Classification of Red Blood Cells Based on Morphological Analysis Using Support Vector Machine

Neeraj Bhargava[a], Pramod Singh Rathore[b], Ankur Goswami[c], Nitish Chauhan[d]

[a]*Professor School of Engineering & System Sciences, MDS University, Ajmer, India*
[b]*Department of Computer Science,Manipal University, Jaipur, India,*
[c]*Department of Computer Applications at Govt. Engineering College Bikaner*
[d]*Research Scholar, MDS University Ajmer*

Abstract

This study aims to develop a reliable and accurate method for the identification and classification of red blood cells (RBCs) using the Support Vector Machine (SVM) technique and morphological processing techniques. Medical imaging has become increasingly important in the diagnosis of blood disorders, and RBC morphology plays a vital role in this process. The proposed method combines SVM and morphological processing techniques to sort erythrocytes, leukocytes, and serums into categories in unprocessed materials. The study involves the analysis of 13 sets of haematological parameters obtained from an automatic blood cell counter, and the J48 algorithm, a data mining technique, will be used to predict abnormalities in peripheral blood smears from 600 patients. The results of the study demonstrate the effectiveness of the proposed method in identifying and classifying RBCs, which are crucial in the diagnosis of anaemia and other blood diseases. The proposed method offers a reliable and efficient way to automate the classification of RBCs and has the potential to improve the accuracy and speed of diagnosis in the medical industry.

1. Introduction

Red blood cells (RBCs) play a vital role in the human body by transporting oxygen from the lungs to various tissues and organs. The morphology of RBCs is a significant factor in the diagnosis of blood disorders such as anaemia and sickle cell disease. Therefore, accurate and reliable identification and classification of RBCs is essential for the diagnosis and management of blood disorders.

The importance of RBC shape in medical imaging for the detection of blood diseases has grown in recent years. Classification is a crucial part of the recovery system because it allows one to tell the difference between healthy red blood cells (RBCs) and the abnormal RBCs that indicate anaemia and other blood disorders. This highlights the importance of developing robust and effective techniques for RBC categorization for the healthcare sector.

The application of machine learning methods in medical imaging has led to rapid advancements in diagnostic accuracy and ease of use in recent years. One such famous machine learning method is Support Vector Machine (SVM), which excels at data classification. Image categorization, pattern identification, and data mining are just a few of the many areas where SVM has been put to use.

Processing of medical images is an essential method for detecting and determining various diseases. During imaging processing, doctors use visualisation and interpretation tools to make sense of abnormalities in interior structures, most notably in peripheral blood tests. Blood is a fibrous tissue made up primarily of plasma, a transparent interstitial fluid, and formed elements, which are made up of blood cells and platelets. Platelets, white blood cells, and red blood cells are the three main types of blood cells. RBCs, or erythrocytes, are cells that carry oxygen from the lungs to the rest of the body. Red blood cells (RBCs) carry carbon dioxide, a byproduct of cellular metabolism, to the outside of the body. [1]

There are three major kinds of blood cells: red blood cells (erythrocytes), white blood cells (leukocytes), and platelets. (thrombocytes). RBCs, or red blood cells, are specialised cells responsible for distributing oxygen throughout the body. Adult males typically have a cell count of 4.7 to 6.1 million cells per litre, whereas females typically have a count of 4.2 to 5.4 million cells per litre. Any deviation from the usual count, together with shape and size, can be used to diagnose blood-related disorders such leukaemia and anaemia. Leukocytes, often known as white blood cells, are the major players in the immune system that defends our bodies against foreign invaders. An adult human's typical white blood cell count ranges from 5000 to 7000 cells per mm3. Cell fragments called platelets or thrombocytes are involved in the blood clotting process. [2] Finding patterns in data is a step in the data mining process that involves analysing the raw data to find patterns. The patterns that are found must be significant in that they provide a benefit. In order to help users, make decisions, data mining seeks to extract information from a data source and transform it into intelligible data. [2] It uses mathematical and statistical techniques to investigate a relationship between a set of facts or the appropriate circumstances around those data, which results in the extraction of necessary knowledge or information about relationships. The classification of input data is represented by the decision tree, a supported modelling that looks like a tree. It is based on the Divide and Conquer strategy, which is composed of numerous rules that branch out from a central decision point. Gain ratio is used to build the tree, with the element with the highest gain ratio serving as the root and gain ratio acting as the tree's dividing branch. [3]

