White Blood Cell, Leukocyte, Identification

ECSE 4540 – Final Project

Mitchell Phillips - 661060944

5/1/2017

Table of Contents

Introduction

Hematologists often investigate blood smears to diagnose diseases by detecting and identifying white blood cells (WBCs). Traditional techniques to do this are somewhat time-consuming and methods do exist to automate the process. Companies like CellaVision [1] offer customers proprietary software and equipment for WBC identification, and a simplified solution to perform classification. This paper investigates several methods to automate WBC identification and proposes a new algorithm to detect WBCs from nucleus locations. The proposed algorithm features preliminary cytoplasm identification as well, with intended applications for classification. The algorithm would be useful for lab assistants and hematologists who inspect large sets of blood smears and prefer to use an alternative to current methods such as CellaVision.

Background

There exist multiple types of WBCs, each of which can be featured given an image of a blood smear. An example is shown below in Figure 1 - Different Types of Leukocytes and Test Image1. In the image, there are 6 large purple blobs. These are the white blood cells; however, not all of these cells are the same. There are five different types of white blood cells; each of which can be distinguished based on their appearance: (1.) Neutrophils, (2.) Lymphocytes, (3.) Monocytes, (4.) Eosinophils, and (5.) Basophils. The different types are labelled in the figure below.

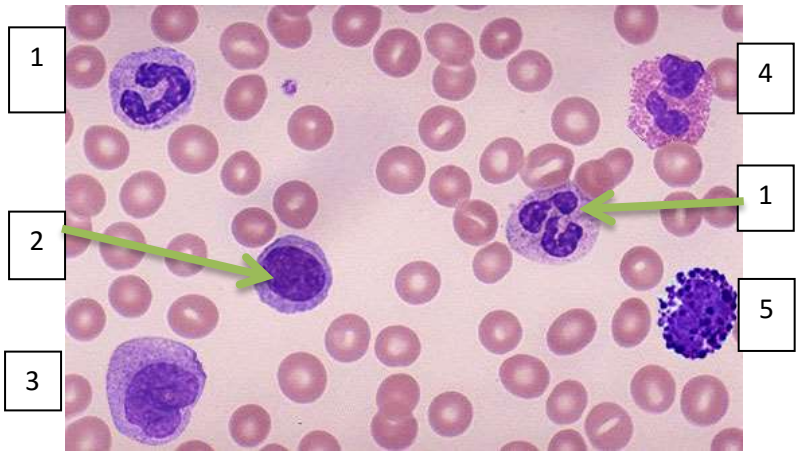

Figure 1 - Different Types of Leukocytes and Test Image [2]

From the above figure, there are 6 large purple blobs, showcasing the five different types of WBCs that exist. Neutrophils can be identified by their pinched-in nucleus (darker purple region) which may have a horseshoe or sausage link appearance. Lymphocytes contain a large, generally round and uniform nucleus. Monocytes are distinguishable based on their kidney bean shape of the nucleus and large amounts of cytoplasm. Eosinophils are similar to neutrophils in appearance, but with large granules in the cytoplasm. Basophils are the easiest to distinguish as they are characterized by their multiple, large granules. These five types of leukocytes can then be split into two distinct groups based on the granules;

1.) Granulocytes - Basophil, Eosinophil, and Neutrophil
2.) Agranulocytes - Lymphocytes and Monocytes [3]

Related Work

Automated leukocyte identification is a popular topic as current techniques for identifying white blood cells and investigating blood smears are rather time consuming. In Mohammed et al. [4], a nuclei segmentation algorithm was proposed using gray scale contrast enhancement and filtering. Sadeghian et al. investigated the application of an active contour model to perform nuclei segmentation and Zack thresholding to identify cytoplasm locations [5]. Prinyakupt and Pluempitwiriyawej investigated segmentation and WBC categorization through the use of linear and naïve Bayes classifiers, with preprocessing focusing on nucleus segmentation in the RGB color space [6]. Common image processing techniques and methods used to perform leukocyte identification and segmentation include: edge detection, morphological opening and closing, thresholding (including global thresholding / Otsu's method), contrast enhancement, and histogram equalization. The methods used and steps of execution are extremely important when detecting WBCs from a blood smear as each play a significant role in the procedure.

Initial approaches in this paper first utilized the nucleus segmentation and WBC identification algorithms from [4] and [6]. However, preliminary tests indicated that they were not as robust as they claimed to be. One image may work great in one, but terribly in the other (and vice-versa). Additionally, the only noticeable adjustably parameter in both algorithms was the structuring element size for the morphological processes. It was desired then, that the proposed algorithm in this paper can be tunable in such a way that a user may only need to manually inspect one to three images, determine the corresponding tuning parameters, and run the rest of their data set without worry.

Data Collection

The blood smear images used for the duration of this procedure were obtained from various online sources, with the largest data set being contributed by Mostafa Mohamed from Mathworks File Exchange [7]. Other sources for images include The American Society of Hematology (ASH) Image Bank [8] and the Bloodline Image Atlas [9]. While developing the algorithms, only a handful of images were considered, whereas testing featured a random batch of 40 – 50 images. The development images where chosen based on their clarity or if certain features were present (either optimal or suboptimal). During testing, the algorithm would be tweaked using the random sample set from Mohamed's data bank. Several hand selected images from Bloodline were also used to check for robustness. The images from Bloodline needed to be hand selected due to the inclusion of unwanted features in some images (eg. strange cross hairs, orange tint, already segmented images etc.).

In the intermediate results, the image from [2] was used for the initial construction of the WBC identification algorithm, with current testing be performed using 35+ images from Mohamad's data set (it is intended to use a batch of 50 while discarding any repeats). It was decided to use Mohamad's set due to having both a large amount of images, but also consisting of strictly blood smears.

Technological Approach

The proposed algorithm attempts to improve upon the methods suggested in [4] and [6], and is heavily based off the pre-processing procedure in [6]. For both brevity and to prevent redundancy, a flow diagram and subsequent Pseudo code is presented below before breaking down the algorithm in the following section. It is noted that MATLAB and the corresponding Image Processing Toolbox were used for implementation.

