

A comparison of automated classification techniques for image processing in video internet of things

Roojwan Sc Hawezi^{a,*}, Farah Sami Khoshaba^a, Shahab Wahhab Kareem^{a,b}

^a Department of Information System Engineering, Erbil Technical Engineering College, Erbil Polytechnic University, Erbil, Iraq

^b Department of information technology, College of Engineering and Computer Science, Lebanese French University, Kurdistan Iraq

ARTICLE INFO

Keywords:

Digital image
Features extraction
Machine learning
Segmentation
Video internet of things

ABSTRACT

The counts of various types of white blood cells give vital information for identifying a variety of ailments utilizing the video internet of things (VIoT). This technique needs to be automated to save time and minimize counting errors. This paper aims to apply recent image processing techniques to identify and categorize white blood cells (WBCs) found in peripheral blood and to develop a system that would automatically detect and evaluate WBCs. The proposed method consists of steps such as segmentation scanning, feature extraction, and blood cell categorization. First, we segment the cell pictures, which entails grouping white blood cells into clusters. The second stage consists of scanning each input image and preparing the dataset. The third stage involves determining the texture and contour of a scanned picture. Finally, different machine learning algorithms are used to classify the results based on these criteria.

1. Introduction

White blood cells, also known as leukocytes, play a vital role in keeping the body free of infection. Although these cells are colorless, we can apply special stains on the blood to colorize them and make them visible under the microscope [1]. The white blood cells can flow by poking out one end of their body and pulling the remainder of their body along. They are the blood's "soldiers," battling germs and other foreign invaders [2]. White blood cells can squeeze through microscopic blood capillaries, leaving the bloodstream to enter other tissues under attack by foreign invaders. The majority of white blood cells are produced in the red marrow of bones. Some are also produced in other parts of the body by specific glands. Every cubic inch of blood contains between 4 and 11 thousand leukocytes in a healthy person. An infection causes the marrow and specific glands in the body to produce extra white blood cells. Medical technicians are trained to look for and count white blood cells in blood samples to determine whether someone has an illness [3–5]. The five types of white blood cells are all present in white blood cells. White blood cells (leukocytes) include neutrophils, eosinophils, basophils, lymphocytes, and monocytes [6].

This paper presented the automated methods for classifying the input blood cell images from the Video Internet of Things (VIoT). We have designed the image processing and machine learning algorithms for that purpose in this paper. The experimental results are validated using the publicly available research images dataset. Section 2 presents the concepts of image processing. Section 3 presents the machine learning techniques. Section 4 presents the experimental results and discussions. Section 5 presents the conclusion and

* Corresponding author.

E-mail addresses: roojwan.hawezi@epu.edu.iq (R.S. Hawezi), farah.xoshiabi@epu.edu.iq (F.S. Khoshaba), shahab.kareem@epu.edu.iq (S.W. Kareem).

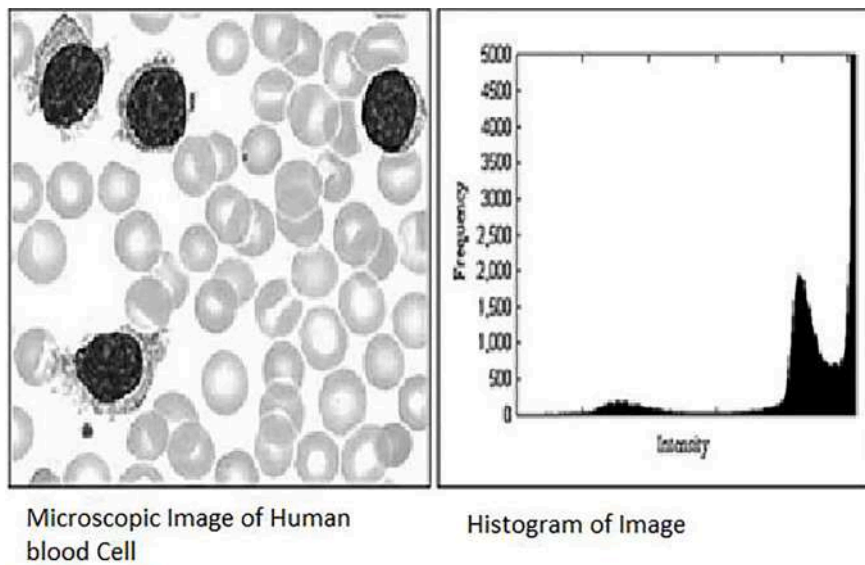


Fig. 1. Example of human blood cell image.

suggestions.

