Trends in Health Informatics

www.thi.reapress.com

THI. Vol. 3, No. 1 (2026) 74–87.

Paper Type: Original Article

White Blood Cell Segmentation via Invariant Blood Image Characteristics and Marker-Controlled Watershed

Lorenzo Cevallos-Torres*

Brigham Young University–Idaho, USA; cev18002@byui.edu.

Citation:

Received: 26 April 2025 *Revised: 21 July 2025* *Accepted: 11 September 2025*	Cevallos-Torres, L. (2026). White blood cell segmentation via invariant blood image characteristics and marker-controlled watershed. *Trends in Health Informatics, 3*(1), 74-87.

Abstract

Accurate segmentation of White Blood Cells (WBCs) is considered one of the most critical steps in automated hematological disease diagnosis systems and blood cell classification. However, variations caused by imaging conditions, staining techniques, illumination intensity, and the presence of overlapping cells significantly degrade the accuracy and generalizability of many existing methods. In this study, a robust approach for WBC segmentation in microscopic blood smear images is proposed based on three invariant characteristics of peripheral blood images. In the proposed method, the background mask is initially extracted using the M channel of the CMYK color space combined with the Zack thresholding algorithm. Subsequently, WBC nuclei are identified, and Red Blood Cell (RBC) regions are determined by simultaneously exploiting background and nuclear information. Then, preliminary WBC regions are extracted using the color characteristics of RBCs. To separate clustered WBCs, a modified Marker-Controlled Watershed algorithm based on distance transform and a refined nuclear mask is employed to prevent over-segmentation. Finally, WBC-like artifacts are eliminated using geometric features. The performance of the proposed method is evaluated on three standard datasets with different imaging conditions and staining protocols. The results demonstrate that the proposed method can provide accurate and stable WBC segmentation independent of the dataset type, while exhibiting strong robustness against variations in imaging conditions and the presence of overlapping cells.

Keywords: White blood cell segmentation, Blood microscopic images, CMYK color space, Zack algorithm, Watershed algorithm, Image processing.

1 | Introduction

Automatic analysis of microscopic blood images has become one of the important research topics in the field of medical image processing and Computer-Aided Diagnosis (CAD) systems over the past two decades.

Corresponding Author: cev18002@byui.edu

https://doi.org/10.22105/thi.v3i1.51

Differential White Blood Cell Count (DWBCC) is one of the most common hematological tests that provides valuable information regarding the immune system status, various infections, inflammatory diseases, autoimmune disorders, and hematological malignancies such as leukemia. However, performing this test manually is not only time-consuming but also highly dependent on the expertise of specialists and may be affected by fatigue, inter-observer variability, and human errors. Therefore, the development of intelligent systems for the automated analysis of blood smear images has emerged as an active research area in medical image processing [1–3]. In most automated hematological disease diagnosis systems, White Blood Cell (WBC) segmentation is considered the first and one of the most critical processing stages, as the accuracy of subsequent steps, including feature extraction, cell classification, and disease diagnosis, directly depends on the quality of segmentation. Any error in identifying cell boundaries or incomplete extraction of the nucleus and cytoplasm regions may lead to inaccurate extracted features and consequently reduce the performance of classification models. Therefore, accurate WBC segmentation has been recognized by many researchers as one of the fundamental challenges in automated blood image analysis systems [2], [4]. Although the detection of WBCs in stained microscopic images may appear straightforward, this process is practically challenging due to several factors. Variations in staining protocols (such as Wright, Giemsa, and Wright–Giemsa), differences in imaging conditions, illumination variations, noise, color deviations caused by laboratory equipment, overlapping cells, the presence of platelets and artifacts, variations in WBC morphology and size, as well as differences among available datasets, cause many segmentation algorithms to perform effectively only under specific conditions and exhibit limited generalization capability when applied to new images [3], [5].

To overcome these challenges, various approaches have been proposed for WBC segmentation. The early generation of these methods was mainly based on intensity thresholding and color information. In these approaches, the color difference between the WBC nucleus and other image components is exploited to separate WBCs from the background and Red Blood Cells (RBCs). Although these methods are computationally simple and fast, their performance is highly dependent on threshold selection and they demonstrate limited robustness against illumination variations, staining differences, and image quality changes [6].

To improve segmentation accuracy, researchers have investigated the use of different color spaces, including HSV, Lab, YCbCr, and CMYK. The main idea behind these approaches is that certain color channels can enhance the contrast between WBC nuclei and other image components, thereby facilitating the segmentation process. However, selecting an appropriate color space remains dependent on dataset characteristics, staining protocols, and imaging conditions, and no single-color space has demonstrated optimal performance across all datasets [5], [7].

In recent years, with the advancement of machine learning techniques, particularly deep learning methods, numerous studies have been conducted for WBC segmentation and classification. Although these approaches have achieved promising performance on several benchmark datasets, recent systematic reviews indicate that challenges such as dependency on image quality, the requirement for large annotated datasets, limited generalization capability, and high computational costs remain unresolved [8], [9].

Alongside deep learning-based approaches, classical image processing techniques continue to be widely employed in blood image analysis systems due to their simplicity, interpretability, computational efficiency, and independence from training procedures. Among these techniques, thresholding methods, morphological operations, distance transform, region growing, and Marker-Controlled Watershed algorithms have been extensively utilized [2], [10]. These methods can accurately extract WBC nuclei and cell regions under controlled imaging conditions; however, their performance is strongly influenced by marker selection, staining uniformity, and the degree of cell overlap.

To reduce the over-segmentation problem, several modified versions of the Watershed algorithm have been proposed. For instance, Abrol et al. developed an improved Marker-Controlled Watershed approach to enhance leukocyte separation accuracy; however, the selection of initial markers remained dependent on

image characteristics [10]. Furthermore, several studies have demonstrated that the performance of Watershed-based algorithms significantly decreases in images with severe illumination variations, non-uniform staining, or clustered cells [2], [8].