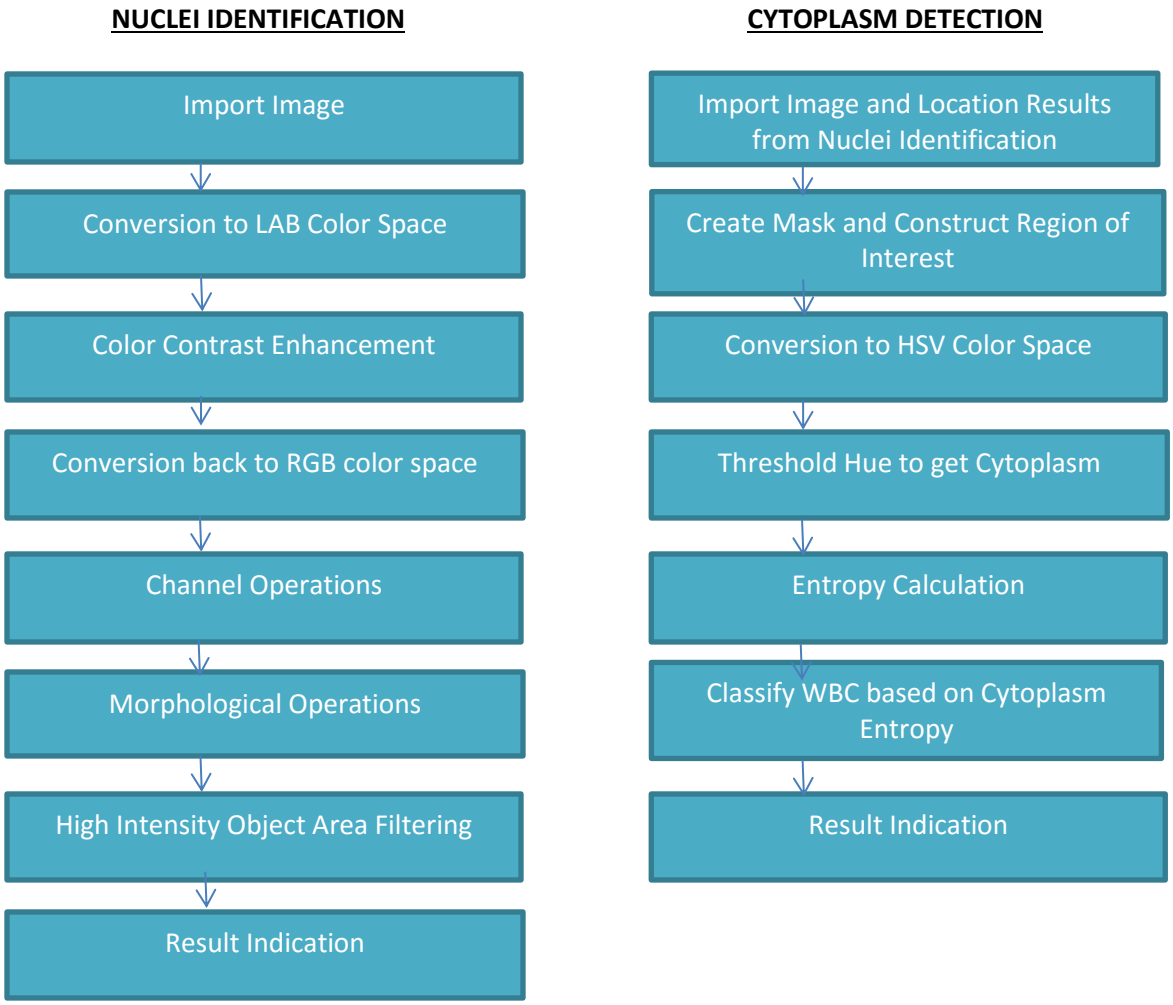

Figure 2- Technological Approach Flow Chart

Figure 2 is the flow chart corresponding to the proposed algorithm. The pseudo code is as follows.

Pseudo Code:

WBC / Nuclei Identification

```
1.) import blood smear image

2.) convert image to LAB color space
3.) obtain normalized luminosity values
4.) adjust luminosty using normalized value by contrast stretching
5.) multiply adjusted value by max luminosity
6.) replace the luminosity of the original image with the adjusted one
7.) convert back to RGB color space

8.) take difference between weighted blue and red channels, divide the result by the green channel
9.) threshold channel operation result by a predefined value (preventative measure)

10.) morphologically open thresheld image
11.) morphologically close image result from opening

12.) for number of high intensity objects in morphologically processed image
13.)     if high intensity object area greater than threshold area
14.)              obtain centroid of object in image
15.)              perform cytoplasm detection using object location
16.)              place circle on original image at the centroid location


16.) return result to user
```

Cytoplasm Detection

```
1.) import blood smear image
2.) import region information from WBC Identification

3.) construct mask from WBC region information
4.) construct region of interest (ROI) on imported image

5.) convert ROI to HSV color space
6.) threshold hue of ROI to get cytoplasm

7.) calculate entropy of cytoplasm
8.) classify WBC based of entropy of cytoplasm

9.) return result to user
```

The contrast enhancement and channel operations are outlined here. Threshold levels, weight decisions, and color space analysis are featured in the following section, with step by step algorithm application discussion.

A color contrast enhancement was performed in LAB color space, utilizing the luminosity of the image. Note the enhancement technique used is summarized in the tutorial section on Mathworks' website [10]. After converting the enhanced image back to RGB color space, the channels were manipulated in such a way that results in obtaining a binary image, where high intensity values are featured at WBC nucleus locations.

LAB color space allows all the colors of the spectrum to be accessed, where L is the value corresponding to luminance, and a & b are the components for color [11]. Here, only the luminance values are considered. Once an image was converted, the luminosity values were normalized so they were in the range from 0 to 1.

[1]

$$L_{max} = 100 \text{ and } L_{norm} = \frac{L_{im}}{L_{max}}$$

L_{max} is the maximum obtainable luminance value and $L_{im}$is the luminance value of the image of interest. The contrast stretching procedure then involves setting the luminance value to a range such that 1 % of the pixel intensities of the normalized luminance values are saturated at both the low and high ends, and scaling this result by the maximum value.

[2]

$$L_{im}{}' = \text{stretched}(L_{norm}) * L_{max}$$

Using$L_{im}{}'$, the image is converted back to RGB color space where the channel operations occur.

The channel operations involve taking a weighted difference between the blue and red channels, and then dividing the result by the green.