2. Image processing

As direct digital imaging technology has become more common in medical diagnosis, digital image processing is becoming increasingly important in healthcare. Real-world data is frequently incomplete and noisy, and it is also likely to be over or under filled, redundant, or erroneous [7]. Furthermore, typical conventional techniques, like linear regression, are especially prone to these predictors. As a result, before running the model, data analysis, including pre-processing, is necessary [8]. While data preparation frequently refers to the transformation of characteristics into another collection to employ an appropriate data mining or machine learning technology. Preprocessing seldom occurs throughout the data collection procedure [9,10]. Classifier development is a popular area of study in data mining and machine learning. Thousands of submissions have been received; nonetheless, the quality of the learned models is governed by the essence of the training data. Poor models will arise if the training data is untrustworthy, regardless of whether the classifier inducer is utilized [11]. Image visualization refers to all types of image alterations to this matrix that results in ideal image generation. Image analysis, in terms of processing, extends much beyond simple quantitative measures, and it is used for both tangible and abstract interpretations of biological pictures. One must be aware of the content and type of images, and this information must be included in the algorithms at a high level of abstraction. As a result, image analysis is a relatively narrow field, and improved approaches can easily be adapted to other application areas. Picture management involves all methods for efficiently storing, communicating, archiving, transferring, and retrieving picture data. As a result, telemedicine techniques have been included in image administration [12]. Images can be studied on a high or low level, and image processing is referred to as high-level image processing when the content of the image is known ahead of time. Fig. 1 shows a microscopic view of a human blood cell with four WBCs and several RBCs.

The aim of blood cell image transmission is generally to enhance the present image to reduce noise. The elimination of the image's backdrop, which contains red blood cells, platelets, and other objects, is required for cell segmentation. White blood cells, which are the objects of power, grow as a result of the segmentation process [13]. After proper segmentation, the whole white blood cell, including the nucleus and cytoplasm, should be retrieved. The form, texture, size, and content ratio of the nucleus in the cell are all factors to consider when identifying a cell. Image segmentation is the most critical stage of image processing and a vital technology, and it has a significant influence on future processing. Image segmentation has advanced significantly in recent years, and various novel segmenting methods have been devised recently. The most powerful algorithms have defects of their own. Given the universe's complexity, segmenting and counting cell images is a challenging undertaking. There are several Morphological Processing (MP) statements in various disciplines of biological image processing. Noise reduction, smoothing, and other methods of filtering, classification, segmentation, and pattern recognition are applied to both binary and grayscale images.

3. Classification algorithms

Machine learning (ML) is a collection of forecasting algorithms. These machine learning algorithms are not only simpler to employ, but they also outperform classic mathematical approaches [14]. Classification is a machine learning problem that entails picking scores from a large variety of disparate data. It is one of the classification methods that predicts based on a given input. A training set, which