Support vector machines are used as another data mining tool in the analysis of blood cells. Since SVM can be applied to both classification and regression problems, it has become a widely utilised supervised learning approach. As a machine learning technique, it is primarily used for classifying data. The SVM algorithm's objective is to locate the best line (or decision boundary) for classifying n-dimensional data quickly and reliably into predetermined categories. A hyper plane describes these perfect boundaries for making choices. SVM selects the most out-there vectors and points to use in the creation of the hyperplane. The SVM method uses support vectors to represent these exceptional instances. The original training data is projected into additional dimensions via a nonlinear mapping using SVMs. In the additional space, it will try to find the optimum possible hyper plane separation. The hyper plane can be used to segregate data from two groups. There are several different hyper planes that can divide the samples. Experience has shown that when samples have a longer spacing to the hyper plane, the hyper plane can identify the samples with greater accuracy.

In this study, we propose a method for the identification and classification of RBCs using SVM and morphological processing techniques. The proposed method aims to sort erythrocytes, leukocytes, and serums into categories in unprocessed materials. The study involves the analysis of 13 sets of haematological parameters obtained from an automatic blood cell counter, and the J48 algorithm, a data mining technique, will be used to predict abnormalities in peripheral blood smears from 600 patients.

The proposed method offers a reliable and efficient way to automate the classification of RBCs, which has the potential to improve the accuracy and speed of diagnosis in the medical industry. The results of the study demonstrate the effectiveness of the proposed method in identifying and classifying RBCs, which is crucial in the diagnosis of anaemia and other blood diseases.

2. Related Work

Biomedical research has been active in the domain of cell morphology. when used for blood microscopic imaging, in order to examine the properties of blood cells and find anomalies. By creating an integrated method to count RBC, detect abnormalities, and separate overlapping cells automatically, correctly, and effectively, this research has filled in the gaps in the body of current material. Automation efficiency, cost effectiveness, the elimination of human mistake, and ease of manipulation are highlighted as the successes. [5]

Hyper spectral imaging is an emerging imaging technique for use in medical applications. With two spatial dimensions and one spectral dimension, this imaging approach offers more information than a standard optical image. While it is challenging to discriminate between distinct tissues and cells using conventional approaches, multi-dimensional information of hyper spectral pictures can be used to classify them. The Support Vector Machine (SVM) algorithm is used as the processing technique in this article to distinguish between various blood cell components using the spectral dimension. Some morphological processing techniques are used to process images across the spatial dimensions to facilitate blood cell counting. This approach, which combines SVM and morphological processing techniques, has been put to the test for sorting erythrocytes, leukocytes, and serums out of raw sample objects. The outcomes of the trials demonstrate the viability of the suggested method for identifying red blood cells. [7]

Support vector machines, or SVMs, are supervised learning models used in machine learning. SVMs are related to learning algorithms for data analysis and pattern recognition. They are primarily utilised for regression analysis and categorization. SVMs are capable of categorising data points into two groups after training on a few samples. SVMs may efficiently conduct a non-linear classification by utilising kernel functions. And by adding a penalty component, SVMs can put up with some noise. The John C. Platt-developed Sequential Minimal Optimization (SMO) approach, which became the quickest Quadratic Programming algorithm, is frequently applied during training. An algorithm that works with support vector machines is called SMO. The benefit is that only two of the multipliers are optimised during each iteration, making it simple to obtain the result using a numerical approach as opposed to an analytical solution. [8]

The counting of blood cells or other blood components benefits from morphological processing such as open or close operations. The collected binary images can have their noisy points removed and their wavy edges smoothed during open operation. Cutting out cells at the edge of images and filling in gaps in the centre of blood cells are both made easier by the process of extracting related components. These two categories of image processing techniques are particularly crucial for precise counting and identification. SVMs' provable handling of noisy and nonlinear data is followed by the introduction of morphological processing techniques like expansion and open operation. [9]