[3]

$$im_{co} = \frac{W_B * B - W_R * R}{G}$$

B represents the pixel intensities in the blue channel, *R* are the intensities in the red, and *G* are the intensities in the green. W_B and W_R are the weighting factors of the blue and red channels respectively. For the duration of the experiment and the results presented in this paper, weighting factors of $W_B = 1$ and $W_R = 0.75$ were used. These were determined through trial and error, and gave sufficient results highlighting the WBCs. The following section documents the decision to perform the above weighting procedure, in addition to the application of the remaining standard image processing techniques.

Intermediate Results & Technological Approach Expansion

The following section features the intermediate steps and results for the proposed algorithm, expanding on the details presented in the Technological Approach Section. Both strengths and weaknesses of the current algorithm are presented.

RGB Investigation

Following [6], the grayscale and the individual RGB color channels were investigated.

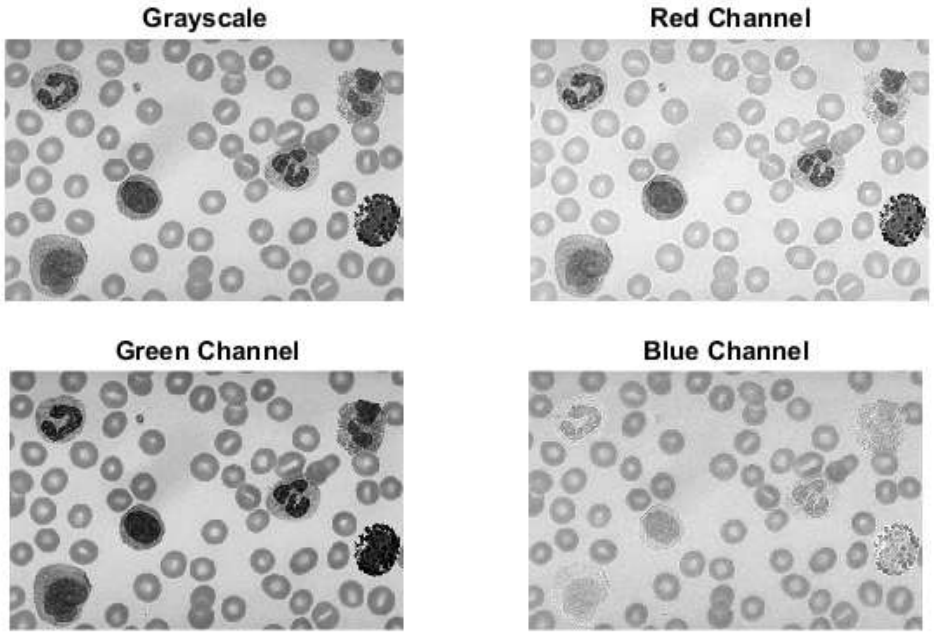

Figure 3 - RGB and Grayscale Channels for Test Image

Figure 3 contains the grayscale, red, green, and blue color channels of the image shown in Figure 1. It is observed that in the test image, the WBCs contain large intensities in the blue channel; however that is true for most of the image. This is indicated by the fact that the majority of the image is made up of lighter shades of gray in the blue channel image. Conversely, in the grayscale, red, and green channels, the WBCs are characterized by lower intensity values. The intensity behavior can be attributed to the violet color of the WBCs from the original image. It is noted that the dark characterization however is only true for the nuclei, not the cytoplasm.

To gain a better understanding of the color distribution of the blood smears, a color histogram for the test image was constructed.

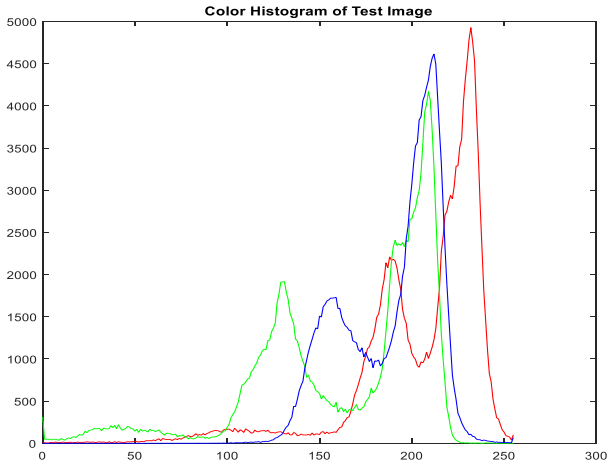

Figure 4 - Color Histogram of Test Image

Figure 4 is the color histogram of the test image. From the color histogram, three distinct sections can be determined. The three sections correspond to the nuclei, the cytoplasm & red blood cells, and the background. Each of these sections can be interpreted from the color histogram based on the channel peaks. The easiest section to determine is the background which is identified by the largest peaks in each color channel.

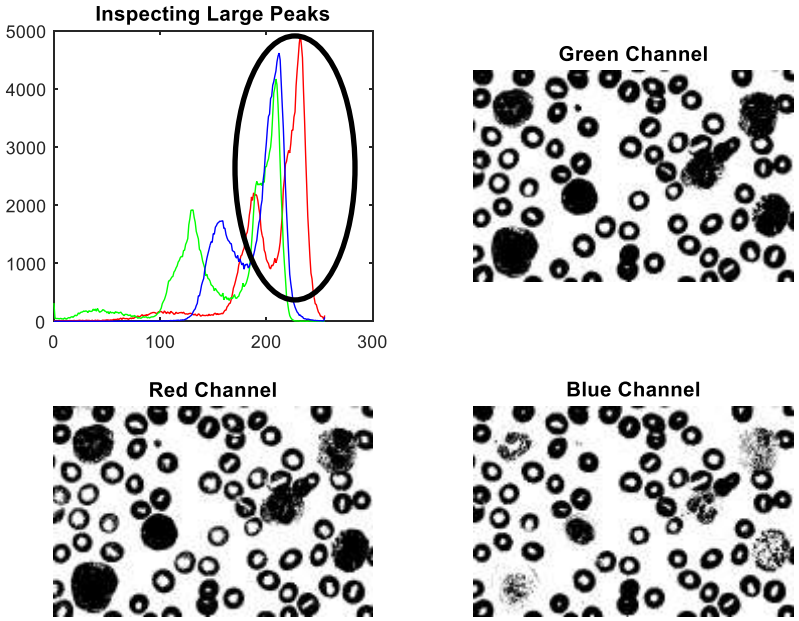

Figure 5 - Large Color Histogram Peaks

Thresholding the color channels at their respected peaks results in Figure 5. The red and green channels produce the anticipated results, but the blue channel produces additional high intensity values in cytoplasm locations. This behavior can be observed in Figure 3 as the cytoplasm contains the highest (brightest) intensity values in the blue channel.