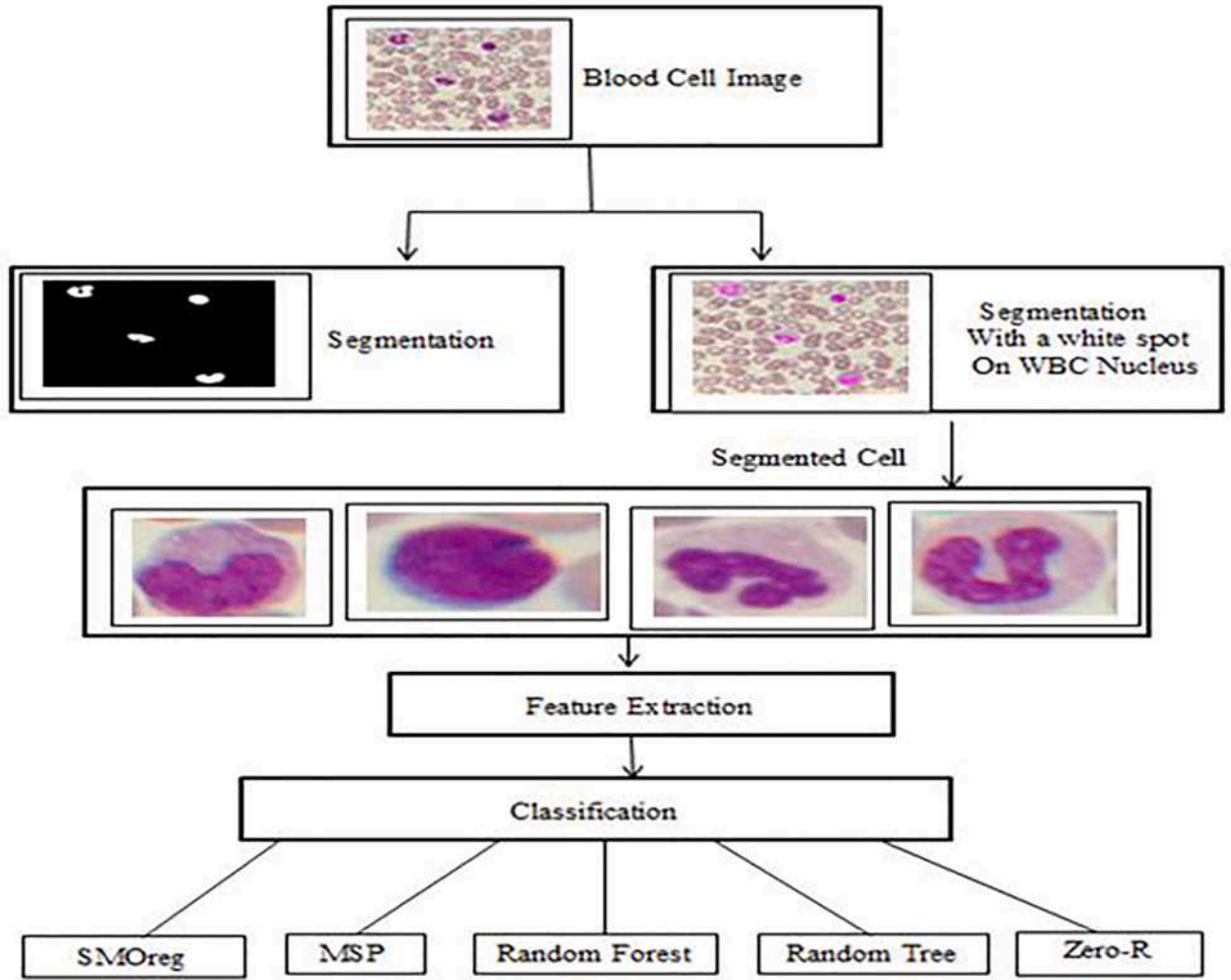


Fig. 2. Architecture of proposed blood cell classification system.

comprises attributes and a specific outcome, is used in a technique for predicting results. This collection is frequently a specific object or some other predicted characteristic. Classification is accomplished by recognizing various categories in an image and allocating unique pixels within the picture to the categories.

Classification algorithms are classified into two types: unsupervised and supervised. Unless data for analyzing a dataset or picture is available, supervised classification use phantom markers derived from training samples. Unsupervised classification determines spectral classifications in a multiband image without the assistance of an analyst. The Image Classification algorithms give cluster-building technology, the capacity to analyze the quality of the clusters, and the ability to transfer the clusters to classification algorithms. Throughout our investigations, we have enhanced a variety of computation methodologies. Categorizing a set of data into distinct classes demands a detailed grasp of the classifications' connection to the data. Cell categorization is particularly vital in the medical context [15]. Five distinct machine learning algorithms are studied in this study. The mathematical model behind these methods is depicted in the next section.

4. Algorithms and results analysis

4.1. Methodology

Pre-processing is a stage that involves improving an image from one that has been gathered. It also entails putting the image into a sequence for producing the primary image and fitting it to the next estimate [16–22]. The proposed approach consists of three steps in this paper: (1) Segmentation and Scanning WBCs from a blood smear image, (2) Feature extraction to obtain the database, and (3) Classification of blood cells into one of five classes (Fig. 2). In the following, we explain each of these steps briefly.

A critical white blood cell segmentation algorithm is necessary for an autonomous white blood cell classification system to function effectively. The following is found in a good segmentation:

- Pixels belonging to the same class have similar multimodal grayscale values, forming a linked region.
- The values of adjacent pixels belonging to distinct groups differ. Working on objects (containing several pixels) in the image rather than single pixels is believed to be the most important stage in analyzing WBC images.

Feature extraction has morphological operations, where it does the role of extracting features out of WBCs that contain vital information. The nucleus region, as well as the entire cell, is included in the form feature. Homogeneity, contrast, and entropy are texture features that are extracted from the segmentation images. For classification, we have used five classifiers such as Support Vector Machin with Regression (SMOreg), Model tree classifier (M5P), random forest, random tree, and Zero-R.