In recent years, Convolutional Neural Networks (CNNs), U-Net architectures, Attention U-Net, Mask R-CNN, and other deep learning models have been extensively applied for WBC segmentation. These approaches achieve high accuracy on public datasets by automatically learning complex image features [9], [11]. Nevertheless, most of these models require large amounts of annotated data, considerable training time, powerful computational hardware, and careful hyperparameter optimization. Moreover, recent investigations have shown that the performance of many deep learning models decreases when evaluated on images acquired from different laboratories, imaging devices, or staining protocols, indicating that generalization remains a major challenge in this field [8], [9].

A review of previous studies reveals that most existing methods have focused on improving a specific stage of the segmentation process; for example, some studies emphasize accurate nucleus extraction, others focus on separating overlapping cells, and some aim to enhance classification performance [2], [8], [11]. In contrast, only a limited number of studies have attempted to develop an integrated and imaging-condition-independent framework that simultaneously exploits stable characteristics of blood smear images for background extraction, nucleus detection, RBC removal, adjacent cell separation, and false region elimination. This issue has been highlighted as an important research gap in recent reviews [8], [9]. Although numerous WBC segmentation methods have been introduced in recent years, achieving a robust and generalizable approach capable of providing stable and reliable performance under diverse imaging conditions and across different datasets remains an open challenge [3], [8], [9]. Many classical methods, including thresholding, morphological operations, clustering, and Watershed-based techniques, have been specifically optimized for particular datasets or imaging conditions and require parameter adjustment when staining type, illumination intensity, image quality, or cell overlap characteristics change [2], [4], [10]. On the other hand, although deep learning-based methods have reported high accuracy on training datasets, their dependence on large annotated datasets, high computational requirements, and reduced generalization ability when applied to images from different laboratories and imaging systems remain significant limitations [8], [9], [11].

Previous studies indicate that despite considerable progress, developing an accurate and stable segmentation method that maintains consistent performance under different imaging conditions and across diverse datasets remains a significant challenge. Many classical approaches focus on improving a specific segmentation stage, such as nucleus extraction or separation of adjacent cells, and require parameter recalibration when image quality or staining conditions vary [2], [5]. Conversely, although deep learning approaches have demonstrated high accuracy on certain datasets, their dependence on extensive training data, computational complexity, and limited transferability to images acquired from different laboratories restrict their practical applicability [6–8]. Furthermore, in many existing studies, different segmentation stages are performed independently, and the informative relationship between background extraction, nucleus detection, RBC removal, and adjacent cell separation is not fully exploited. Consequently, errors generated in early stages may propagate to subsequent steps and negatively affect the final segmentation accuracy.

In this study, a sequential framework for WBC segmentation in peripheral blood smear images is proposed, where the output of each stage is utilized as guiding information for the subsequent stage. In the proposed approach, the image background is initially extracted using the M channel of the CMYK color space and the Zack thresholding algorithm. Subsequently, WBC nuclei are detected, and based on the information obtained from these two stages, RBCs are separated from other image regions. Then, initial WBC regions are extracted, and an improved version of the Marker-Controlled Watershed algorithm based on distance transform is employed to separate adjacent cells. Finally, false regions are removed according to geometric characteristics, generating the final WBC segmentation mask.

The main contributions of this study are summarized as follows:

I. A sequential and integrated framework is introduced for WBC segmentation, in which the extracted information from each stage is utilized as guidance for the following stage.

II. The combination of background, nucleus, and RBC information is employed to improve the WBC segmentation process.

III. An improved Marker-Controlled Watershed algorithm is proposed to enhance the separation of clustered and adjacent WBCs.

IV. The proposed method is evaluated on multiple standard datasets to investigate its accuracy, robustness, and generalization capability.

2 | Proposed Method

Microscopic blood smear images are generally acquired as color images and contain four meaningful regions, including the WBC nucleus, WBC cytoplasm, RBCs, and background. Although the color characteristics of these regions may vary depending on imaging conditions and staining protocols, three common assumptions exist for microscopic blood smear images that remain invariant under different acquisition conditions and staining variations. These assumptions are described as follows:

Assumption 1: The darkest purple region in microscopic blood smear images corresponds to the nuclei of WBCs.

Assumption 2: The brightest region in these images represents the background area.

Assumption 3: No nucleus exists within the regions corresponding to RBCs in microscopic blood smear images.

Due to the invariance of these three assumptions against changes in imaging conditions and staining procedures, they are employed in this study to design a robust framework for WBC segmentation. In this context, robustness refers to the capability of the proposed segmentation method to accurately segment WBC regions across various images acquired from different laboratories with diverse staining protocols and imaging systems, regardless of the specific image characteristics. An algorithm can be considered robust when both its overall structure, the interaction between its different components, and the employed techniques in each processing stage maintain stable performance under varying conditions. In the proposed WBC segmentation algorithm, the aforementioned three assumptions are utilized to construct a robust framework. Furthermore, at each stage of the algorithm, reliable and robust methods available in the corresponding field are employed to enhance the stability and generalization capability of the segmentation process. Overall, the proposed segmentation algorithm consists of six main stages. *Fig. 1* illustrates the block diagram of the proposed framework, and each stage is described in detail in the following sections.