The middle peaks of the color histogram belong to the red blood cells for all the channels and the cytoplasm in the red & green channels. This is verified by Figure 6 below.

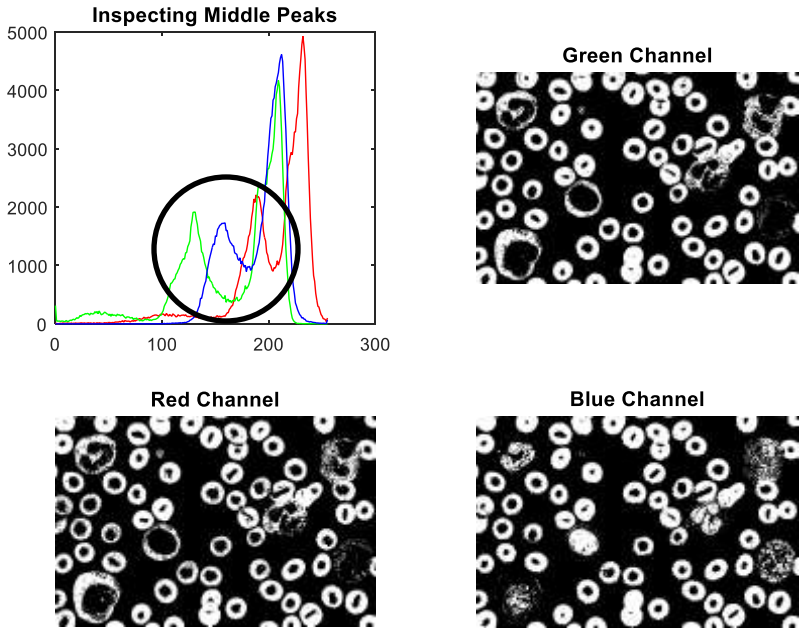

Figure 6 - Middle Color Histogram Peaks

It is observed in the figure above that the red blood cells do indeed produce high intensity values in each of the color channels. For the red and green channels, the cytoplasm is also highlighted. However, in the blue channel, both the red blood cells and portions of the nuclei are indicated.

The smallest peaks of the red and green channels represent the nuclei locations in the image. This is shown below with the blue channel pictured as well for consistency.

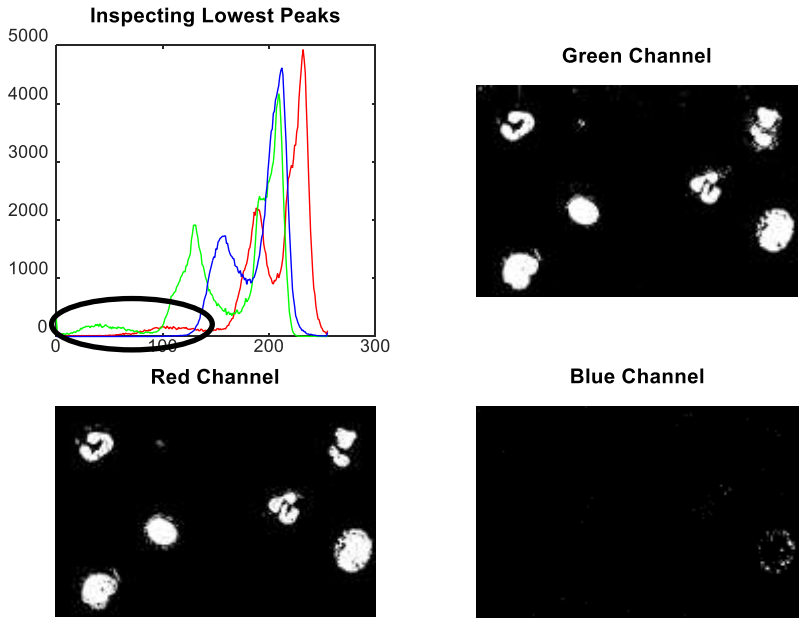

Figure 7 - Lowest Color Histogram Peaks

Figure 7 contains the lowest histogram peaks where the nuclei are visualized in the red and green channels. The blue channel produces little to no high intensity values at the lowest peaks, which can be credited to the fact that the considered image contains mostly high blue channel intensities. Results from the color distribution analysis were helpful in the development of the proposed algorithm.

PPA Analysis

Using a similar process as [6], nucleus segmentation was first performed by averaging the red and blue channel intensities over the green channel. This was done because the green channel contains the lowest pixel intensities at the nucleus (as indicated in Figure 3). However, after trying to replicate the rest of Prinyakupt and Pluempitiwiriyawej's algorithm (PPA), the results were not as robust as initially anticipated. When using selected images from Bloodline [9] and the data set by Mohamed [7], there were large inaccuracies. Results varied substantially from not being able to identify any white blood cells, to picking out way more than what should be present.