4.2. Experimental evaluation

The only way to learn about this new study is to see and evaluate the methodology that went into making it possible (mean absolute error, correlation coefficient, root mean square deviation, root relative squared error relative absolute error). The correlation coefficient is the first of the criteria to examine; the relationship between two variables is described by the correlation coefficient, which presents the strength and the orientation of the correlation. It is a value that can be either negative or positive. When the absolute value is close to 1, the link is stronger. The correlation coefficient is null because there is no linear relationship between the variables. The mean absolute error (MAE) metric involves the calculation of the difference between the actual (Z_t) and the fitted (Z'_t) data points and dividing this difference by the square root of the number of data points. If the cost of forecast errors is proportional to the absolute size of the forecast errors, MAE is the best calculation.

$$MAE = \frac{1}{N} \sum_t^N |Z'_t - Z_t| = \frac{1}{N} \sum_t^N |e_t| \quad (1)$$

To start, assume that the forecasts of models 1 and 2 are equal to $e1_t$, $e2_t$ where $t = 1, 2, \dots, m$. On the date axis, loss for model 1 is calculated as: $|e1_t| - |e2_t|$, whereas for model 2, loss is calculated as: $date = 1 - e1_t - e2_t - e2_t$. MAE is used as a measure of prediction loss [23,24]. An important measure in the modeling field is the root mean square deviation (RMSD) also known as the Root Mean Square Error (RMSE). The RMSD is used to combine all of the individual differences (also known as residuals) into a single predictive power score. In computer and statistical contexts, the square root of the mean squared error (also known as the RMSD of a model prediction for the estimated variable X) is referred to as the root mean square error (RMSDE):

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{mo del,i})^2}{n}} \quad (2)$$

The authors represent observed data and the model represents predicted values at time/location, where X_{obs} represents that data. RMSE values will have units, which prevents direct comparison of RMSD values for phosphorus concentrations with RMSE values for chlorophyll-a concentrations. On the other hand, the RMSD values can be used to compare model performance between models and as a way to compare model performance while under test to model performance in a calibration period. The impact of each error on RMSD is proportional to the squared error size, so greater errors have a disproportionately significant impact on RMSD. As a result, RMSD is susceptible to outliers [24].

Table 1
Evaluation of WBC images classification.

	SMOreg	M5P	Random Forest	Random Tree	Zero-R
Correlation coefficient	0.0357	0.315	0.433	0.1141	-0.4
Mean absolute error	1.5311	1.176	1.0968	1.26	1.2329
Root mean squared error	1.777	1.3887	1.2847	1.8493	1.4405
Relative absolute error	124.19%	95.39%	88.96%	102.20%	100%
Root relative squared error	123.36%	96.40%	89.19%	128.38%	100%

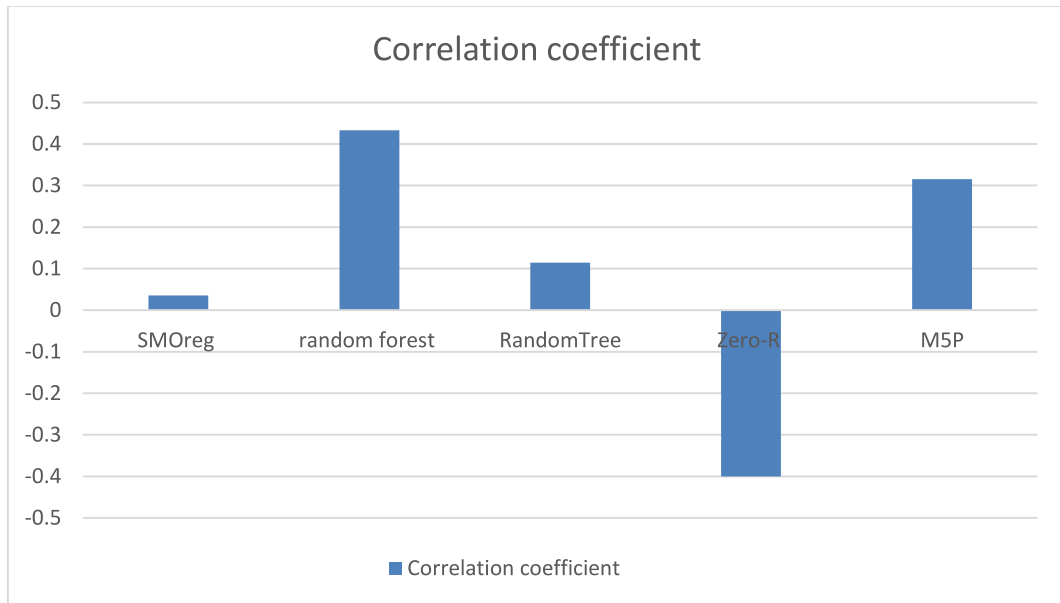


Fig. 3. Correlation Coefficient analysis.

