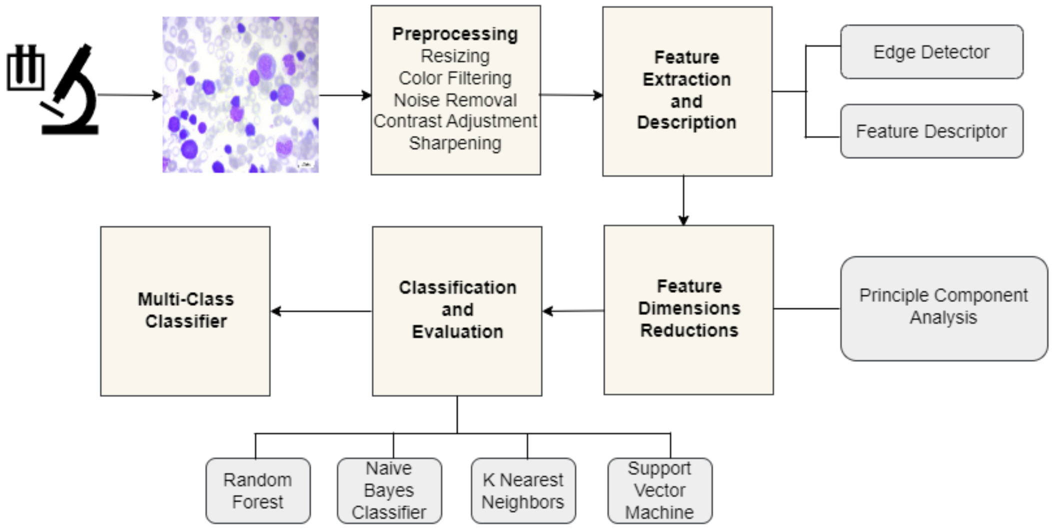

Figure 1. Proposed Framework.

3.1. Dataset Description

The All-IDB1 [13] and Own acquired photos from pathology are used to get microscopic images of white blood cells for the suggested technique. There are 1208 photos in this dataset. 549 of these photos are benign, while 659 of them are malignant. There are roughly 93000 blood components in it. Lymphocytes have been labeled in this dataset. There are 5510 lymphoblasts among the components. Figure 2 describes the sample images from the dataset with various different conditions. These images are used for extracting the features and then final classification.

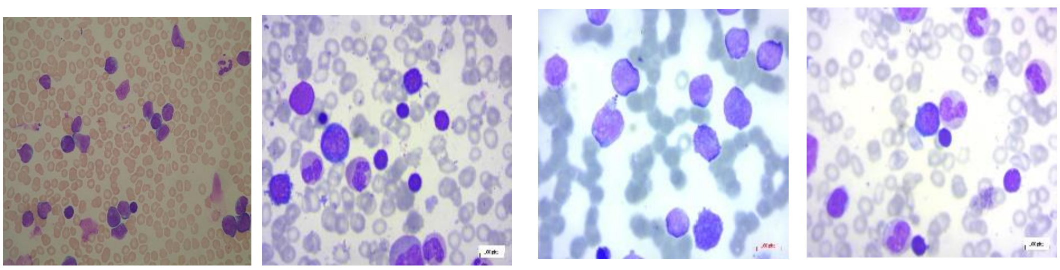

Figure 2. Sample Images from datasets.

3.2. Image Pre-processing

The term "image pre-processing" refers to the preliminary steps taken with an image before any further processing is done. If information is quantified by entropy, then these actions have the opposite effect of increasing the image's information value. The goal of pre-processing is to improve the quality of the picture data by reducing artifacts such as noise and enhancing features like the contrast that will be used in subsequent analysis. When processing a picture, redundancy can be quite helpful. Adjacent pixels representing the same physical object have similar or identical brightness values. If a deformed pixel is located in the image, it can be restored by taking the average of the values of the pixels immediately surrounding it. One way to classify picture pre-processing methods is by the size of the pixel neighborhood used in the calculation of the new pixel brightness. During the pre-processing phase, we eliminated the backdrop, cut off the surplus of blood flow, enhanced the image, reduced the noise, filtered it, and sharpened it.