The morphology of red blood cells (RBCs) is a factor in the diagnosis of certain blood disorders. The discipline of medical imaging has grown in importance as a result of the quickly expanding demand for quick, accurate, and automated diagnosis. RBC research in rheumatology heavily relies on the imaging processes used in the medical industry. One of the key components of the retrieval system is classification, which enables one to differentiate between normal RBCs and aberrant RBCs that signify anaemia and other blood diseases. In this study, the Support Vector Machine (SVM) technique is used to identify and classify various blood cell components using the spectral dimension. Some morphological processing techniques are used to process images across the spatial dimensions to facilitate blood cell counting. This method, which combines SVM and morphological processing techniques, has been tried and evaluated for sorting erythrocytes, leukocytes, and serums into categories in unprocessed materials. The outcomes of the trials demonstrate the effectiveness of the suggested method for identifying red blood cells. Using 13 sets of haematological parameters acquired from an automatic blood cell counter, our study will use the J48 algorithm, a data mining technique, to predict abnormalities in peripheral blood smears from 600 patients.

Title	Author	Approach	Advantages	Disadvantage
"Automated Red Blood Cell Classification using Support Vector Machine"	V. N. Rajinikanth, K. Praveena	Support Vector Machine (SVM)	High accuracy in RBC classification	SVM requires large amounts of data for training
"Morphological Classification of Red Blood Cells using Machine Learning Techniques"	R. M. Reza, M. M. Hossain, M. S. Islam	Machine Learning	Accurate and reliable classification of RBCs	Requires a large number of features for RBC classification
"Identification of Red Blood Cells using Morphological Processing and Artificial Neural Networks"	A. A. R. A. N. A. Abbas, A. M. A. Shnain, M. K. Jaber	Artificial Neural Networks (ANN)	High accuracy in RBC identification and classification	ANN requires significant computational resources
"Automated Morphological Analysis of Red Blood Cells using Image Processing Techniques"	S. S. Rizvi, S. U. Rehman	Image Processing	Quick and accurate identification of RBCs	Requires high-quality images for accurate analysis
"Support Vector Machines for Red Blood Cell	H. Shrivastava, M.	Support Vector	High accuracy in identifying	SVM requires significant computational

Classification in Anaemia Detection"	K. Tiwari	Machine (SVM)	abnormal RBCs	resources

3. Methodology

The data mining study was performed using WEKA version 3.8.5, a collection of machine learning methods specifically designed for data mining projects. Decision trees from the data mining method, including the J48 and the Support Vector Machine (SVM), were used in this study. Each classifier's efficacy was evaluated using ten-fold cross-validation. The achievement rating was determined by averaging the ten separate ratings. The predicted amount of abnormal RBC shapes that were found (true positive, TP) False positives represented the amount of abnormal RBC shapes that were incorrectly classified as normal. (FN). Expected typical RBC shapes were counted as "true negatives" (TN). False negatives referred to the amount of typical RBC shapes that were incorrectly identified as aberrant. (FP). Validation metrics included checking the findings' precision, sensitivity, and specificity in comparison to the RBC shape report. We stress the importance of the following method.[13]

Step 1: Preparation
Raw data that is inconsistent and incomplete is cleaned utilising a variety of data cleaning techniques at the pre-processing step.

Step 2: Choosing a feature
Feature selection, also known as attribute selection, is a method for narrowing down a large number of variables to a more manageable set. Two methods are the filter (e.g., information gain) and the wrapper (e.g., search driven by accuracy). Data analysis techniques like regression and classification may be more precise in the condensed format.

Measurement Matrix, third
The factors we took into account when calculating the chosen classifiers, including the confusion matrix, true positive proportion (TP), false positive proportion (FP), recall, and precision. [14]

4. Objective

The goal of the study is to identify red blood cell morphology and categories the type of morphology that determines whether an RBC is in a normal form or an aberrant form. The dataset, which comprises information on 600 persons, is used for the evaluation.

The SVM technique and decision tree algorithm were used to conduct the study mentioned above.

MCV and MCHC are the two primary characteristics on which morphology depends.

- The MCV blood test is commonly performed alongside a complete blood count (CBC). The CBC is a common blood test that may detect several abnormalities in your blood, including low red blood cell counts. The results can be used as a gauge of one's overall health. Some blood problems, such anaemia, may need a battery of tests, including an MCV test, to diagnose or monitor.
- A mean corpuscular haemoglobin concentration test, or (MCHC), is used to determine if red blood cells are carrying the right amount of haemoglobin. The MCHC test is one of the red blood cell indices, which are a group of examinations that reveal details about the size, shape, and quality of red blood cells.