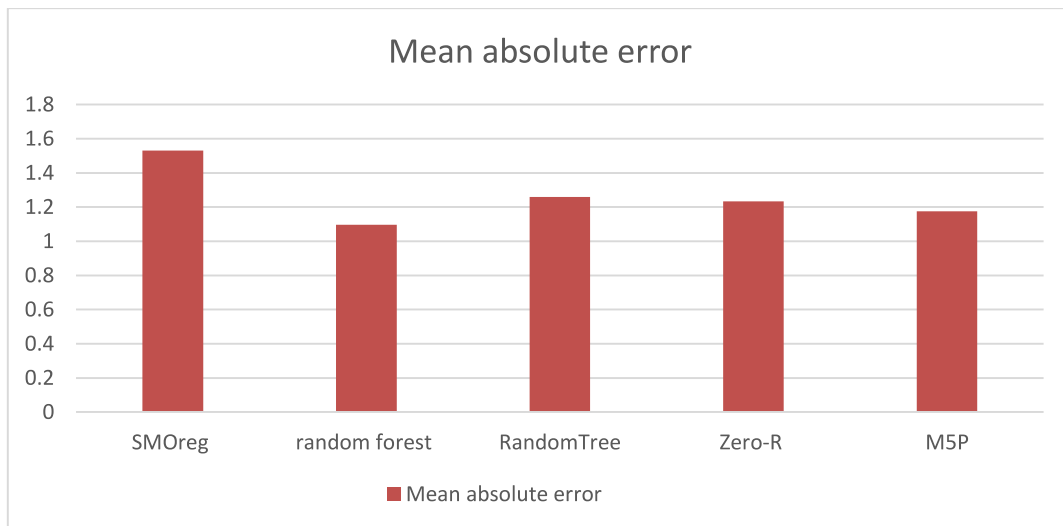


Fig. 4. Mean Absolute error analysis.

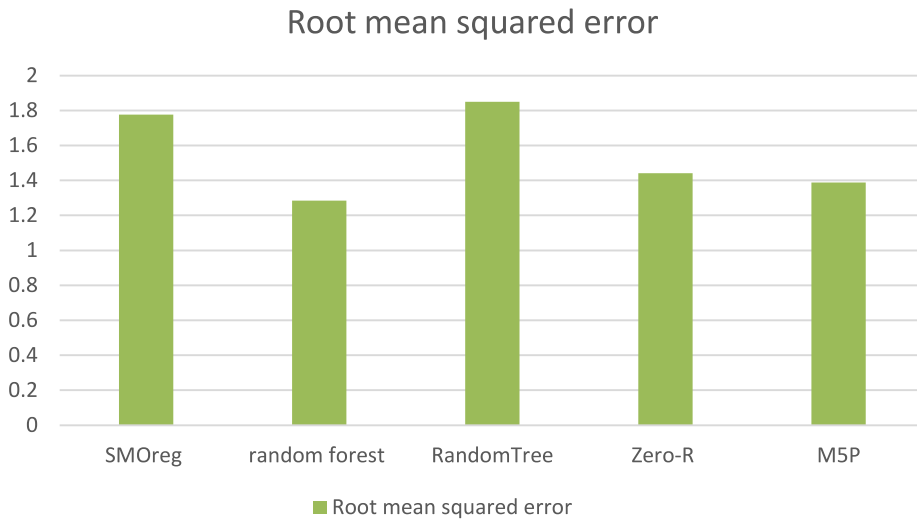


Fig. 5. Root mean square error analysis.