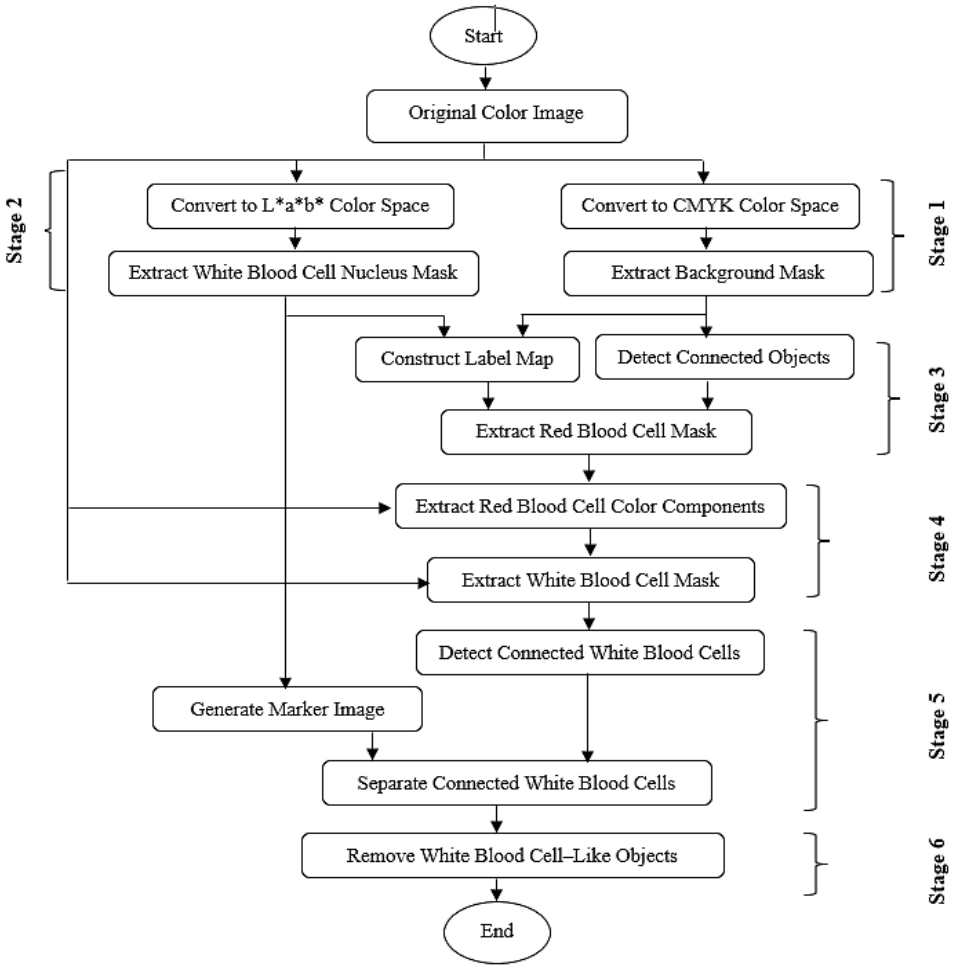

Fig. 1. Block diagram of the proposed segmentation algorithm.

2.1 | Background Mask Extraction

The objective of this stage is to separate image components from the background region. According to the second assumption, the visually brightest region of the image corresponds to the background. Therefore, background separation and mask extraction should also be performed in a robust manner, similar to the approach reported in [12].

Although the background color in many blood smear images is close to white, which facilitates background separation, its appearance may vary in some images due to staining procedures, imaging conditions, and noise. Furthermore, because of the types of dyes used in blood smear staining, components with purple color characteristics generally exist with a considerable proportion in most blood cells. The M channel of the CMYK color space effectively represents different purple components present in the image. Therefore, in this channel, blood cells can be represented with high contrast relative to the background.

Accordingly, the proposed method in this stage employs the M channel of the CMYK color space combined with the Zack thresholding algorithm [13] to achieve robust background extraction. *Fig. 2* illustrates examples of blood smear image histograms obtained from their grayscale representations.

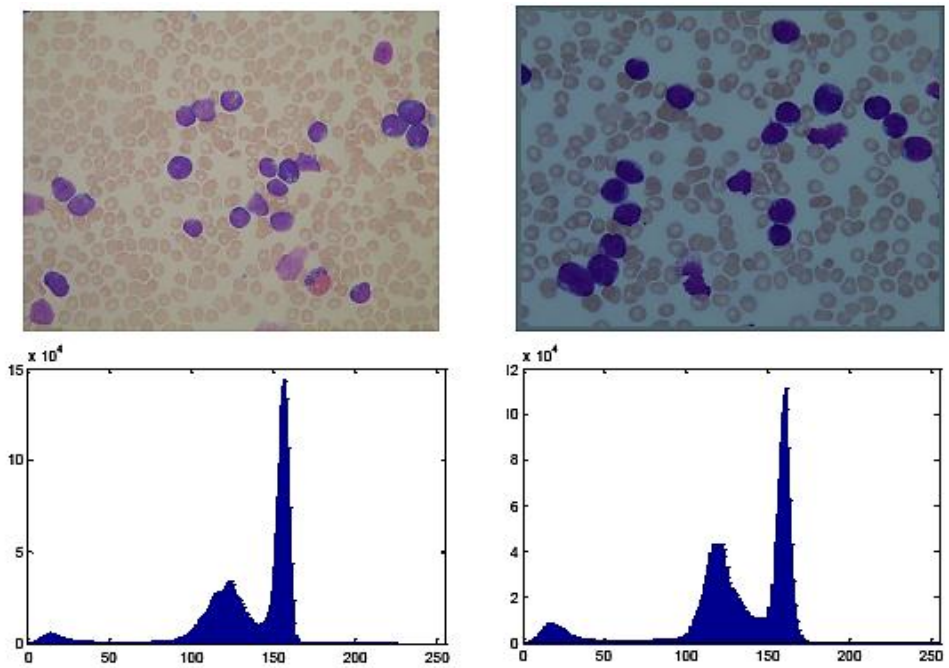

Fig. 2. Two examples of microscopic blood smear image histograms.