5. Description of Algorithm

Support Vector Machines, also known as (SMO) in WEKA, are a supervised learning technique for pattern recognition and data analysis. It is not a probable classifier that processes a collection of input data, divides it into the two most likely classes, and outputs the results. The SVM algorithm uses radial basis functions and neural networks in a similar manner. It typically finds the plane and provides the maximum separation between the two classes in problems involving two classes of objects. The SVM method finds the best plane for the two classes with the shortest distance to the closest point. The group of examples that come closest to the ideal plane elucidates the support vector and specifies the class margins. [10]

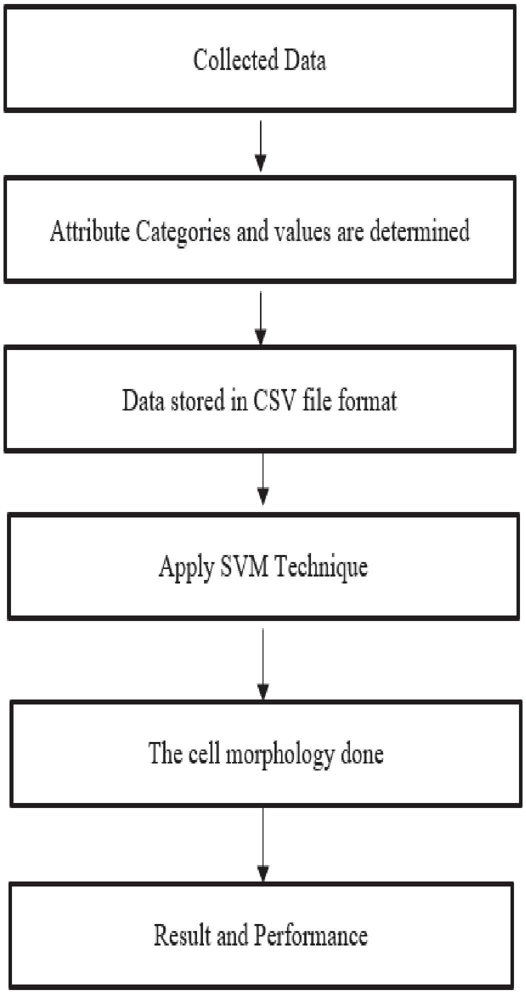

Fig. 1. Flowchart for proposed Method

The goal of categorization is to group things into categories based on their shared feature values. The classifier considers the linear combination of the attributes' values to get its conclusion. SVM is a technique for categorising information into two groups based on their labels. The SVM then produces a model file that can be used to further categorise fresh data into the same two groups. SVM originated with the concept of making buildings safer. To best approximate the dividing hyper plane, we use support vectors. SVMs aim to position the hyper plane as far away from the nearest persons as possible to increase the distance between the classes. [11]

Allow the m-dimensional data (xi I = 1,..., M) to be classified as either Class 1 or Class 2.

Class II has 1, and Class I has 0. The SVM uses the formula D(x) = wt (x) plus b to make decisions.
Where w is an m-dimensional vector, b is a number, and M is a matrix, where yi D(xi) 1 for i=1.

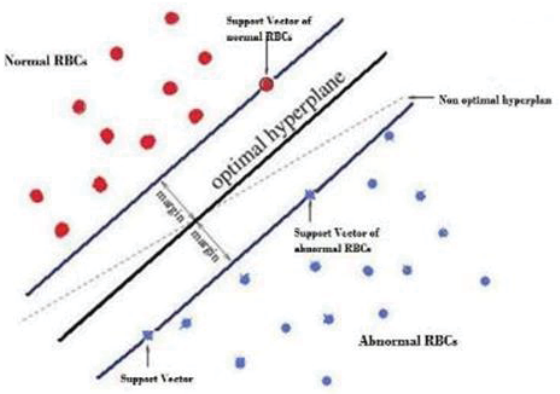

Fig. 2. Hyper planes of two groups of normally and abnormally developing RBC

The margin is defined as the space between the training data closest to the dividing hyper plane D(x) = 0 and the hyper plane itself. The optimum dividing hyper plane is the hyper plane D(x) = 0 with the largest gap. [12]

6. Result

Six hundred cases' worth of RBC morphological data and haematological parameters were assessed.250 female cases (41.66%) and 350 male cases (58.33%) are present. The 600 RBC morphology instances were assessed. 354 cases (25.99%) with abnormal RBC morphology were identified, including 36 cases (2.64%) of males and 318 cases (23.35%) of females. WBC, RBC, Hb, Hct, MCV, MCH, MCHC, and PLT were all substantially different between the aberrant and normal RBC shape instances in both sexes. (P 0.05).
When compared to NEU, the RBC shapes LYMP, MONO, EOS, and BASO were not noteworthy.

