
Blood Cancer Identification on Peripheral Blood Smear (PBS) Images Using HSV Feature Extraction**Irohito Nozomi¹, Febri Aldi²**irohito_nozomi@upiyptk.ac.id¹, febri_aldi@upiyptk.ac.id²^{1,2} University of Putra Indonesia YPTK Padang

Article Information

Received : 22 Jun 2024

Revised : 3 Jul 2024

Accepted : 25 Jul 2024

KeywordsBlood Cancer, Feature
Extraction, PBS, SVM**Abstract**

Blood cancer is a category of diseases that have an impact on the development and operation of blood cells. Due to the complexity and diversity of these diseases, proper diagnosis is required before starting treatment. Medical imaging techniques have undergone significant advances in recent years, especially in peripheral blood smear (PBS) image processing. The aim of this study was to uncover how important the extraction of PBS image features is for the diagnosis of blood cancer. Feature extraction is essential to detect anomalies in blood cells in terms of blood cancer detection. The method used is feature extraction based on hue and saturation values (HSV) and uses Machine Support Vector Machine (SVM) machine learning algorithms in classifying malignant and benign PBS images. PBS image data used in this study was 100 images, consisting of 50 malignant PBS images, and 50 benign PBS images. Through the application of HSV feature extraction techniques and PBS image analysis, SVM algorithms can uncover latent indicators of blood cancer and facilitate timely and precise diagnosis. With the SVM technique, classification accuracy can be achieved by 92%. These results demonstrate the potential effectiveness of this feature extraction method. Extraction of HSV features may alter the diagnosis of blood cancer with additional research and application in clinical settings.

A. Introduction

Hematological cancer, another name for blood cancer, is a category of diseases that impact the development and operation of blood cells [1]. Due to the complexity and diversity of these diseases, proper diagnosis is required before starting treatment. Medical imaging techniques have undergone significant advances in recent years, especially in peripheral blood smear (PBS) image processing [2]. PBS provides important details regarding the structure and arrangement of blood cells, which helps in the recognition and categorization of various forms of blood cancer. We will investigate the importance of PBS image feature extraction for proper diagnosis of blood cancer in this work.

Feature extraction is an important method in digital image processing that requires searching and obtaining related data from image or video data [3]. Feature extraction is essential for detecting anomalies in blood cells in terms of blood cancer detection [4]. Algorithms on digital image processing can extract important aspects that aid in blood cancer diagnosis from PBS images by examining the shape, texture, color, and other properties of blood cells.

As shown in Figure 1, peripheral blood smear images (PBS) are microscopic images of blood samples that are often used to diagnose blood diseases, such as blood cancer [5]. Medical practitioners can spot any abnormality or abnormality thanks to these photographs, which offer in-depth information about the structure and nature of blood cells [6]. Computer vision algorithms have the ability to identify and categorize different types of blood cells from PBS images, which can help with early identification and precise diagnosis of blood cancers.

Various attributes were extracted from PBS images with machine learning algorithms in the context of blood cancer screening. These traits include texture-based traits such as homogeneity and granularity in cells as well as shape-based traits such as size, roundness, and irregularity [7]. In addition, hue and saturation (two color-based characteristics) are critical in detecting anomalies in blood cells. Digital image processing algorithms can improve the accuracy of blood cancer diagnosis by generating comprehensive blood cell profiles by combining these various aspects [8].

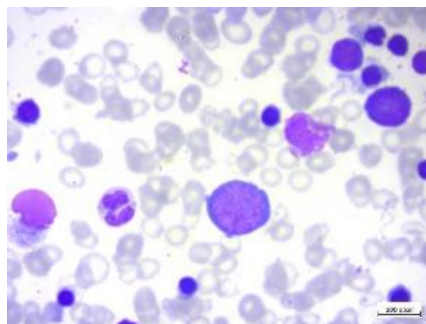


Figure 1. PBS Image

Hue and saturation are important color-based features extracted from PBS images that can be used to identify abnormalities in blood cells [9]. In healthy blood cells, hue and saturation values remain within a certain range [10]. However, if there is blood cancer or other blood disorders, this value may deviate from the normal

range. By analyzing the hue and saturation values of blood cells in PBS images, machine learning algorithms can detect and classify abnormal cells, aiding in early detection and precise diagnosis of blood cancers [11].