3.3. Feature Extraction

Before being used for further categorization, the obtained visual data must first go through a process called feature extraction, during which it is turned into a specific collection of features and labeled. At this point, the characteristics of objects that have been segmented from either the entire image or specific areas of the image are retrieved and recognized. In other words, the process of creating a set of features from the visual input in order to recognize patterns is what we mean when we talk about feature extraction. The objects in the image can each be parsed for a variety of characteristics, including shape characteristics (such as area, perimeter, and solidity), texture characteristics (such as homogeneity, energy, angular second, and others), statistical characteristics (such as mean, skewness, and variance), geometrical characteristics (such as perimeter, area, compactness, and symmetry), and color characteristics.

These characteristics can be extracted from the objects in the image. Because blast cells (ROI) carry a multitude of information, including information about their cytoplasm and nucleus, the features extraction stage is critical for identifying the kind of acute leukemia. The retrieved photos show rounded cells. For convenience during feature detection, an intelligent edge detection algorithm is applied to each image. A Gaussian filter is used to smooth out the image and get rid of the noise. Next, we calculate the intensity variations within the image. The two-fold threshold is then applied to locate possible image boundaries. Hysteresis is then used to follow the contours of the image. To finish the edge identification process, weak edges that are not related to strong edges are also suppressed. Figure 3 describes the step-by-step effect on the sample images. Figure 3 (a) is a sample image after the processing and color filtering (b) after performing the operation. Figure 3 (c) shows the effect after feature extraction process and (d) shows after feature description process.

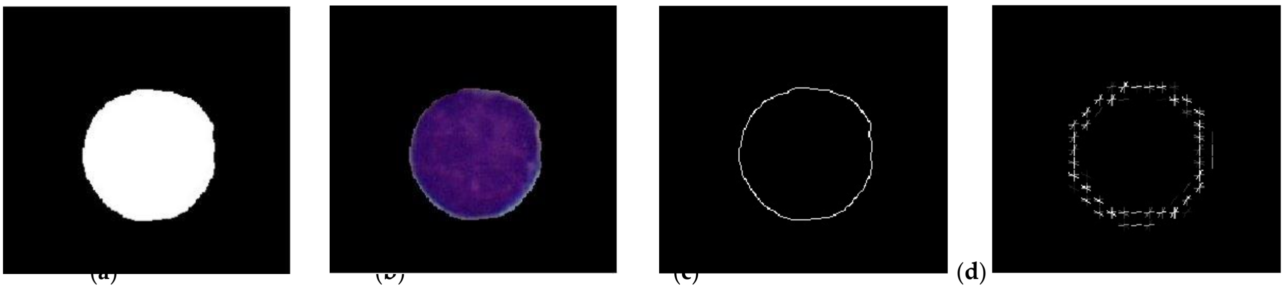
(d)

Figure 3. Sample Images after processing (**a**) Image after color filtering; (**b**) Image after the operation; (**c**) Image after feature extraction; (**d**) Image after feature description.

3.4. Feature Dimension Reduction

Extraction of the features, which comes after segmentation, is a crucial step for accurate classification. Features are adjectives that describe an image and indicate their inherent similarities. The classifier then uses these features and their labels to match various photos and categorize them into distinct classes. Using the HOG feature descriptor, each image is converted into a feature vector, a 1-dimensional array. The speed may be affected by working with this feature vector's 352836 dimensions, which is a relatively large number. The resultant array's dimensions or columns are reduced using the PCA (Principal Component Analysis) dimension reduction approach. The purpose of using PCA was to lower the size of the final feature vector and, as a result, enhance performance. In this study, just the top ten principle components are employed for classification, which results in a feature vector with a size of 10.