The histogram of blood smear images generally consists of three main regions, where each region has a relative maximum. In most cases, the region with the largest relative maximum corresponds to the background, the region with the smallest relative maximum represents the nuclei of WBCs, and the remaining region corresponds to RBCs and the cytoplasm of WBCs.

The optimal threshold value for separating the background from other image components is the threshold that can effectively distinguish the background region in the histogram from the other two regions. Depending on the characteristics of the blood smear image, imaging and staining conditions, and the number of cells present in the image, the distance between histogram regions, the relative maximum values, and the width of each histogram region may vary. When different histogram regions become very close or even overlap, selecting an appropriate threshold becomes a challenging task.

One effective strategy to address this problem is the utilization of different color spaces. In fact, color spaces generate new color channels by applying linear or nonlinear transformations to the RGB components of an image. The histograms obtained from these new channels may differ from the grayscale histogram in terms of the distance between regions, relative maximum values, and region widths. Therefore, by selecting an appropriate color space, the separation between histogram regions can be maximized, enabling more accurate threshold selection.

After investigating different color spaces and performing simulation analyses, the M channel of the CMYK color space was found to provide a more suitable histogram for applying the Zack thresholding algorithm. The details of this stage are described step-by-step below.

Step 1. The original RGB image is converted into the CMYK color space, and the M channel is selected as the input image for the next processing stage.

Step 2. The histogram of the image is generated. The histogram is a bar graph in which the horizontal axis represents image intensity values, and the height of each bar indicates the number of pixels corresponding to each intensity level.

Step 3. The locations of the relative maxima in the histogram are identified.

Step 4. The maximum and minimum values among the detected relative maxima are determined, and a straight line is drawn between these two points.

Step 5. The distance between the obtained line and all histogram points is calculated. The histogram point with the maximum distance from this line is selected as the threshold value.

All the above steps are illustrated in *Fig. 3*.

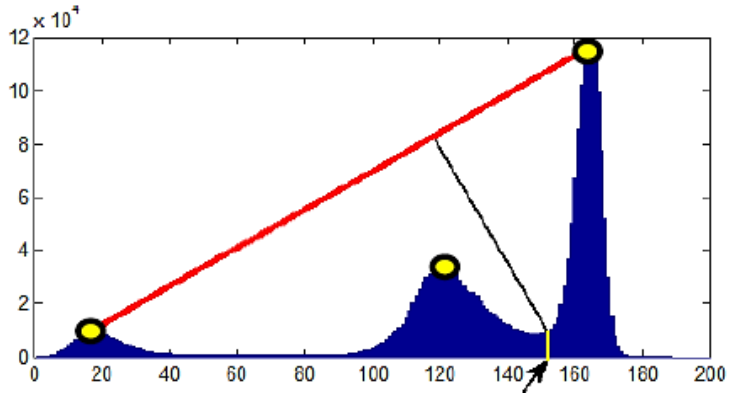

Fig. 3. Zack algorithm for threshold determination [13].

Step 6. The background regions are extracted by binarizing the image using the obtained threshold value.

2.2 | White Blood Cell Nucleus Segmentation

The objective of this stage is to identify the regions corresponding to WBC nuclei based on the first assumption. This stage is one of the most important steps in the segmentation process, and any error at this stage can significantly affect the final segmentation results. Therefore, selecting a robust approach is of great importance. In this study, the method reported in reference [12] is employed for WBC nucleus segmentation.

2.3 | Extraction of Red Blood Cell Regions

The objective of this stage is to determine the regions corresponding to RBCs. According to the third assumption, no nucleus exists within RBC regions. After completing the two previous stages, both the nucleus regions and the cellular components separated from the background are identified. Therefore, by exploiting the information obtained from these two stages, most RBC regions can be easily determined.

However, since blood cells in microscopic smear images are not always spatially separated and different types of blood cells are frequently attached or overlapped, all RBC regions cannot be completely identified at this stage. The proposed RBC detection procedure is described step-by-step as follows.

Step 1. Using the binary background mask obtained in the first stage and the binary nucleus mask obtained in the second stage, the blood smear image can be divided into three regions. Assuming that nucleus regions are represented by value 1 and other pixels by value 0 in the nucleus mask, while cellular components are represented by value 1 and the background by value 0 in the background mask, a label map can be generated according to the following rules:

I. Regions where both the background mask and the nucleus mask are equal to zero are considered background regions.

II. Regions where both the background mask and the nucleus mask are equal to one are considered nucleus regions.

III. The remaining regions are labeled as RBC and WBC cytoplasm regions.

Step 2. Connected components are extracted using the background mask. Then, by using the generated label map, connected components that do not contain any nucleus region are identified as RBC regions.

2.4 | Extraction of Initial White Blood Cell Regions

The objective of this stage is to extract WBC regions using the color information of RBCs. In the previous stage, all connected components containing only RBCs were identified based on the third assumption. Therefore, the remaining connected components represent candidate regions that most likely contain WBCs. However, these candidate regions do not exclusively consist of WBCs, and RBCs may also exist within these regions. Therefore, RBC regions can be effectively removed by exploiting their color characteristics. The details of the proposed approach are described below.

Step 1. Let $V = \{V_1, V_2, \dots, V_M\}$ represent the set of color vectors belonging to RBCs, and let $W = \{W_1, W_2, \dots, W_n\}$ represent the set of color vectors associated with connected components in the image. The set of color vectors corresponding to WBCs is obtained from the vectors that exist in set W but do not belong to set V.

Step 2. The candidate connected components of WBC regions are scanned, and regions whose color vectors belong to the set obtained in the previous step are identified and segmented as WBCs.