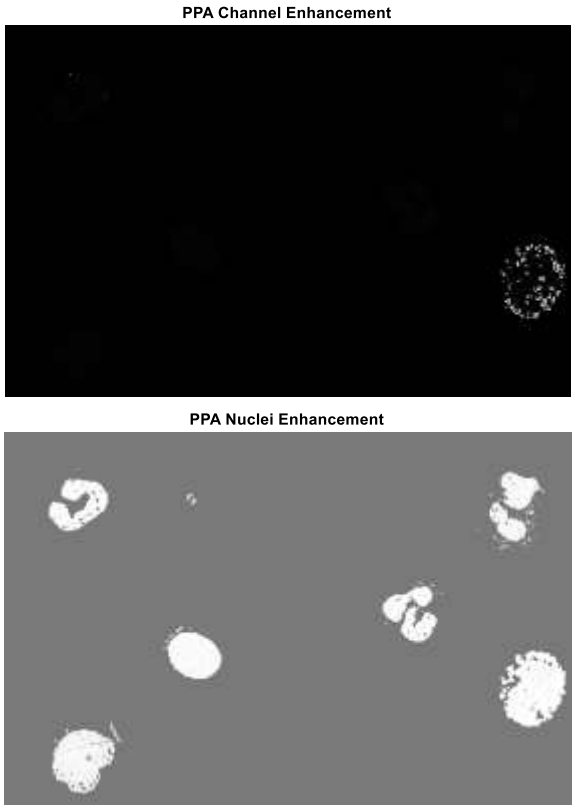

Figure 8 - PPA Channel Enhancement & Histogram Equalization

Figure 8 showcases a key step for PPA and the believed cause of its failure. Histogram equalization is performed on the channel operation of averaging the red and blue channel intensities over the green channel. From the results, it is observed that the channel operation does not produce significant, visual results, which is why the histogram equalization is applied. For the test image, PPA works well and produces acceptable results. However, this is not always the case. This is demonstrated on a random sample of test images from Mohamed's data set.

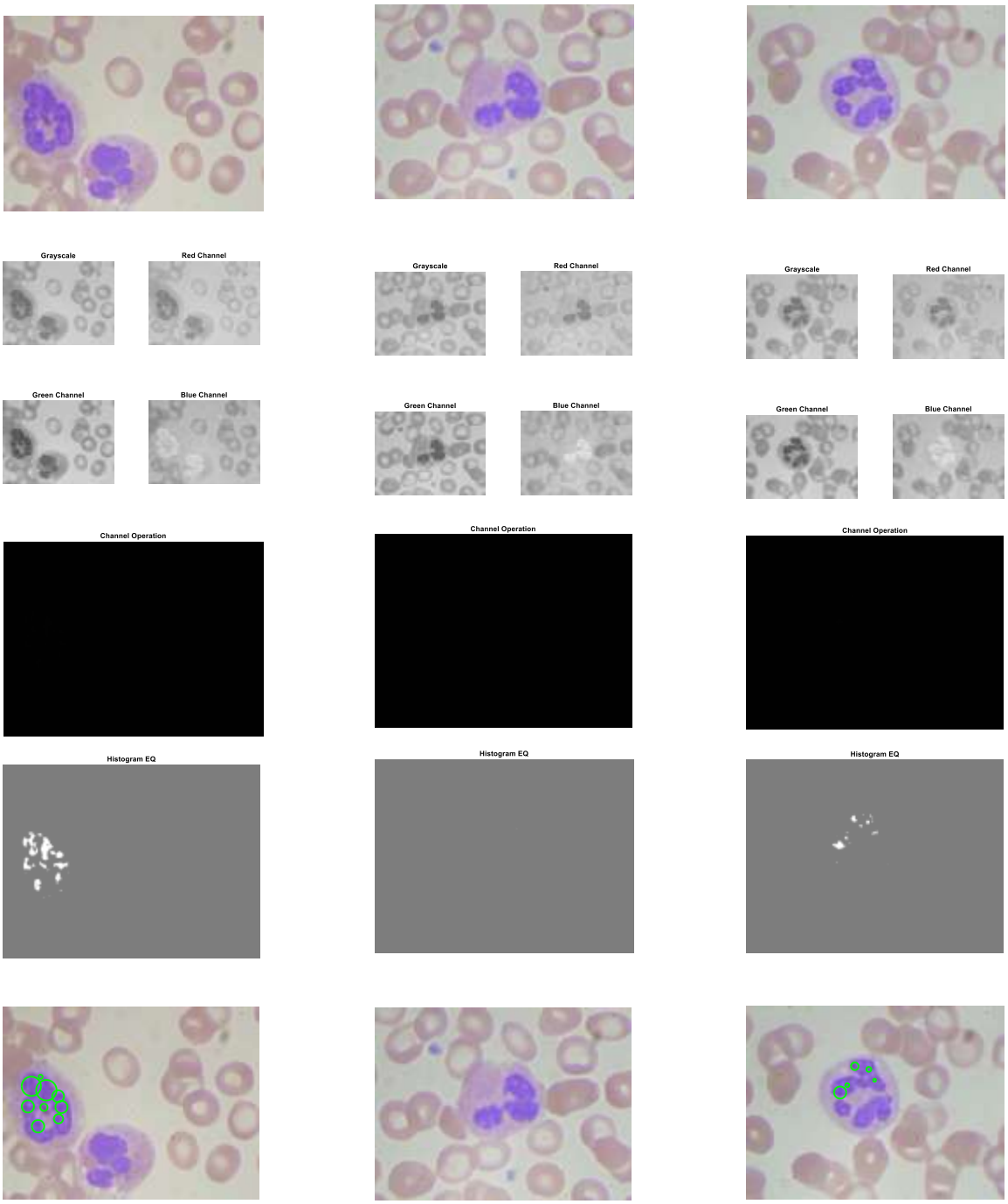

Figure 9 - PPA Results

Based on the obtained results, it is observed that several errors occur. The over estimation of the WBCs may be attributed to the structuring element size of the morphological process (not pictured and occurs after equalization); however, even if the structuring element was changed, the results from the histogram equalization and corresponding channel operations are of concern and lead to the majority of issues. It is believed that the cause of the issue is both the blurriness of the blood smear and the weak contrast of the WBCs.

Proposed Algorithum – Nuclei Identification

The proposed algorithm instead uses the idea of initially applying a contrast enhancement to the image, similarly to the solution introduced in Mohamed et al. [4]. Capitalizing on the strong violet colors of the nuclei, a color contrast enhancement was performed in LAB color space where the luminosity values were manipulated to adjust the image [10]. After adjustment, the resulting image was obtained.

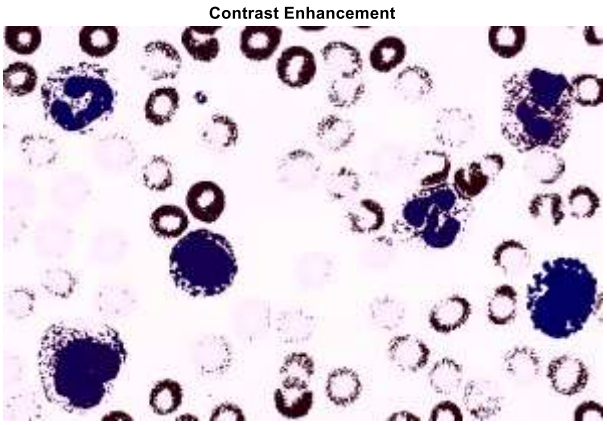

Figure 10 - Contrast Enhancement

Immediately, it is observed in Figure 10 that the nuclei now appear to have a strong, dark blue color. It is noted that the difference in color between the nuclei and the rest of the image is stronger than before. Similarly to PPA and due to the occurrence of the change in color, the RGB color channels were investigated once again. This is presented below.