3.5. Classification

With so many options available, we've decided to employ a wide range of strategies in this role. To begin, several models were used to perform the classification, some of which used the whole feature vector while others used feature selection to reduce the dimensionality. Classification models such as Naive Bayes, SVM, K-Nearest Neighbour with varying values of K, and random forest have all been tried and tested so far. To begin, we compared the performance of K-Nearest Neighbour, Naive Bayes, and Random Forest, three sequential forward feature selection methods.

Since there is no test set in the dataset, we split it up into five equal portions. Four are used during instruction and two during testing using cross-validation. Matching accuracy, precision, recall, and f1-score are calculated for each test performed on each fold. The results are then averaged to make a model evaluation. Separately evaluating each "fold" of the experiment improves the reliability of the results.

3.6. Multi-class classification

After feature extraction, choosing the appropriate classifier is a crucial step that requires taking into account both the input and the expected outcome. Even though many classifiers are binary classifiers, multi-class classification can be accomplished with them. In the suggested work, we divided leukocytes into five classes using a multi-class categorization. This is as a result of the variety of blood smear pictures for which it is impractical to train a single classifier due to its poor performance. It has been demonstrated through experimentation that multi-class classifiers outperform conventional techniques.

4. Result and Discussion

Both qualitative and quantitative aspects of the experimental results of the suggested hybrid model classification methodology are provided. Using the information we acquired, we put the suggested strategy to the test. For the aim of diagnosis, the leukocytes were specifically segmented so that the structure and colour of their nuclei could be seen clearly. The underlying truth is contrasted with the suggested technique. This algorithm is capable of accurately locating and dividing the five kinds of leukocytes. Three metrics are used to evaluate the proposed segmentation technique: false positive rate (FPR), false negative rate (FNR), and F- measure. Table II illustrates the result obtained from the proposed model. Table 1 focuses on the performance metrics of various classifiers used in the research. Table 1 of comparative analysis describes the accuracy, precision, recall and F-score values for classifiers. Support vector machine shows the best results in terms of performance metrics comparative to other classifiers. Figure 4 describes the comparison graph for analysis purpose. This figure focuses on performance of classifier in terms of performance metrics defined in the research.

Table 1. Comparative Analysis.

Classifier	Precision (%)	Recall (%)	F-score (%)	Accuracy (%)

Random Forest	98.2	97.60	97.60	97.60
Support Vector Machine	99.00	98.80	98.80	98.85
K-Nearest Neighbors	97.30	96.60	96.60	97.10
Naïve Bayes	98.20	97.80	97.80	98.08

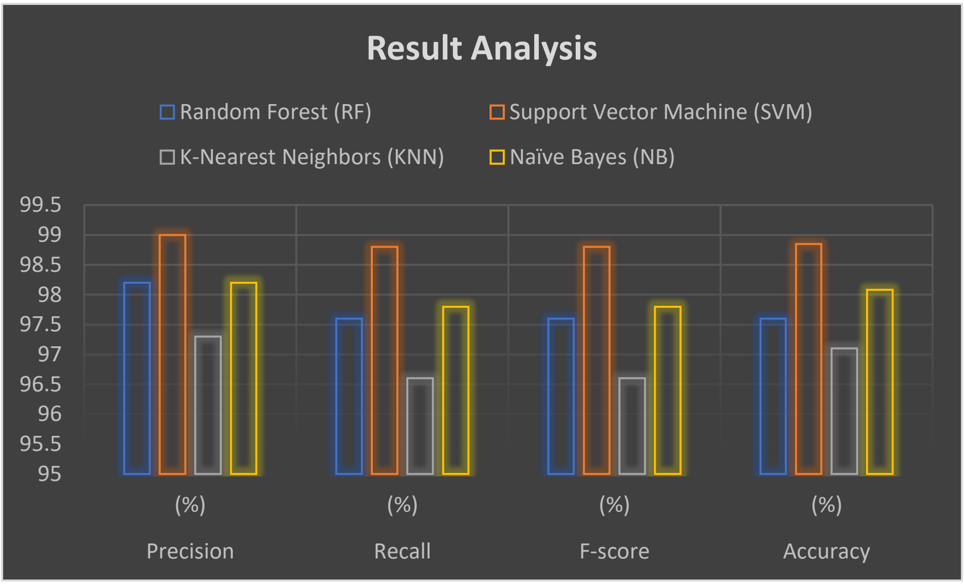

Figure 4. Result Analysis.