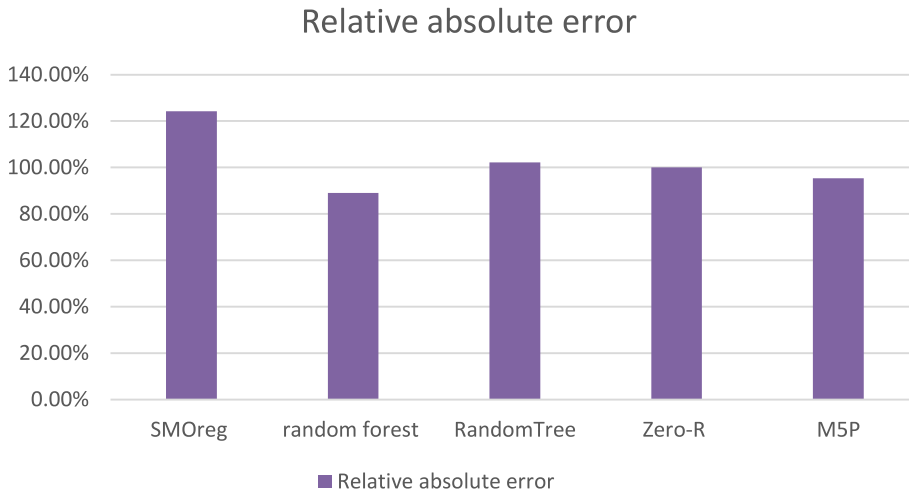


Fig. 6. Relative absolute error analysis.

The absolute deviation acquired from the prediction model is known as relative absolute error (RSE). The accuracy of a prediction is inversely proportional to it. Lower RSE means better accuracy:

$$RSE = \frac{\sum_{i=1}^n |f_i - y_i|}{\sum_{i=1}^n |f'_i - y_i|} \quad (3)$$

To calculate root relative squared error (RRSE), simply multiply root square by root and square root, then square the result:

$$RRSE = \frac{\sum_{i=1}^n |f_i - y_i|^2}{\sum_{i=1}^n |f'_i - y_i|^2} \quad (4)$$

A negatively associated relationship exists between RRSE and prediction accuracy. It is possible to improve the forecast accuracy by reducing the RRSE [25]. The microscopic images used in this paper are five types of WBC images. The image for WBC was taken from the central public health Laboratory in Duhok. Images are captured from smear slides by a Nikon 50i microscope, equipped with a Nikon color camera DP5M.

After applying the mentioned classification algorithm in section III, we present the result of different criteria as shown in Table 1. The output of the classification algorithms is classified into five different models (Basophil, Eosinophil, Lymphocyte, Monocyte, and Neutrophil) the result of the classification is shown in the Figures below. Depended on the Correlation coefficient shown in Fig. 3, the Mean Absolute Error shown in Fig. 4, the Root mean square error in Fig. 5, the Relative absolute error in Fig. 6, and the Root relative squared error in Fig. 7 the K* classifier is better than others mention algorithms and the second better algorithms are Decision table,

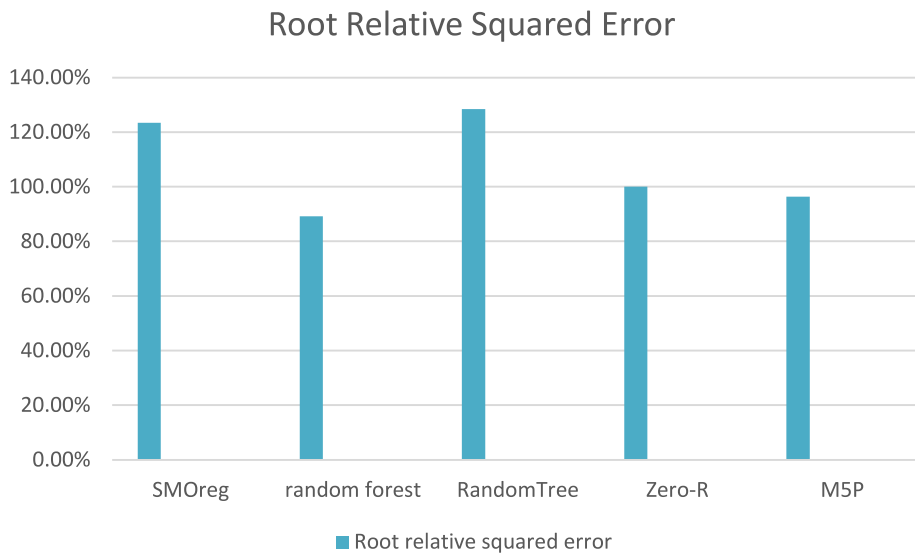


Fig. 7. Root relative squared error analysis.

then Additive regressive, Bagging and input mapped classifier.