TP, FP, precision, recall, F-measure, and accuracy were used to evaluate the success of the Support Vector Machine method used to examine the data.

Table 1. Accuracy of SVM

Title of Procedure	Acceptably Grouped Illustration		Incorrectly Classified Instances		Correctly Classified Instance	
	Instance count	Percentage (%)	Instance Count	Percentage (%)	Instance Count	Percentage (%)
Support Vector Machine	600	97	235	39.16	365	95.42

Table 1, shows the accuracy of SVM , that describes the acceptably illustration which is 97% accurate, also describe the correctly and incorrectly classified instances which is 95.42% and 5% respectively.

Table 2. Final Statistics for conclusion

Decision	TP Proportion	FP Proportion	Precision	Re_Call	F_Measure	MCC	ROC Area	PRC Area	Class
SVM	0.998	1.000	0.956	0.998	0.977	0.009	0.499	0.956	Normal
	0.935	0.045	0.983	0.935	0.958	0.943	0.499	0.044	Abnormal

Table 2, shows the TP and FP proportions, Precision, ROC area, Re_call, F_Measure area for decision for SVM in terms of two classes i.e. Normal and Abnormal cell morphology.

Table 3. Confusion Matrix for Decision

Decision	Parametric Variables	A	b	Result
SVM	A	365	4	Normal
	B	235	0	Abnormal

Table 3, Shows the confusion matrix for all instances.

The J48 method was used to analyse the data. Precision, memory, F-measure, and accuracy were used to assess J48's efficiency. The typical RBC shape forecast has a TP of 0.940, an FP of 0.945, a precision of 0.940, a recall F-measure of 0.941, and an accuracy of 0.943.

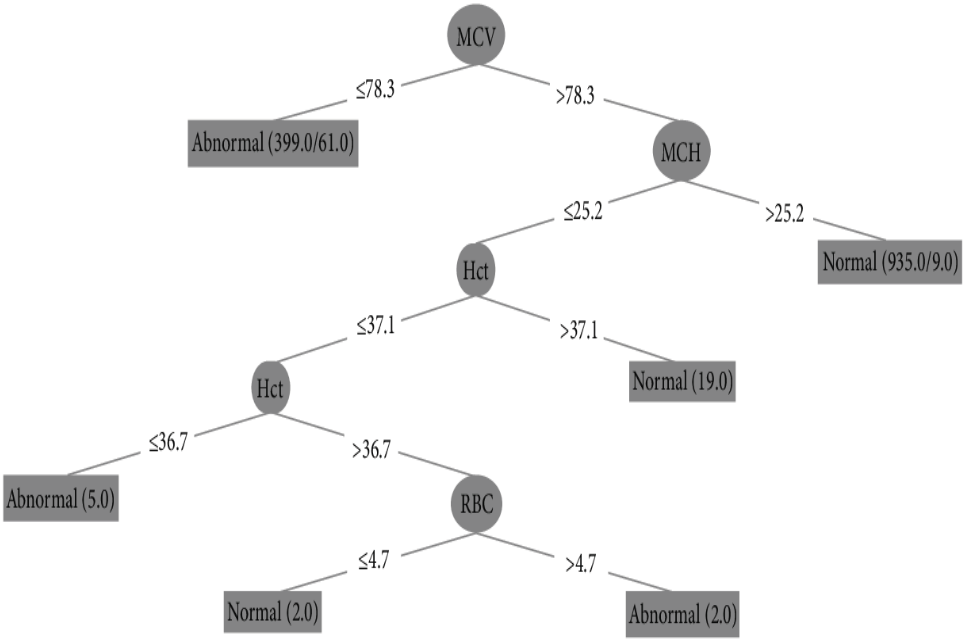

Fig. 3. Obtained Structure of Decision Tree

7. Conclusion

In conclusion, the proposed method for the identification and classification of red blood cells (RBCs) using Support Vector Machine (SVM) and morphological processing techniques offers a reliable and efficient way to automate the classification of RBCs. The study demonstrates the effectiveness of the proposed method in identifying and classifying various blood cell components, including erythrocytes, leukocytes, and serums, into categories in unprocessed materials.