5. Conclusion

Leukaemia is a type of blood cancer that commonly affects children and adults. The type of cancer and the extent of its dissemination throughout the body affect leukemia treatment. For the patient to receive the right care and heal, the disease must be identified as soon as feasible. This study presents a novel strategy for the totally automatic identification and classification of leukocytes utilizing microscopic images. The purpose of this work is to provide an automated technique to support medical activity in the detection of acute lymphocytic leukemia (ALL). In order to do so, we have presented a revolutionary approach. The suggested approach successfully separates WBCs from blood smear images and correctly categorizes each segmented cell, according to experimental data. When compared to other classifiers, the suggested classifier was shown to have greater accuracy. Multi-class classifiers improve overall accuracy for classifying different leukocyte subtypes. Additionally, increasing the dataset's size will be necessary to provide the classification model with more examples to use during the training phase and to enable us to apply a validation method other than the 10-fold Cross-Validation.

References

1. Minetto, Dorini and Leite. White blood cell segmentation using morphological operators and scale-space analysis. SIBGRAPI 07: Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing. 2007.
2. Deng, Liao. An accurate segmentation method for white blood cell images. IEEE International Symposium on Biomedical Imaging. 2002; 245–248.
3. Yang L, Meer P, Foran DJ. Unsupervised segmentation based on robust estimation and color active contour models. IEEE Trans.Inf. Technol.Biomed. 9. 2005; 475–486.
4. BM Silva, JJPC Rodrigues, I de la Torre Díez, M López-Coronado, and K Saleem. Mobile-health: A review of current state in 2015. J. Biomed. 2015; 265-272.
5. Yi F, Chongxun Z, Chen P, Li L, White blood cell image segmentation using on-line trained neural network. In: Proceedings of 27th International Conference on Engineering in Medicine and Biology Society. 2006; 6476–6479.
6. Yampri P, Pintavirooj C, Daochai S, Teartulakarn S. White blood cell classification based on the combination of Eigen cell and parametric feature detection. In: IEEE Conference Industrial Electronics and Applications. 2006; 1–4.