5. Conclusion

Blood is made up of three types of cells: platelets, white blood cells, and red blood cells. Red blood cells transfer oxygen from the heart to all tissues and transport carbon dioxide. They can account for up to 50% of total blood volume. WBCs play a crucial role in the immune system since they are the body's main line of defense against infections and diseases. As a result, the accurate categorization of WBCs is critical and in high demand. WBCs are classified into two categories based on the appearance of the cytoplasm. Granulocytes, which comprise Basophils, Eosinophils, and Neutrophils, are the first kind. Lymphocytes and monocytes are included in the second category, known as Agranulocytes. Millions of individuals are impacted by leukemia, which is a type of cancer. We proposed a process for specific identification of WBCs using image classification, improvement, and segmentation algorithms. Different methods exist which can be utilized to perform specific classification steps. We used (Root-mean-square deviation, mean absolute error, Correlation coefficient, Relative absolute error, and Root relative squared error) classification methods for classifying the WBC images into five groups (Lymphocyte, Basophil, Eosinophil, Monocyte, and Neutrophil). The result showed that Random Forest is the best of the above algorithms for the input dataset. Additionally, it also utilizes morphological operations to increase the image corrode and also to improve as far as the potential of the primary image of nuclei. Certainly, intensity, shape, texture, and color were also employed while identifying features.

Declaration of Competing Interest

None.

References

- [1] Sharif Muhammad, Amin Javaria, Siddiqa Ayesha, Khan Habib Ullah, Malik Muhammad Sheraz Arshad, Anjum Muhammad Almas, et al. Recognition of Different Types of Leukocytes using YOLOv2 and Optimized Bag-of-Features. *IEEE Access*; 2020. p. 167448–59. September 2020.
- [2] Cömert Zafer, Budak Ümit, Akbulut Yaman, Şengür Abdulkadir. White Blood Cell Classification Based on Shape and Deep Features. In: *International Artificial Intelligence and Data Processing Symposium (IDAP)*; 2019.
- [3] Tareef A, Song Y, Feng D, Chen M, Cai W. Automated multi-stage segmentation of white blood cells via optimizing color processing. In: *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*; 2017. p. 565–8.
- [4] Ratley Astha, Minj Jasmine, Patre Pooja. Leukemia Disease Detection and Classification Using Machine Learning Approaches: A Review. In: *First International Conference on Power Control and Computing Technologies (ICPC2T)*; 2020.
- [5] Su M, Cheng C, Wang P. A Neural-Network-Based Approach to White Blood Cell Classification. *The Scientific World Journal* 2014.
- [6] Ismael Sami H, Kareem Shahab Wahhab, Almkhtar Firas H. Medical image classification using different machine learning algorithms. *AL-Rafidain J Comput Sci Math* 2020:135–47.
- [7] Hegde Roopa B, Prasad Keerthana, Hebbar Harishchandra, Singh Brij Mohan. Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. *Biocybern Biomed Eng* 2019:382–92.
- [8] Sholeh FI. White blood cell segmentation for fresh blood smear images. In: *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*; 2013. p. 425–9.
- [9] Wibawa MS. A Comparison Study Between Deep Learning and Conventional Machine Learning on White Blood Cells Classification. In: *International Conference on Orange Technologies (ICOT)*; 2018. p. 1–6.
- [10] Zhang D, Zhang P, Wang L. Cell Counting Algorithm Based on YOLOv3 and Image Density Estimation. In: *IEEE 4th International Conference on Signal and Image Processing (ICSIP)*; 2019. p. 920–4.