Previous studies on PBS image feature extraction, such as the one conducted by [12], reported that better performance, with an accuracy rate of 99.65%, was achieved by classifying trained and improved ResNet50 models. It can be argued that the suggested approach can help in differentiating different groups of ALL and in establishing the diagnosis as well as the course of therapy for laboratory personnel. According to a different study conducted by [13], ten well-known CNN architectures (EfficientNet, MobileNetV3, VGG-19, Xception, InceptionV3, ResNet50V2, VGG-16, NASNetLarge, InceptionResNetV2, and DenseNet201) were used for feature extraction of various data classes after color threshold-based segmentation in the HSV color space by designing a two-channel network. DenseNet201 has the best performance in diagnosis and classification of these ten models. In the end, a model built on the basis of this cutting-edge technology was produced and submitted. The accuracy, sensitivity, and specificity of this deep learning-based model are 99.85, 99.52, and 99.89%, respectively. The suggested approach can help in identifying all harmless situations.

Classifiers trained with LeuFeatx deep features achieved 96.15% accuracy in the [14] study that used ALL_IDB2 data set for binary classification experiments. This method is better than other sophisticated methods reported in the literature. Then, according to research by [15], fifteen different subtypes of white blood cells and platelets as well as morphological abnormalities were classified by deep neural networks trained on synthetic bloodstain data sets. A hybrid method combining deep learning approaches and image processing is suggested for platelet classification. The average macro precision and accuracy of platelet classification increased with this method, increasing from 76.6% to 97.6% and 82.6% to 98.6%, respectively.

White blood cells are classified into six groups, according to different studies by [16]: neutrophils, eosinophils, basophils, lymphocytes, and aberrant cells. For the classification of white blood cells, they present a comparison of deep learning techniques and conventional image processing techniques. An average accuracy of 99.8% was achieved by evaluating the neural network classifier's output for handmade features. In addition, they use full training of convolutional neural networks and transfer learning techniques for categorization. Accuracy of about 99% is achieved when CNN has completed its training.

An automated approach to identifying cell nuclei and extracting leukocytes from images of peripheral blood smears with color and lighting variations was given in the study by [17] under different titles. Leukocyte detection is carried out with an active contour approach, while nuclear detection is carried out by morphological and arithmetic processes. The findings showed that leukocytes and nuclei with Dice scores of 0.96 and 0.97, respectively, were found using the recommended approach. The overall sensitivity of this method is approximately 96%. According to further research, [18] presented an image processing technique resistant to changes in brightness and hue to identify white blood cell nuclei. The recommended approach uses five data sets created by adding brightness changes to the original photo, in addition to the other two data sets. They also compared the results of the suggested

approach with four cutting-edge techniques. These findings demonstrate the precision of the recommended method of detecting atomic nuclei, with a Dice coefficient of 0.965 and an average accuracy of 0.99.

With regard to the above background, we tried to extract features on PBS images based on hue and saturation values and used the Support Vector Machine (SVM) machine learning algorithm to classify malignant and benign PBS images.

B. Research Method

Image processing in extracting PBS image features based on hue and saturation components is carried out to obtain characteristics of the image, so that benign or malignant PBS images can be classified. The first stage of research after collecting PBS image data is to pre-process the PBS image, namely converting RGB to LAB, extracting component "a" from the LAB image, thresholding component "a", and performing morphological operations. Then convert the RGB image to HSV, extract the hue and saturation components. So it can calculate the average value of hue and saturation. The last step is to conduct training on images that have gone through the pre-processing process. This research method is presented in Figure 2 below.