2.5 | Separation of Overlapping White Blood Cells

The objective of this stage is to identify and separate clustered or overlapping WBCs in the image. In the previous stage, when multiple WBCs are attached to each other, their regions appear as a single connected component, which cannot be effectively used for feature extraction. Consequently, extracting features from such regions does not provide meaningful information.

In many existing WBC segmentation and classification methods, this stage is ignored [2], [14], [15] and images without overlapping WBCs are selected to avoid the need for cell separation. Although selecting such images may lead to improved numerical results, it does not represent real clinical conditions. Therefore, overlapping cells should be detected and separated to enable the extraction of individual features from each cell for subsequent classification.

The Watershed algorithm is one of the powerful image segmentation techniques and is widely used for separating connected components in images. Since this algorithm performs segmentation based on the number and locations of local minima in the image, directly applying Watershed to grayscale blood smear images or their gradient images often produces unsatisfactory results. This is mainly due to the large number of local minima present in such images, resulting in excessive and inaccurate segmentation [10]. To overcome this limitation, a marker image is commonly employed. A marker image is a binary image in which selected pixels belonging to each object are highlighted. In fact, the marker image represents the approximate centers of connected objects and provides a simplified representation of their overall structure. However, selecting an appropriate marker image for images containing overlapping objects is not a straightforward task. The accuracy of the Watershed algorithm is highly dependent on the quality of the selected markers. One effective strategy for generating marker images, particularly for images containing connected and overlapping objects, is the use of the distance transform. This transform requires an initial binary image and calculates the distance of each pixel to the nearest background pixel. The resulting values are represented as a new grayscale image, where local maxima obtained from the distance transform are used to generate the marker image.

Therefore, selecting an appropriate initial binary image and generating a suitable marker image play a crucial role in achieving accurate separation of connected image components. After this stage, the marker image is combined with the original grayscale image. In this context, combination refers to modifying the original image such that the markers become local minima in the processed image. However, using the original image or its gradient often does not provide satisfactory results. Therefore, when the objective is to separate connected objects, the complement of the distance transform image is used instead of the original grayscale image for combination with the marker image.

In this stage, the marker image is carefully selected so that the separation boundaries generated by the Watershed algorithm can more accurately distinguish individual image components. In many Watershed-based approaches, the WBC mask is used for distance transformation and marker generation. However, when more than two WBCs are connected or when the shared boundary between two WBCs is large, the binary WBC mask does not generate an effective marker image.

In this study, instead of using the complete WBC mask, a modified WBC nucleus mask is employed for marker generation. This modification ultimately improves the separation of overlapping WBCs. The detailed procedure of this stage is presented in the following section.

Step 1. The circularity feature of each object identified as a WBC is calculated (the corresponding equation is provided).

Step 2. Objects with a circularity value lower than 0.8 are labeled as clustered WBCs according to reference [16].

Step 3. The distance transform is applied to the mask of clustered WBCs. By calculating the local maxima of the resulting distance transform image, the number of overlapping cells and their approximate centers are identified as small regions.

Step 4. The nucleus regions of clustered WBCs are examined to determine the number of markers within each nucleus region. If only one or fewer markers exist in a nucleus region, the nucleus region is replaced by its convex hull. Otherwise, the original nucleus region remains unchanged. This strategy prevents over-segmentation of WBCs with irregular-shaped nuclei.

Step 5. The distance transform is applied to the newly generated mask, and the marker image is obtained from the local maxima of the resulting distance transform image.

Step 6. The marker image is combined with the complement of the distance transform image obtained in *Step 3*, and the Watershed algorithm [17] is applied to the resulting image.

Step 7. The clustered WBCs are separated using the Watershed boundaries obtained from the previous step.

Fig. 4 illustrates the block diagram of the proposed algorithm for separating overlapping WBCs.

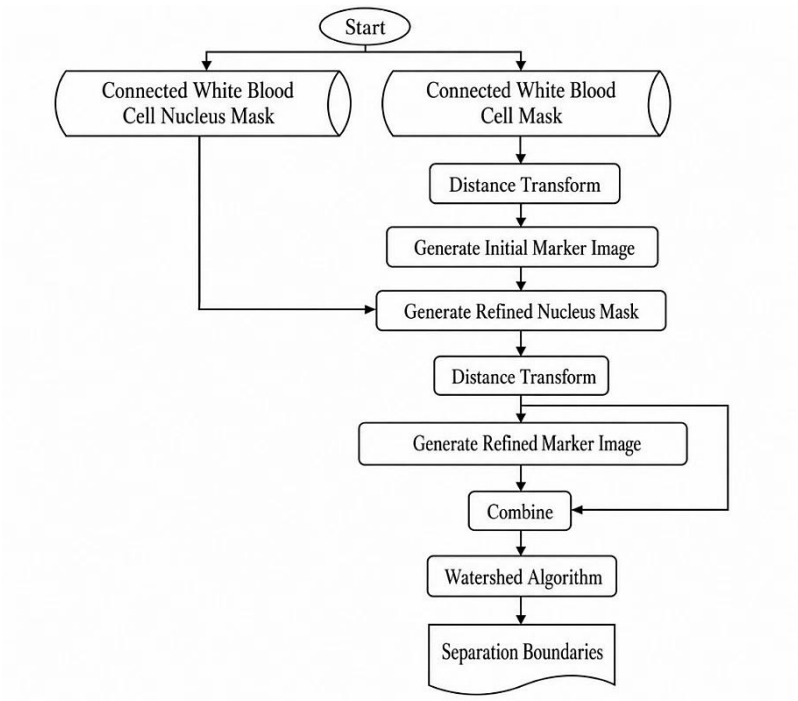

Fig. 4. Block diagram of the proposed algorithm for separating overlapping cells.