7. Chinwaraphat S, Sanpanich A, Pintavirooj C, Sangworasil M, Tosranon P. A modified fuzzy clustering for white blood cell segmentation. In: Proceedings of the Third International of Symposium on Biomedical Engineering. 2008; 356–359.
8. FI Sholeh. White blood cell segmentation for fresh blood smear images. Proc. Int. Conf. Adv. Computer Sci. Inf. Syst. 2013; 425-429.
9. S Nazlibilek, D Karacor, T Ercan, MH Sazli, O Kalender and Y Ege. Automatic segmentation, counting, size determination and classification of white blood cells. 2014; 58-65.
10. Won C, Nam JY, Choe Y. Segmenting cell images: a deterministic relaxation approach. In Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, Springer Berlin Heidelberg. 2004; 281–291
11. S Ravikumar. Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine. Nanomed, Biotechnol. 2016; 985-989.
12. Dorini LB, Minetto R, Leite NJ. White blood cell segmentation using morphological operators and scale-space analysis. IEEE XX Brazilian Symposium on Computer Graphics and Image Processing. 2007; 294–304
13. R. D. Labati, V. Piuri, and F. Scotti, “All-IDB: The acute lymphoblastic leukemia image database for image processing,” in 2011 18th IEEE International Conference on Image Processing. IEEE, sep 2011, pp. 2045–2048. [Online]. Available: http://ieeexplore.ieee.org/document/6115881/
14. Morea, M. P., & Sugandhib, R. (2021). A Review on Systematic Investigation of Leucocytes Identification and Classification Techniques for Microscopic Blood Smear.
15. NIC Marzuki, NH Mahmood and MAA Razak. Segmentation of white blood cell nucleus using active contour. 2015; 115-118.
16. BC Ko, JW Gim and JY Nam. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. Micron. 2011; 695-705.
17. F Zamani and R Safabakhsh. An unsupervised GVF snake approach for white blood cell segmentation based on nucleus. Proc. 8th Int. Conf. Signal Process. 2006; 1453-1456.
18. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, Larson RA. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert. 2010; 453-474.
19. Gokbuget N and Hoelzer D. Treatment of adult acute lymphoblastic leukemia. Paper presented at the Seminars in hematology. 2009; 64-75.
20. J. Prinyakupt and C. Pluempitiwiriyawej, ‘‘Segmentation of white blood cells and comparison of cell morphology by linear and naíve Bayes classifiers,’’ Biomed. Eng. OnLine, vol. 14, no. 1, p. 63, 2015
21. Bodzas, A.; Kodytek, P.; Zidek, J. Automated detection of acute lymphoblastic leukemia from microscopic images based on human visual perception. Front. Bioeng. Biotechnol. 2020, 8, 1005
22. LB Dorini, R Minetto and NJ Leite. Semiautomatic white blood cell segmentation based on multiscale analysis. IEEE J. Biomed. Health Information. 2013; 250-256.
23. SF Bikhet, AM Darwish, HATolba and SI Shaheen. Segmentation and classification of white blood cells. Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. 2000; 2259-2261.
24. Lim G C C, Rampal S, & Yahaya H. Cancer Incidence in Peninsular Malaysia- The Third Report of the National Cancer Registry, Malaysia. National Cancer Registry. 2008.
25. Nilanjan Ray. A concave cost formulation for parametric curve fitting: Detection of leukocytes from intravital microscopy images. IEEE International Conference on Image Processing. 2010.
26. Anjali Gautam, Priyanka Singh, Balasubramanian Raman, Harvendra Bhadauria. Automatic classification of leukocytes using morphological features and Naïve Bayes classifier. IEEE Region 10 Conference. 2016.
27. Filip Novoselnik, Ratko Grbic, Irena Galic, Filip Doric. Automatic White Blood Cell Detection and Identification Using Convolutional Neural Network. IEEE International Conference on SST. 2018.
28. S Dhiravida Janani, R Muthu Selvi, G Mlndhu. Blood Cell Detection and Counting Using Convolutional Sparse Dictionary Learning. IEEE International Conference on Current Trends towards Converging Technologies. 2018.
29. Anjali Gautam, Harvindra Bhadauria. Classification of white blood cells based on morphological features. IEEE International Conference on Advances in Computing, Communications and Informatics. 2014.
30. Muhammad Sajjad, Siraj Khan, Muhammad Shoaib, Hazrat Ali, Zahoor Jan, Khan Muhammad, Irfan Mehmood. Computer Aided System for Leukocytes Classification and Segmentation in Blood Smear Images. IEEE International Conference on Frontiers of Information Technology. 2016.
31. Shubham Manik, Lalit Mohan Saini, Nikhil Vadera. Counting and classification of white blood cell using Artificial Neural Network. IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems. 2016.
32. Zahra Khandan Khadem Alreza, Alireza Karimian. Design a new algorithm to count white blood cells for classification Leukemic Blood Image using machine vision system. IEEE International Conference on Computer and Knowledge Engineering. 2016.
33. Bruno C Gregorio da Silva, Ricardo J. Ferrari. Detection of Leukocytes in Intravital Video Microscopy Based on the Analysis of Hessian Matrix Eigenvalues. SIBGRAPI Conference on Graphics, Patterns and Images. 2015.

Disclaimer/