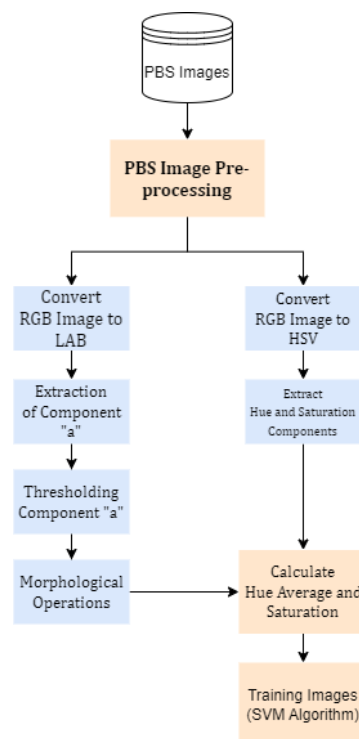


Figure 2. Research Method

Based on the research method in Figure 2 above, the below will be explained step by step in completing this research.

Image Input

The initial process in this study was to input PBS images as many as 50 benign images and 50 malignant images with JPG extensions obtained from kaggle sites.

Figure 3 is an example of a malignant PBS image and figure 4 is an example of a benign PBS image.

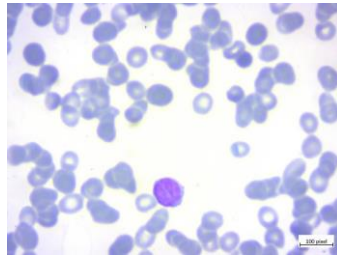


Figure 3. Malignant PBS Image

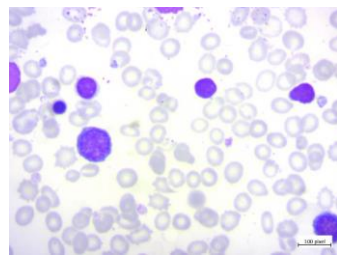


Figure 4. Benign PBS Image

Figures 3 and 4 above show the difference between benign and malignant white blood cells. So the next 100 PBS images are loaded, a pre-processing process will be explained after this.

PBS Image Pre-processing

Convert RGB Image to LAB

Color conversion from RGB color space (red, green, and blue) to LAB color space (lightness, A, and B) is a common transformation process used in digital color management. The BAL alert model is one of several warning models intended to provide warnings in a way that is more in line with human perception [19]. This conversion is useful because the LAB alert range provides information about lightness, basic warning (A), and secondary warning (B), making it easier to analyze and manipulate oranges. In the LAB alert model, component "a" is responsible for the warning of the red-green axis [20].

Extraction of Component "a"

The value of the channel pixel or component "a" of each pixel in the LAB image must be extracted to extract component "a" from the color space of the LAB [21]. The red to green color variation is represented by the "a" component on the LAB color model. Color variations are indicated by positive values for red and negative values for green. In many image processing applications, particularly in color analysis and object detection, extraction of the "a" component is helpful [22].

Thresholding Component "a"

A direct image processing method called threshold is used to split an image into two depending on a predetermined threshold [23]. Here, we will apply a threshold to component "a" in the LAB image to distinguish objects based on the difference in red and green. Through this method, binary images will be generated, with pixels that meet the criteria to be white (maximum value) and those that do not become black (minimum value).

Morphological Operations

By removing small noise, merging pieces, or smoothing object boundaries, morphological techniques can improve segmentation results [24]. One morphological technique used in image processing to fill gaps or small holes in objects contained in binary images is called "holes" [25]. By maintaining the integrity of the corresponding object, the morphological procedure of "holes" aims to improve segmentation or detection of objects. We can ensure that the objects in the binary image have the correct integrity and are prepared for the following stages of image analysis or processing by performing the morphological operation "hole" [26].

Convert RGB Image to HSV

A typical technique in digital image processing is the conversion of images from RGB color systems (Red, Green, Blue) to HSV (Hue, Saturation, Value) color formats. It is often easier to understand how color information is separated in the HSV color space based on attributes such as hue, saturation, and value than when colors are represented in the RGB color space [27].