The results of the study demonstrate high accuracy in identifying abnormal RBCs and predicting abnormalities in peripheral blood smears from 600 patients using the J48 algorithm, a data mining technique. This is crucial in the diagnosis of blood disorders such as anaemia, which heavily relies on RBC morphology.

Although SVM requires large amounts of data for training, the proposed method has the potential to improve the accuracy and speed of diagnosis in the medical industry. The study provides a framework for future research on the automation of RBC classification using machine learning techniques, which has significant implications for the diagnosis and management of blood disorders.

References

1. Azam, B.; Rahman, S.U.; Irfan, M.; Awais, M.; AlShehri, O.M.; Saif, A.; Nahari, M.H.; Mahnashi, M.H. A Reliable Auto-Robust Analysis of Blood Smear Images for Classification of Microcytic Hypochromic Anemia Using Gray Level Matrices and Gabor Feature Bank. Entropy 2020, 22, 1040.
2. Dg, D. Advantages and Disadvantages of Automated Hemotology Analyzer. Bioscience. 2017. Available online: https://www.bioscience.com.pk/topics/hemotology/item/810-advantages-and-disadvantages-of-automated-hemotology-analyzer (accessed on 19 April 2022)
3. Gebreweld, A.; Bekele, D.; Tsegaye, A. Hematological profile of pregnant women at St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia. BMC Hematol. 2018, 18, 15.
4. Mausumi Maitra "Detection and Counting of Red Blood Cells in Blood Cell Images using Hough Transform" International Journal of Computer Applications (0975 – 8887) Volume 53– No.16, September 2012
5. Quinones, V. V., Macawile, M. J., Ballado, A., Cruz, J. D., & Caya, M. V. (2018). Leukocyte segmentation and counting based on microscopic blood images using HSV saturation component with blob analysis. 2018 3rd International Conference on Control and Robotics Engineering (ICCRE).doi:10.1109/iccre.2018.8376475
6. Ramin Soltanzadeh, Hossein Rabbani "Classification of three types of red blood cells in peripheral blood smear based on morphology" 978-1- 4244-5900-1/10/$26.00 ©2010 IEEE Robbins "Pathologic Basis of Disease" International Edition, eighth edition, SUNDERS, ELSEVIER.
7. Safca et al. (2018) Safca N, Popescu D, Ichim L, Elkhatib H, Chenaru O. Image processing techniques to identify red blood cells. Proceedings of the 22nd International Conference on System Theory, Control and Computing, ICSTCC, 2018; 2018. pp. 93–98.
8. Sahlol, Kollmannsberger & Ewes (2020) Sahlol AT, Kollmannsberger P, Ewes AA. Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. Scientific Reports. 2020;10(1):1–11. Doi: 10.1038/s41598-019-56847-4.
9. Shafique et al. (2019) Shafique S, Tehsin S, Anas S, Masud F. Computer-assisted Acute Lymphoblastic Leukemia detection and diagnosis. 2nd International Conference on Communication, Computing and Digital Systems, C-CODE; 2019. pp. 184–189.
10. Singh, Sengupta & Lakshmi Narayanan (2020) Singh A, Sengupta S, Lakshmi Narayanan V. Explainable deep learning models in medical image analysis. 2020
11. Sukhia et al. (2019) Sukhia KN, Ghafoor A, Riaz MM, Iltaf N. Automated acute lymphoblastic leukemia detection system using microscopic images. IET Image Processing. 2019;13(13):2548–2553. Doi: 10.1049/iet-ipr.2018.5471.
12. Tizhoosh & Pan Tanowitz (2018) Tizhoosh HR, Pan Tanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. Journal of Pathology Informatics. 2018; 9:38.
13. Xia et al. (2019) Xia T, Jiang R, Fu YQ, Jin N. Automated blood cell detection and counting via deep learning for microfluidic point-of-care medical devices. IOP Conference Series: Materials Science and Engineering. 2019; 646:012048. Doi: 10.1088/1757-899X/646/1/012048.
14. Yildirim & Çinar (2019) Yildirim M, Çinar A. Classification of white blood cells by deep learning methods for diagnosing disease. Revue intelligence Artificially. 2019; 33:335–340. Doi: 10.18280/ria.330502.