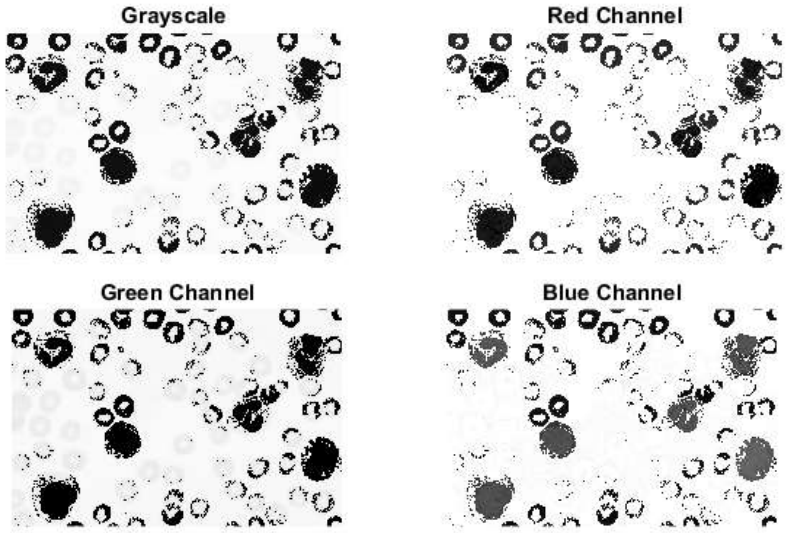

Figure 11 - RGB Channels after Contrast Enhancement

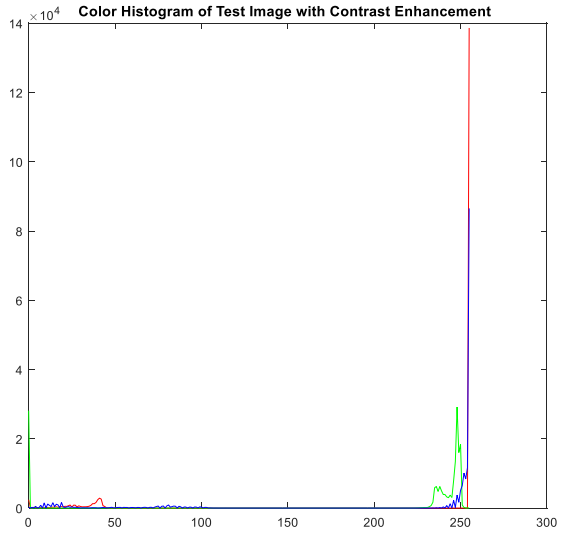

Figure 12 - Color Histogram after Contrast Enhancement

Figure 11 contains the grayscale rendition of the enhanced image in addition to each of the respected RBG color channels. Figure 12 is the corresponding color histogram after contrast enhancement. The obtained results were not as expected as the blue channel does not feature as strong of a response that was hoped for at the nucleus locations. Instead very strong blue channel intensities occur in the background, which can be verified in the color histogram. Based on the color histogram, all the color channels are either very high or very low (hence the contrast enhancement). Both the red and green channels contain very low intensities at nucleus locations, with high intensity values in the unoccupied areas. Using the information obtained from Figure 11 and Figure 12, a similar approach as PPA was deployed. As opposed to averaging the pixel values however, it was decided to weight the channels, and divide the difference between the weighted blue and red channels by the green channel (as presented in the previous section). This resulted in having only high pixel intensity values at the WBC locations. Since operations were performed using a uint8 data type in MATLAB, results were bounded by 0 or 255, producing a binary image. The selected weights for the channels were based on the experimental results from the procedure. Several images are presented below demonstrating the different weighting effects.