- [11] Ruberto Cecilia Di, Putzu Lorenzo. White Blood Cells Identification and Classification from Leukemic Blood Image. In: International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2013); 2013.
- [12] Sundara SM, Aarthi R. Segmentation and Evaluation of White Blood Cells using Segmentation Algorithms. In: 3rd International Conference on Trends in Electronics and Informatics (ICOEI); 2019. p. 1143–6.
- [13] Saade P, Jammal RE, Hayek SE, Zeid JA, Falou O, Azar D, et al. Computer-aided Detection of White Blood Cells Using Geometric Features and Color. In: 9th Cairo International Biomedical Engineering Conference (CIBEC); 2018. p. 142–5.
- [14] Rezatofighi SH, Soltanian-Zadeh H, Sharifian R, Zoroofi RA. A New Approach to White Blood Cell Nucleus Segmentation Based on Gram-Schmidt Orthogonalization. In: International Conference on Digital Image Processing; 2009. p. 107–11.
- [15] Salih Mohammed A, Wahhab Kareem S, azzawi Ahmed khazal al, Sivaram M. Time series prediction using SRE-NAR and SRE-ADALINE. J Adv Res Dyn Control Syst 2018.
- [16] Huang P, Wang Jing, Zhang Jian, Shen Yajuan, Liu Cong, Song Weiqing, et al. Attention-aware residual network based manifold learning for white blood cells classification. IEEE J Biomed Health Informatics 2020.
- [17] Yampri P, Pintavirooj C, Daochai S, Teartulakarn S. White Blood Cell Classification based on the Combination of Eigen Cell and Parametric Feature Detection. In: 1ST IEEE Conference on Industrial Electronics and Applications; 2006.
- [18] Throngnumchai Kan, Lomvisai Pitchayakorn, Tantasirin Chayanan, Phasukkit Pattarapong. Classification of White blood cell using Deep Convolutional Neural Network. In: The 2019 Biomedical Engineering International Conference (BMEiCON-2019); 2019.
- [19] Rajendran Sonu, Kumar DrESuresh. Leukocytes Classification and Segmentation in Microscopic Blood Smear Image. In: 2nd International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICT); 2019.
- [20] Raphael Riya T, Joy K R. Segmentation and Classification Techniques of Leukemia Using Image Processing: an Overview. In: International Conference on Intelligent Sustainable Systems (ICISS); 2019.
- [21] Sahlol AT, Kollmannsberger P, Ewees AA. Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. Scient Rep 2020.
- [22] Şengür A, Akbulut Y, Budak Ü, Cömert Z. White Blood Cell Classification Based on Shape and Deep Features. In: International Artificial Intelligence and Data Processing Symposium (IDAP); 2019. p. 1–4.
- [23] Kantor Joshua M, Richmond Christ D, Bliss Daniel W, Correll Bill. Mean-Squared-Error Prediction for Bayesian Direction-of-Arrival Estimation. IEEE Trans Signal Process 2013;61(19).
- [24] Sharma M, Bhave A, Janghel RR. White Blood Cell Classification Using Convolutional Neural Network. In: Wang J, Reddy G, Prasad V, Reddy V, editors. Soft Computing and Signal Processing. Advances in Intelligent Systems and Computing. Singapore: Springer; 2019. vol 900.
- [25] Al-Jumur Sardar MRK, Wahhab Kareem Shahab, z.yousif Raghad. Predicting temperature of erbil city applying deep learning and neural network. Indonesian J Electr Eng Comput Sci 2021:944–52.



Dr. Roojwan Scddeek Hawezi: I got the BSc degree from Baghdad University in Physical science, then I got a higher Diploma in Computer science. Back in 2003, I got my MSc degree then finally the Ph.D. in 2007 in Computer science from Baghdad too. I'm a specialist in Digital Image Processing. I'm the former head of the Information System Eng. (ISE) Department <https://scholar.google.com/citations?user=oAVYtCYAAAAJ&hl=en> <https://publons.com/researcher/5060847/roojwan-hawezi/> <https://www.scopus.com/home.uri> <https://orcid.org/0000-0002-1597-2895>.



Farah Sami khoshaba: She got the B.Sc. degree from University of Technology / Baghdad in the science of Computer and Control engineering in 1999, and worked as an engineer at the same department, as she was in charge of control laboratory and process laboratory. Then in 2006, she got the M.Sc. degree at the same department in the specialization of Mechatronics. She's a member of academic staff with a grade of lecturer at the Information System Engineering (ISE) Department (2011-till now). She teaches two different courses for undergraduate students and a course for diploma students in the same department. Her research interests include Control, Robotics, Computer Control, Path Planning and Machine Learning. You can contact her at email: (farah.xoshibi@epu.edu.iq) https://scholar.google.com/citations?hl=en&user=_Ei8uI4AAAAJ <https://publons.com/researcher/5103169/farah-khoshaba/> <https://orcid.org/0000-0003-2007-2207>.



Dr. Shahab Wahhab Kareem: Shahab Wahhab Kareem his BSc in Control and Computer Engineering from the University of Technology Baghdad in 2001, MSc in Software Engineering from Salahadeen University in 2009, and Ph.D. in Yasar University Izmir, Turkey in 2020. My research interests include Machine learning and BIG DATA. I'm a lecturer at the Information System Eng. (ISE) Department (2011-till now). He can be contacted at email: shahab.kareem@epu.edu.iq or shahabwk@yahoo.com. <https://scholar.google.com/citations?user=XOTpQSQAAAAJ&hl=en> <https://www.scopus.com/authid/detail.uri?authorId=57205541046> <https://publons.com/dashboard/summary/> <https://orcid.org/0000-0002-7362-4653>.