Extract Hue and Saturation Components

Ensuring images have been converted to the HSV color space is the first step in extracting the Hue (H) and Saturation (S) components from images in the HSV color system. First things first, conversion needs to be done if the original image is within the RGB color space. Next, get access to the S and H sections. The three components of each pixel in an HSV image are Value (V), Saturation (S), and Hue (H). It is necessary to get the value of each pixel to extract the H and S components. Next, calculate the values of the Hue (H) and Saturation (S) components by iterating over each pixel in the image. Once extraction is complete, the H and S components can be saved in an appropriate format or used for more complex analyses, including object recognition, color segmentation, or feature detection. The average value of S is determined using equation 2, and the average value of H is determined using equation 1.

$$t(n, 1) = \frac{\sum_{i=1}^m \sum_{j=1}^n h(i, j)}{\sum_{i=1}^m \sum_{j=1}^n bw(i, j)} \quad (1)$$

$$t(n, 2) = \frac{\sum_{i=1}^m \sum_{j=1}^n s(i, j)}{\sum_{i=1}^m \sum_{j=1}^n bw(i, j)} \quad (2)$$

Where $t(n,1)$ is the n th element of the vector t (Train). $h(i,j)$ are the values of the i row and j column of Hue. $s(i,j)$ are the values of the i row and j column of

saturation. $bw(i,j)$ are the values of the i row and j column of the binary image. M is the number of rows in the image, while n is the number of columns in the image.

Image Training

Benign and malignant PBS images are used for RGB image training after undergoing pre-processing. SVM is the algorithm used in this training. For classification and regression tasks, machine learning techniques such as Support Vector Machine (SVM) algorithms are used [28]. The main goal of SVM is to identify the most effective hyperplane for dividing training data classes. SVM can categorize new data by using this hyperplane as a decision limit [29]. Equation 3 can be used to get an accuracy value after the SVM algorithm generates the training results.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

C. Result and Discussion

The results and discussion that will be presented in this section are the results of step by step in the PBS image processing process, as explained in the research methods section. Here we explain the results and discussion in each process.

Convert RGB Image to LAB

PBS input images, which are essentially RGB images, are converted to LAB to separate base and secondary colors. Figure 5 is the input image and Figure 6 is the converted image.

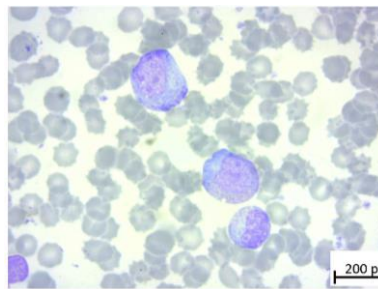


Figure 5. PBS Input Image

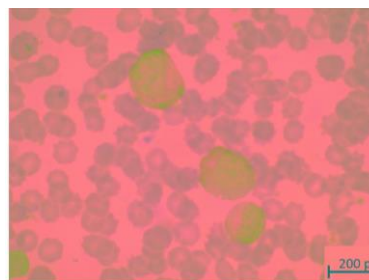


Figure 6. LAB Converted Image

The converted image has three components: L (brightness), a (red-green), and b (yellow-blue). The LAB image is extracted on component a. Thus producing the extract results presented in Figure 7 below.

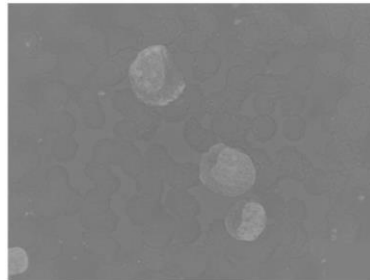


Figure 7. Image of Component Extract Results 'a'

The image produced after the extract on component a only has red to green color information from the original image. Then a thresholding process is carried out on component a. The thresholding results are presented in Figure 8.