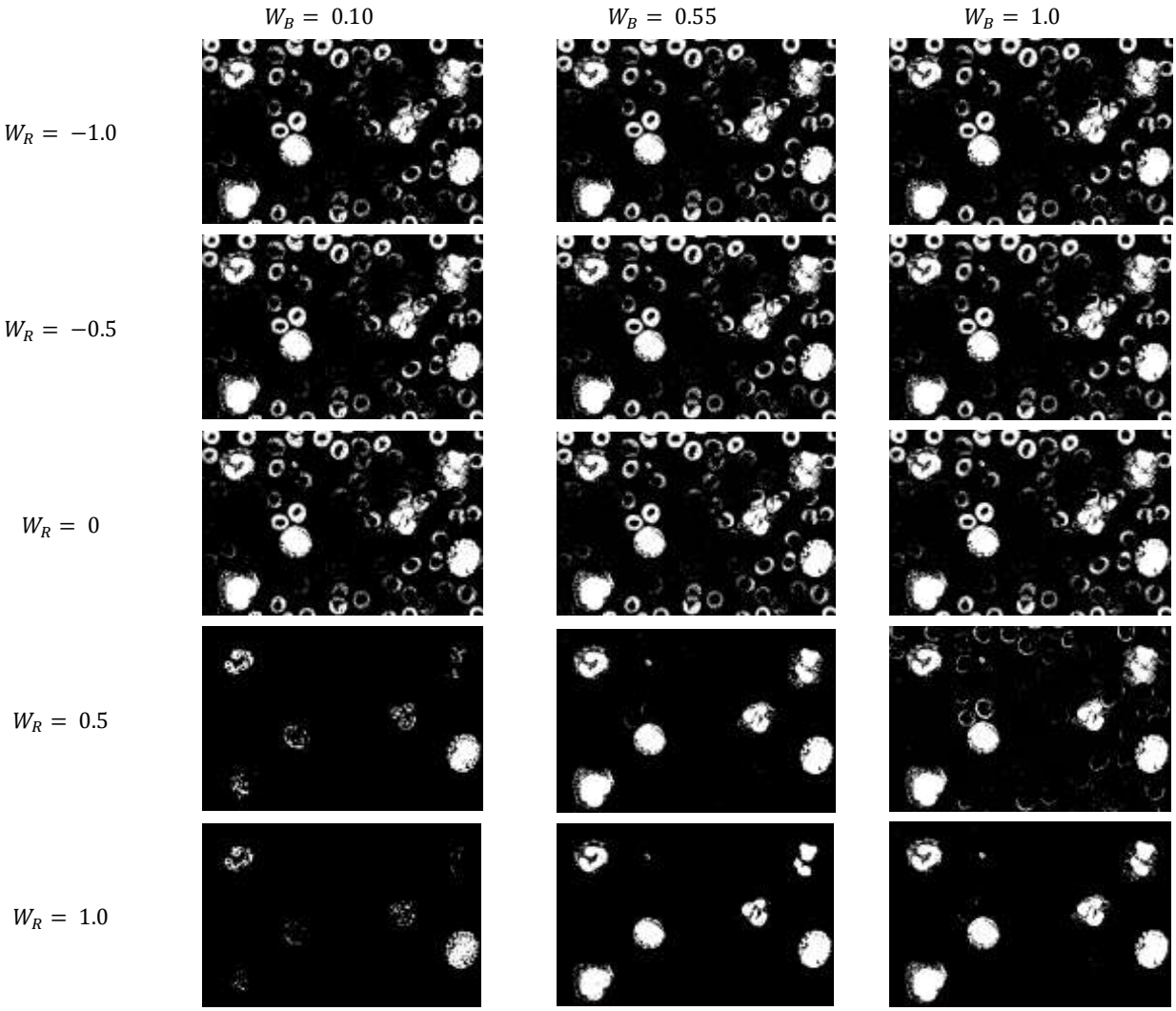

Figure 13 - Channel Weighting Procedure

Figure 13 contains several images used to determine the corresponding channel weights for the proposed algorithm. The choice of weights makes an impact on the resulting image and is crucial for the procedure. Based on the experimental procedure performed, it was decided to set $W_B = 1.0$ and $W_R = 0.75$. The resulting image after channel operations is as follows.

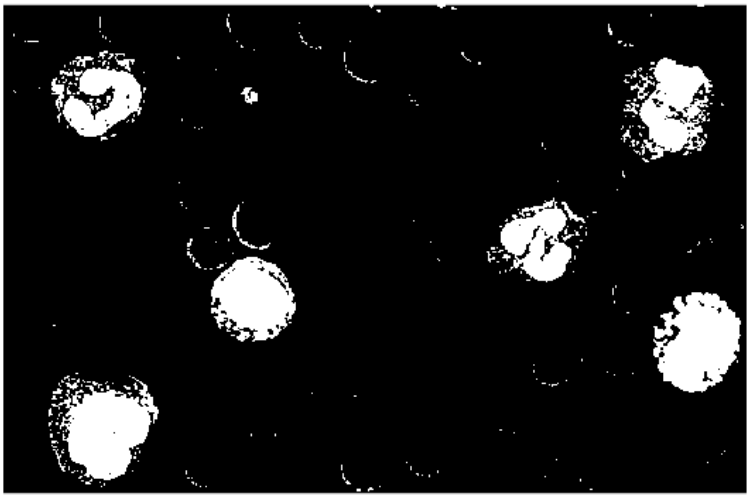

Figure 14 - Binary Image after Channel Operations

Figure 14 is the binary image after performing color channel operations. From inspection, all 6 leukocytes remain. Although the above image produces similar results to PPA, the strength of the proposed algorithm shines when using Mohamed's data set.

Once the binary image was obtained, morphological transformations were implemented to remove any unwanted small objects or small holes from the nuclei, hence both the opening and closing processes. The constructed algorithm is designed to be robust so a user may be able to adjust the structuring element sizes involved during the morphological processes. This may be changed depending on the resolution of the imported image(s) and the relative size of the WBC(s). The effects of applying morphological opening and subsequent closing are presented below.

Figure 15 - Morphological Processing

Figure 15 is the morphologically transformed binary image using the channel operations image as an input. Based on the image obtained and comparing it with Figure 10, it is seen that while both pixel islands and holes have been removed, some distinguishing characteristics of the nuclei have also been lost. Removing certain characteristics of the WBCs will be problematic when

trying to perform classification of the different WBC types. Despite the lost in characteristics, the proposed algorithm is still able to perform WBC identification. The last step of the process in the algorithm is filtering out high intensity objects based on their relative area in the image and a tuning parameter set by the user. If the high intensity object area is less than a certain area, the object is mapped to a low intensity and is not considered a WBC. Once this is done, the results are indicated on the original image and relayed back to the user. The results of the algorithm are shown below.

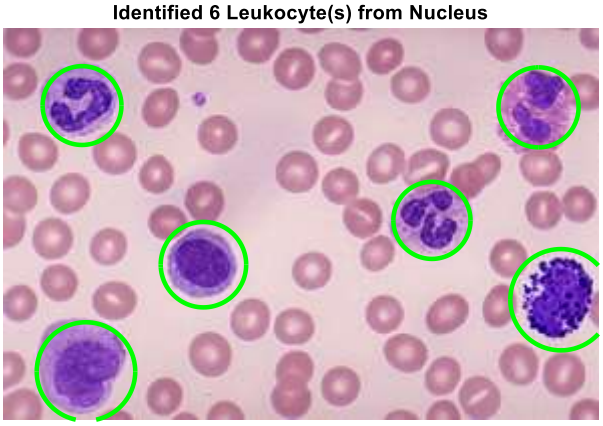

Figure 16 - WBC Identification Results

From Figure 16, the proposed algorithm is capable of performing WBC identification given an image.

After some parameter tweaking, the proposed algorithm was initially tested using several images from Bloodline and Mohamed's data set. In an effort to prevent redundancy, only the Bloodline test images are shown below with Mohamed's set being used and presented throughout the final results section.