2.6 | Removal of White Blood Cell-Like Objects

The objective of this stage is to eliminate colored artifacts and objects that resemble WBCs in terms of color characteristics. *Fig. 5* illustrates examples of such colored artifacts, highlighted by yellow circles.

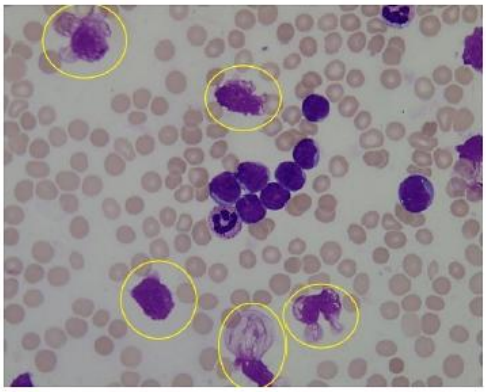

Fig. 5. Examples of colored artifacts resembling white blood cells.

The presence of these artifacts in microscopic blood smear images is inevitable and may increase errors in subsequent classification stages. Therefore, removing these regions is essential to reduce system errors and improve the overall performance.

One of the main characteristics of these WBC-like artifacts is their irregular shape compared with actual WBCs. In many cases, this single feature can be effectively utilized for their removal. Among the previously proposed methods, only reference [16] has specifically addressed these artifacts and demonstrated satisfactory performance in this regard. Therefore, this stage is performed based on the approach presented in [16], as described below.

Step 1. The eccentricity feature is calculated for all extracted objects in the image.

Step 2. According to the experimental results reported in [16], objects with an eccentricity value lower than 0.9 are considered colored artifacts and are removed from the image.

3 | Simulation Results

In this paper, images from three different datasets with varying characteristics were employed to evaluate the different stages of the proposed method. Details of each dataset are described below.

Dataset 1: This dataset consists of 400 images acquired from 100 peripheral blood smears. The blood smears were stained using the Giemsa-Wright method. The images were captured by a digital camera (SONY model No. SSC-DC50AP) mounted on an optical microscope (Microscope-Axioskope 40) equipped with a colorless lens at 100× magnification. All images are color images of size 720 × 576 pixels with 24 bits per pixel, stored in BMP format. The images were collected at the Hematology-Oncology Research Center of Imam Khomeini Hospital (Tehran, Iran) and classified into different types of WBCs by a hematologist. In addition, the nuclear and cytoplasmic regions of the WBCs were manually segmented by an expert. The images in this dataset exhibit significant variation in brightness intensity. This dataset is suitable for nucleus and cytoplasm segmentation of WBCs as well as for the classification of different types of leukocytes [15].

Dataset 2: The microscopic images in this dataset were acquired using a KODAK EasyShare V803 digital camera mounted on a Leica microscope at 100× magnification with an oil-immersion lens. The required blood smears were stained with Giemsa. The images are color images of size 3264 × 2448 pixels with 24 bits per pixel, stored in JPG format. These images were collected by the specialized Hematology Center at Dr. Juan Bruno Zayas Alfonso Hospital in Santiago, Cuba. Healthy and pathological RBCs were classified by a specialist in this dataset. This database can be used for blood smear image segmentation and RBC classification methods.

Dataset 3: The images in this dataset were captured using a CANON Power Shot G5 digital camera attached to a laboratory microscope with variable magnification (ranging from 300× to 500×). The images are color images with 24 bits per pixel, stored in JPG format. This dataset was prepared at the M. Tettamanti Research Center in Monza, Italy, for the detection of Acute Lymphoblastic Leukemia (ALL) using blood sample image processing techniques. It consists of two subsets: ALL-IDB1 and ALL-IDB2.

The first subset (ALL-IDB1) contains 108 images of size 2592 × 1949 pixels, encompassing a total of 3900 blood elements. Overall, there are 510 lymphoblasts across all images, whose center coordinates were determined by an expert. In each image, the presence or absence of at least one lymphoblast is also indicated. The number of images containing lymphoblasts is equal to the number of images without lymphoblasts. This subset can be used to evaluate the algorithm's performance in blood smear segmentation and separation of overlapping cells. Moreover, due to the availability of labeled lymphoblasts, it is suitable for assessing the classification power of the system in ALL detection at both image and cell levels.

Fig. 6 presents the practical implementation of the first six stages of the proposed WBC segmentation algorithm on a sample microscopic blood smear image.

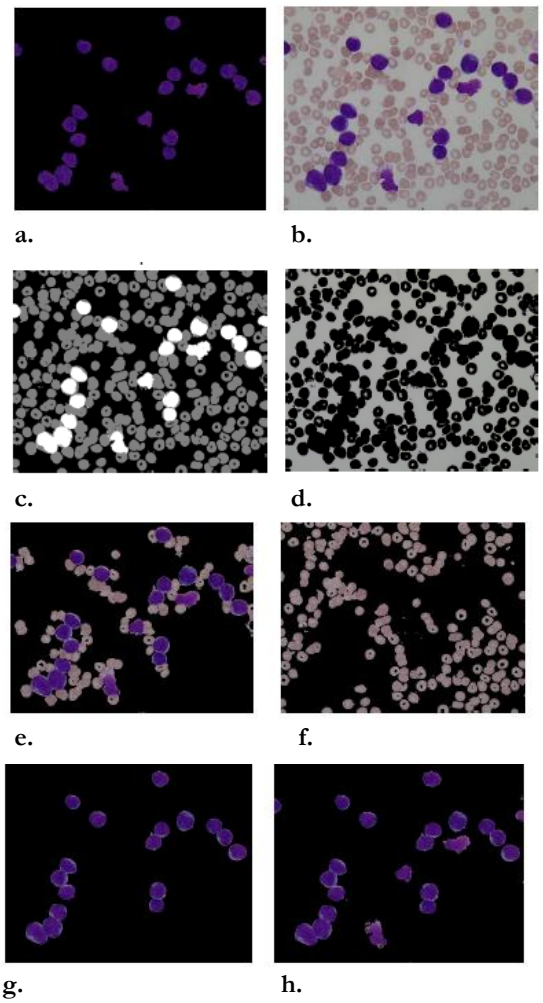

Fig. 6. Step-by-step implementation of the proposed WBC segmentation algorithm on a representative blood smear image; a. microscopic blood smear image, b. white blood cell nuclei, c. background regions, d. generated label map, e. red blood cell regions, f. candidate white blood cell regions, g. segmented white blood cell regions, and h. final white blood cell regions after removing WBC-like objects.