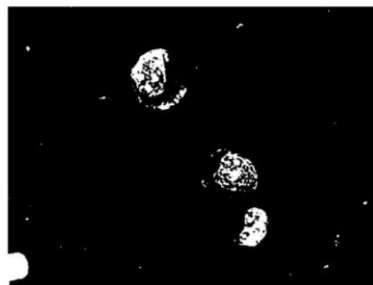


Figure 8. Image of Thresholding Results of Component 'a'

The threshold value used to perform thresholding is 140. All pixels in component "a" that have a value greater than 140 will be considered part of the object of interest. Binary images that have a component value of "a" above 140 will be converted to a logical value of 1 (white), while pixels that have a component value of "a" below or equal to 140 will be converted to a logical value of 0 (black). Next, perform morphological operations of holes to perfect the resulting segmentation results presented in Figure 9.

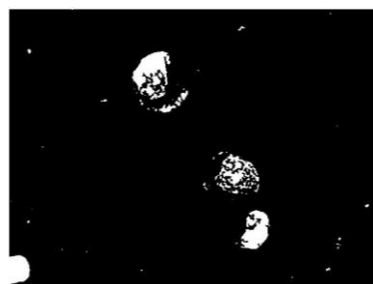


Figure 9. Image of Morphological Holes Surgery Results

The holes in the detected object must be filled, which is the part of the disconnected and unwanted object in the segmentation result.

Convert RGB Image to HSV

The result of RGB PBS to HSV image conversion is a step carried out to calculate the average of the hue and saturation values. Equation 1 is the equation for calculating the average. The results of the image conversion to HSV can be seen in Figure 10.

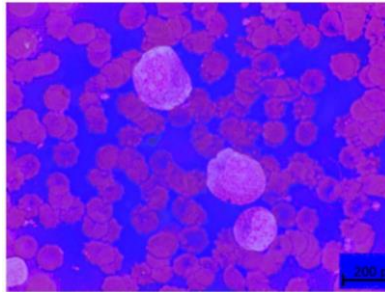


Figure 10. HSV Converted Image

Jumlah nilai rona dan saturasi masing-masing dibagi dengan jumlah nilai gambar biner yang telah disempurnakan pada operasi morfologis. Hasil perhitungan tersebut dapat dilihat pada Tabel 1. Nilai rona dan saturasi pada Tabel 1 adalah karakteristik gambar PBS. Sebanyak 100 gambar PBS yang telah dilatih melihat gambar PBS jinak dari nomor 1 hingga 50. Nilai rona jinak rata-rata adalah 0,68, sedangkan nilai saturasi jinak rata-rata adalah 0,48. Kemudian untuk gambar PBS ganas dari angka 51 sampai 100 menunjukkan nilai rata-rata nilai hue adalah 0,72, sedangkan nilai saturasi rata-rata adalah 0,50. Sehingga terdapat perbedaan yang signifikan antara karakteristik PBS ganas dan PBS jinak.

Table 1. Average Hue and Saturation Values

Image	Hue	Saturation	Type
1 - 50	0,68	0,48	Benign
51 - 100	0,72	0,50	Malignant

Image Training

Based on the hue and saturation values that have been obtained from 100 benign and malignant PBS images, the next step is to conduct training on these images. Training is conducted using the SVM algorithm. The results of the training obtained an accuracy value of 92%. The accuracy value shows good performance against benign and malignant PBS image classification.

D. Conclusion

The ability to extract features from PBS images using hue and saturation values holds great promise for the diagnosis of benign or malignant blood cancers. Through the application of HSV feature extraction techniques and PBS image analysis, SVM algorithms can uncover latent indicators of blood cancer and facilitate timely and

precise diagnosis. With the SVM technique, classification accuracy can be achieved by 92%. These results demonstrate the potential effectiveness of this feature extraction method. Despite barriers and limitations, new developments in these methods and ongoing research provide encouraging prospects. Extraction of HSV features may alter the diagnosis of blood cancer with additional research and application in clinical settings.

E. References

- [1] C. Thakur, P. Nayak, V. Mishra, M. Sharma, and G. K. Saraogi, "Treating blood cancer with nanotechnology: A paradigm shift," *Nano Drug Deliv. Strateg. Treat. Cancers*, pp. 225–243, Jan. 2021, doi: 10.1016/B978-0-12-819793-6.00010-2.
- [2] M. Ghaderzadeh, M. Aria, A. Hosseini, F. Asadi, D. Bashash, and H. Abolghasemi, "A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images," *Int. J. Intell. Syst.*, vol. 37, no. 8, pp. 5113–5133, Aug. 2022, doi: 10.1002/INT.22753.
- [3] J. Jing, T. Gao, W. Zhang, Y. Gao, and C. Sun, "Image Feature Information Extraction for Interest Point Detection: A Comprehensive Review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4694–4712, Apr. 2023, doi: 10.1109/TPAMI.2022.3201185.
- [4] N. M. Deshpande, S. S. Gite, and R. Aluvalu, "Microscopic Analysis of Blood Cells for Disease Detection: A Review," *Intell. Syst. Ref. Libr.*, vol. 206, pp. 125–151, 2022, doi: 10.1007/978-3-030-76732-7_6/COVER.
- [5] J. P. Phukan, H. Kawsar, J. Banerjee, and A. Sinha, "A comparative study of anemia in peripheral blood smear and automated cell counter generated red cell parameters," 2022, doi: 10.4103/ijh.ijh_3_22.
- [6] V. Prasad, D. Sharma, and S. Lata Verma, "Peripheral Blood Smear Examination: An Overview," vol. 19, no. 2, p. 3170, 2022, Accessed: Feb. 26, 2024. [Online]. Available: <http://www.webology.org>
- [7] S. Saleem, J. Amin, M. Sharif, G. A. Mallah, S. Kadry, and A. H. Gandomi, "Leukemia segmentation and classification: A comprehensive survey," *Comput. Biol. Med.*, vol. 150, p. 106028, Nov. 2022, doi: 10.1016/J.COMPBIOMED.2022.106028.
- [8] A. Shah, S. S. Naqvi, K. Naveed, N. Salem, M. A. U. Khan, and K. S. Alimgeer, "Automated Diagnosis of Leukemia: A Comprehensive Review," *IEEE Access*, vol. 9, pp. 132097–132124, 2021, doi: 10.1109/ACCESS.2021.3114059.
- [9] V. Abrol, S. Dhalla, S. Gupta, S. Singh, and A. Mittal, "An Automated Segmentation of Leukocytes Using Modified Watershed Algorithm on Peripheral Blood Smear Images," *Wirel. Pers. Commun.*, vol. 131, no. 1, pp. 197–215, Jul. 2023, doi: 10.1007/S11277-023-10424-1/TABLES/6
- [10] H. A. Muhamad, S. W. Kareem, and A. S. Mohammed, "A Comparative Evaluation of Deep Learning Methods in Automated Classification of White Blood Cell Images," *8th IEC 2022 - Int. Eng. Conf. Towar. Eng. Innov. Sustain.*, pp. 205–211, 2022, doi: 10.1109/IEC54822.2022.9807456.
- [11] C. K. Jha, A. Choubey, M. H. Kolekar, and C. Chakraborty, "Automated acute lymphoblastic leukemia detection using blood smear image analysis," *Mach. Learn. Deep Learn. Tech. Med. Image Recognit.*, pp. 64–75, Dec. 2023, doi: 10.1201/9781003366249-4/Automated-Acute-Lymphoblastic-Leukemia-