Figs. 7-9 illustrate the segmentation results obtained by the proposed algorithm on images from different databases.

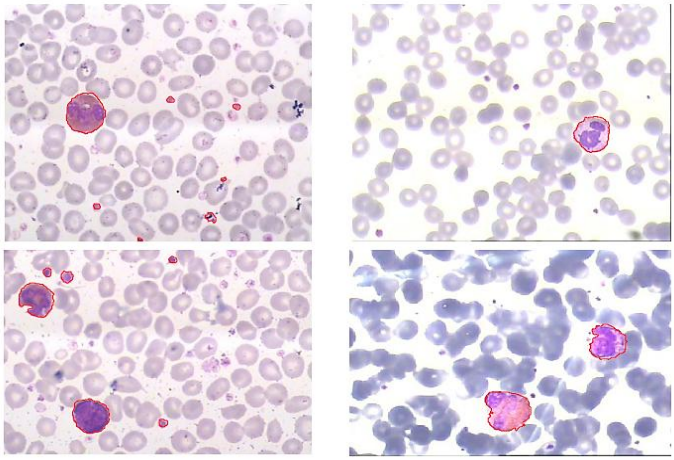

Fig. 7. Segmentation results of the proposed algorithm on four sample images from dataset 1.

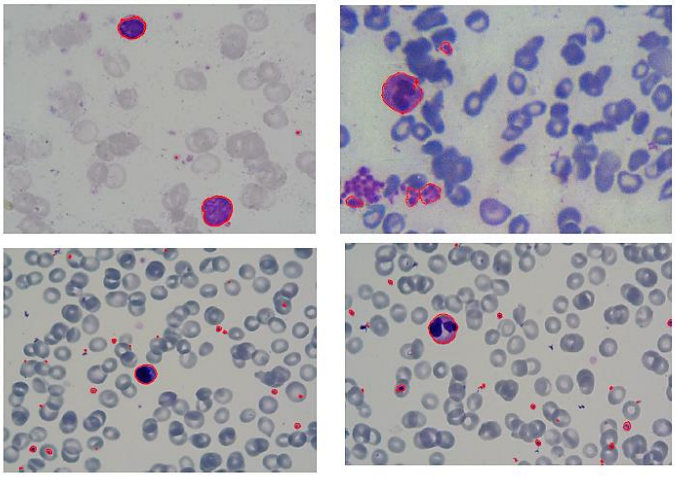

Fig. 8. Segmentation results of the proposed algorithm on four sample images from dataset 2.

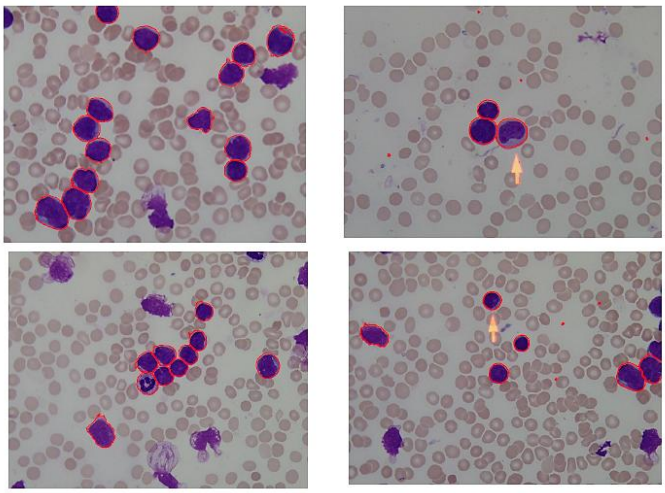

Fig. 9. Segmentation results of the proposed algorithm on four sample images from dataset 3.

As illustrated in *Figs. 7–9*, the proposed framework demonstrates stable segmentation performance across all three datasets, despite considerable variations in image acquisition conditions, staining protocols, image

quality, and imaging devices. In all test images, the initial extraction of the background and WBC nuclei provides reliable information for the subsequent processing stages. The sequential utilization of this information effectively reduces error propagation throughout the segmentation pipeline.

Furthermore, the experimental results indicate that exploiting RBC information before the final WBC extraction substantially eliminates irrelevant regions and improves the accuracy of candidate WBC detection. In addition, the modified Marker-Controlled Watershed algorithm based on the distance transform successfully separates clustered WBCs, even in challenging cases involving extensive cell overlap, while reducing the over-segmentation commonly observed in conventional Watershed-based approaches.

Overall, the results obtained from the three independent datasets demonstrate that the proposed framework maintains stable performance under different imaging and staining conditions without requiring dataset-specific parameter adjustment. These findings indicate that the proposed method possesses satisfactory robustness and generalization capability, making it suitable for automated peripheral blood smear image analysis.

4 | Conclusion

In this study, a robust multi-stage framework for WBC segmentation in microscopic peripheral blood smear images was proposed. Unlike many existing approaches whose performance depends heavily on imaging conditions, staining protocols, or specific datasets, the proposed framework was developed based on three invariant characteristics of peripheral blood images. Furthermore, the output of each processing stage was utilized as guiding information for the subsequent stage, resulting in a coherent segmentation pipeline that minimizes error propagation and improves robustness under varying imaging conditions.