- Detection-Using-Blood-Smear-Image-Analysis-Chandan-Kumar-Jha-Arvind-Choubey-Maheshkumar-Kolekar-Chinmay-Chakraborty.
- [12] E. Özbay, F. A. Özbay, and F. S. Gharehchopogh, "Peripheral Blood Smear Images Classification for Acute Lymphoblastic Leukemia Diagnosis with an Improved Convolutional Neural Network," *J. Bionic Eng.*, pp. 1–17, Oct. 2023, doi: 10.1007/S42235-023-00441-Y/METRICS.
 - [13] M. Ghaderzadeh, F. Asadi, A. Hosseini, D. Bashash, H. Abolghasemi, and A. Roshanpour, "Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review," *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/9933481.
 - [14] P. Rastogi, K. Khanna, and V. Singh, "LeuFeatx: Deep learning-based feature extractor for the diagnosis of acute leukemia from microscopic images of peripheral blood smear," *Comput. Biol. Med.*, vol. 142, p. 105236, Mar. 2022, doi: 10.1016/J.COMPBIOMED.2022.105236.
 - [15] R. Al-qudah and C. Y. Suen, "Improving blood cells classification in peripheral blood smears using enhanced incremental training," *Comput. Biol. Med.*, vol. 131, p. 104265, Apr. 2021, doi: 10.1016/J.COMPBIOMED.2021.104265.
 - [16] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Development of a robust algorithm for detection of nuclei of white blood cells in peripheral blood smear images," *Multimed. Tools Appl.*, vol. 78, no. 13, pp. 17879–17898, Jul. 2019, doi: 10.1007/S11042-018-7107-X/METRICS.
 - [17] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 382–392, Apr. 2019, doi: 10.1016/J.BBE.2019.01.005.
 - [18] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Image Processing Approach for Detection of Leukocytes in Peripheral Blood Smears," *J. Med. Syst.*, vol. 43, no. 5, pp. 1–11, May 2019, doi: 10.1007/S10916-019-1219-3/METRICS.
 - [19] M. A. Ansari and D. K. Singh, "Significance of Color Spaces and Their Selection for Image Processing: A Survey," *Recent Adv. Comput. Sci. Commun.*, vol. 15, no. 7, Mar. 2021, doi: 10.2174/2666255814666210308152108.
 - [20] M. Jawahar, L. J. Anbarasi, S. G. Jasmine, and M. Narendra, "Diabetic Foot Ulcer Segmentation using Color Space Models," pp. 742–747, Jul. 2020, doi: 10.1109/ICCES48766.2020.9138024.
 - [21] Y. Zhang, Z. Dong, K. Zhang, S. Shu, F. Lu, and J. Chen, "Illumination variation-resistant video-based heart rate monitoring using LAB color space," *Opt. Lasers Eng.*, vol. 136, p. 106328, Jan. 2021, doi: 10.1016/J.OPTLASENG.2020.106328.
 - [22] C. H. Yeh et al., "Lightweight Deep Neural Network for Joint Learning of Underwater Object Detection and Color Conversion," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 11, pp. 6129–6143, Nov. 2022, doi: 10.1109/TNNLS.2021.3072414.
 - [23] E. H. Houssein, B. El-din Helmy, D. Oliva, A. A. Elngar, and H. Shaban, "Multi-level Thresholding Image Segmentation Based on Nature-Inspired Optimization Algorithms: A Comprehensive Review," *Stud. Comput. Intell.*, vol. 967, pp. 239–265, 2021, doi: 10.1007/978-3-030-70542-8_11/COVER

- [24] R. A. Lotufo, R. Audigier, A. V. Saúde, and R. C. Machado, "Morphological Image Processing," *Microsc. Image Process. Second Ed.*, pp. 75–117, Jan. 2023, doi: 10.1016/B978-0-12-821049-9.00012-5.
- [25] S. Bhutada et al., "Opening and closing in morphological image processing," <https://wjarr.com/sites/default/files/WJARR-2022-0576.pdf>, vol. 14, no. 3, pp. 687–695, Jun. 2022, doi: 10.30574/WJARR.2022.14.3.0576.
- [26] X. Sun et al., "Moving Object Detection Algorithm based on Improved Visual Background Extractor," *J. Phys. Conf. Ser.*, vol. 1732, no. 1, p. 012078, Jan. 2021, doi: 10.1088/1742-6596/1732/1/012078.
- [27] M. Kamiyama and A. Taguchi, "Color conversion formula with saturation correction from HSI color space to RGB color space," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E104A, no. 7, pp. 1000–1005, 2021, doi: 10.1587/TRANSFUN.2020EAL2087.
- [28] F. Zhang and L. J. O'Donnell, "Support vector regression," *Mach. Learn. Methods Appl. to Brain Disord.*, pp. 123–140, Jan. 2020, doi: 10.1016/B978-0-12-815739-8.00007-9.
- [29] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/J.NEUCOM.2019.10.118.