The proposed framework integrates background extraction using the M channel of the CMYK color space and the Zack thresholding algorithm, WBC nucleus detection, RBC region extraction, preliminary WBC segmentation, separation of clustered cells using a modified Marker-Controlled Watershed algorithm based on the distance transform, and the removal of WBC-like artifacts using geometric features within a unified segmentation strategy.

Experimental evaluation on three publicly available blood smear datasets acquired under different imaging and staining conditions demonstrated that the proposed method provides accurate and stable WBC segmentation without requiring dataset-specific parameter adjustment. The results further indicate that exploiting complementary information from background extraction, nucleus detection, and RBC identification, together with the improved marker generation strategy for the Marker-Controlled Watershed algorithm, significantly enhances the separation of adjacent WBCs while effectively reducing over-segmentation. Overall, the proposed method offers an interpretable, computationally efficient, and robust alternative to existing segmentation approaches and is well suited for automated blood image analysis systems. Future work will focus on integrating the proposed framework with deep learning and transfer learning techniques to further improve segmentation performance on highly challenging blood smear images while preserving the robustness and interpretability of the proposed methodology.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

All data are included in the text.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Rezatofighi, S. H., & Soltanian-Zadeh, H. (2011). Automatic recognition of five types of white blood cells in peripheral blood. *Computerized Medical Imaging and Graphics*, *35*(4), 333–343. https://doi.org/10.1016/j.compmedimag.2011.01.003

[2] Putzu, L., Caocci, G., & Di Ruberto, C. (2014). Leucocyte classification for leukaemia detection using image processing techniques. *Artificial Intelligence in Medicine*, *62*(3), 179–191. https://doi.org/10.1016/j.artmed.2014.09.002

[3] Hegde, R. B., Prasad, K., Hebbar, H., & Singh, B. M. K. (2019). Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. *Biocybernetics and Biomedical Engineering*, *39*(2), 382–392. https://doi.org/10.1016/j.bbe.2019.01.005

[4] Madhloom, H. T., Kareem, S. A., Ariffin, H., Zaidan, A. A., Alanazi, H. O., & Zaidan, B. B. (2010). An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold. *Journal of Applied Sciences*, *10*(11), 959–966. https://ui.adsabs.harvard.edu/link_gateway/2010JApSc..10..959M/doi:10.3923/jas.2010.959.966

[5] Habibzadeh, M., Krzyżak, A., & Fevens, T. (2013). Comparative study of shape, intensity and texture features and support vector machine for white blood cell classification. *Journal of Theoretical and Applied Computer Science*, *7*(1), 20–35. file:///C:/Users/Administrator/Desktop/Comparative_study_of_shape_intensity_and_texture_f.pdf

[6] Otsu, N. (1979). A threshold selection method from gray-level histograms. *Automatica*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

[7] Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing*. Pearson. https://www.cl72.org/090imagePLib/books/Gonzales,Woods-Digital.Image.Processing.4th.Edition.pdf

[8] Shahzad, M., Ali, F., Shirazi, S. H., Rasheed, A., Ahmad, A., Shah, B., & Kwak, D. (2024). Blood cell image segmentation and classification: A systematic review. *PeerJ Computer Science*, *10*, e1813. https://doi.org/10.7717/peerj-cs.1813

[9] Anand, V., Gupta, S., Koundal, D., Alghamdi, W. Y., & Alsharbi, B. M. (2024). Deep learning-based image annotation for leukocyte segmentation and classification of blood cell morphology. *BMC Medical Imaging*, *24*(1), 83. https://doi.org/10.1186/s12880-024-01254-z

[10] Abrol, V., Dhalla, S., Gupta, S., Singh, S., & Mittal, A. (2023). An automated segmentation of leukocytes using modified watershed algorithm on peripheral blood smear images. *Wireless Personal Communications*, *131*(1), 197–215. https://doi.org/10.1007/s11277-023-10424-1

[11] Saidani, O., Umer, M., Alturki, N., Alshardan, A., Kiran, M., Alsubai, S., … & Ashraf, I. (2024). White blood cells classification using multi-fold pre-processing and optimized CNN model. *Scientific Reports*, *14*(1), 3570. https://doi.org/10.1038/s41598-024-52880-0

[12] Zoghi, S. (2025). Robust unsupervised white blood cell nucleus segmentation using intuitionistic fuzzy divergence thresholding in the lab color space. *Trends in Health Informatics*, *2*(2), 116–126. https://doi.org/10.22105/thi.v2i2.49

[13] Zack, G. W., Rogers, W. E., & Latt, S. A. (1977). Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry*, *25*(7), 741–753. https://doi.org/10.1177/25.7.70454

[14] Habibzadeh, M., Krzyzak, A., Fevens, T., & Sadr, A. (2011). Counting of rbcs and wbcs in noisy normal blood smear microscopic images. *Medical Imaging 2011: Computer-Aided Diagnosis* (Vol. 7963, pp. 1009–1019). SPIE. https://doi.org/10.1117/12.878748

[15] Lee, H., & Chen, Y. P. P. (2014). Cell morphology based classification for red cells in blood smear images. *Pattern Recognition Letters*, *49*, 155–161. https://doi.org/10.1016/j.patrec.2014.06.010

[16] Huang, D. C., Hung, K. D., & Chan, Y. K. (2012). A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images. *Journal of Systems and Software*, *85*(9), 2104–2118. https://doi.org/10.1016/j.jss.2012.04.012

[17] Soille, P. (2004). *Morphological image analysis: Principles and applications*. Springer. https://doi.org/10.1007/978-3-662-05088-0