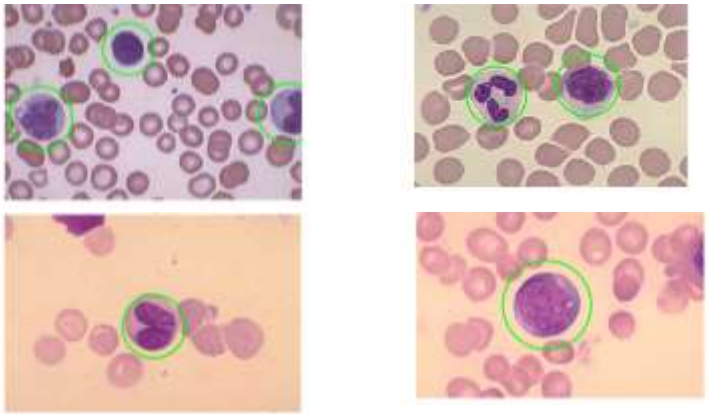

Figure 17 - Bloodline Test Results

Figure 17 demonstrates the proposed algorithm's ability to identify WBCs if a different tint is used. The above results are acceptable.

Proposed Algorithum – Cytoplasm Detection

In an effort to classify the different types of leukocytes, the proposed algorithm investigated the idea of differentiating type by entropy levels of the cytoplasm. The first step in the procedure was to identify the locations of WBCs in a given image. This process is completed by the above Nuclei Identification algorithm. Once the location of a WBC is obtained, a mask is placed around the remaining portions of the image in order to focus solely on the WBC.

Figure 18 - Blood Smear & Corresponding ROI

Figure 18 presents a blood smear image containing a WBC and the corresponding masked image with the highlighted region of interest (ROI). (Note – the cytoplasm detection algorithm was designed using Mohamed's data set and has not been evaluated with other image samples, thus the shift in the presented test image.)

After obtaining the ROI, it was decided to inspect the image in HSV color space due to the previous experimental results and analysis of the RGB color space. Changing to HSV and examining the hue channel produces the following images.

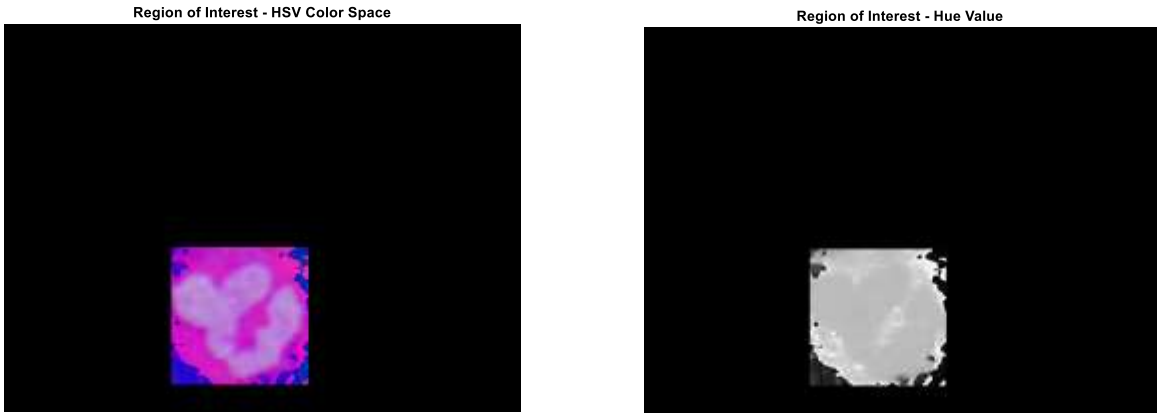

Figure 19 - HSV Color Space Analysis

Based on Figure 19, both the cytoplasm and nucleus can be easily identified in the images. The nucleus is depicted by the gray/pink colors in the left (HSV) image and by the dark gray intensities in the right (hue) image. The cytoplasm is represented by the pink/purple regions in

the HSV image and has the lighter intensities in the hue image. To gain a better understanding of the intensity distribution of the hue channel, the corresponding histogram was obtained.

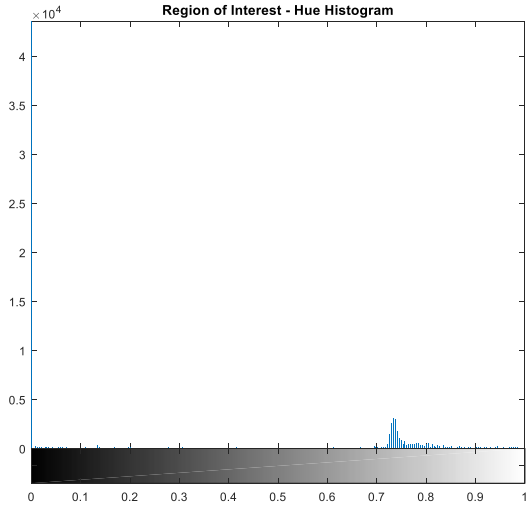

Figure 20- ROI Hue Histogram

Figure 20 is the hue channel histogram of the ROI for a WBC. The results are not surprising as the majority of pixels in Figure 19 are black. What is interesting is that the nucleus corresponds to the intensity peak around 0.73. The higher intensities represent the cytoplasm, with the highest corresponding to any red blood cells that may be featured in the ROI. Using this information, a threshold between 0.75 and 0.85 was applied to the hue channel in order to extract the cytoplasm.

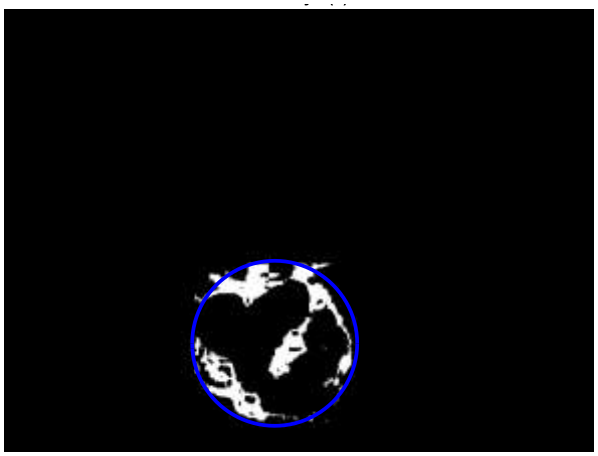

Figure 21 - Extracted WBC Cytoplasm

Figure 21 is the extracted cytoplasm of the WBC based off of the hue intensities given a region of interest. Several cytoplasm extraction results are presented below.

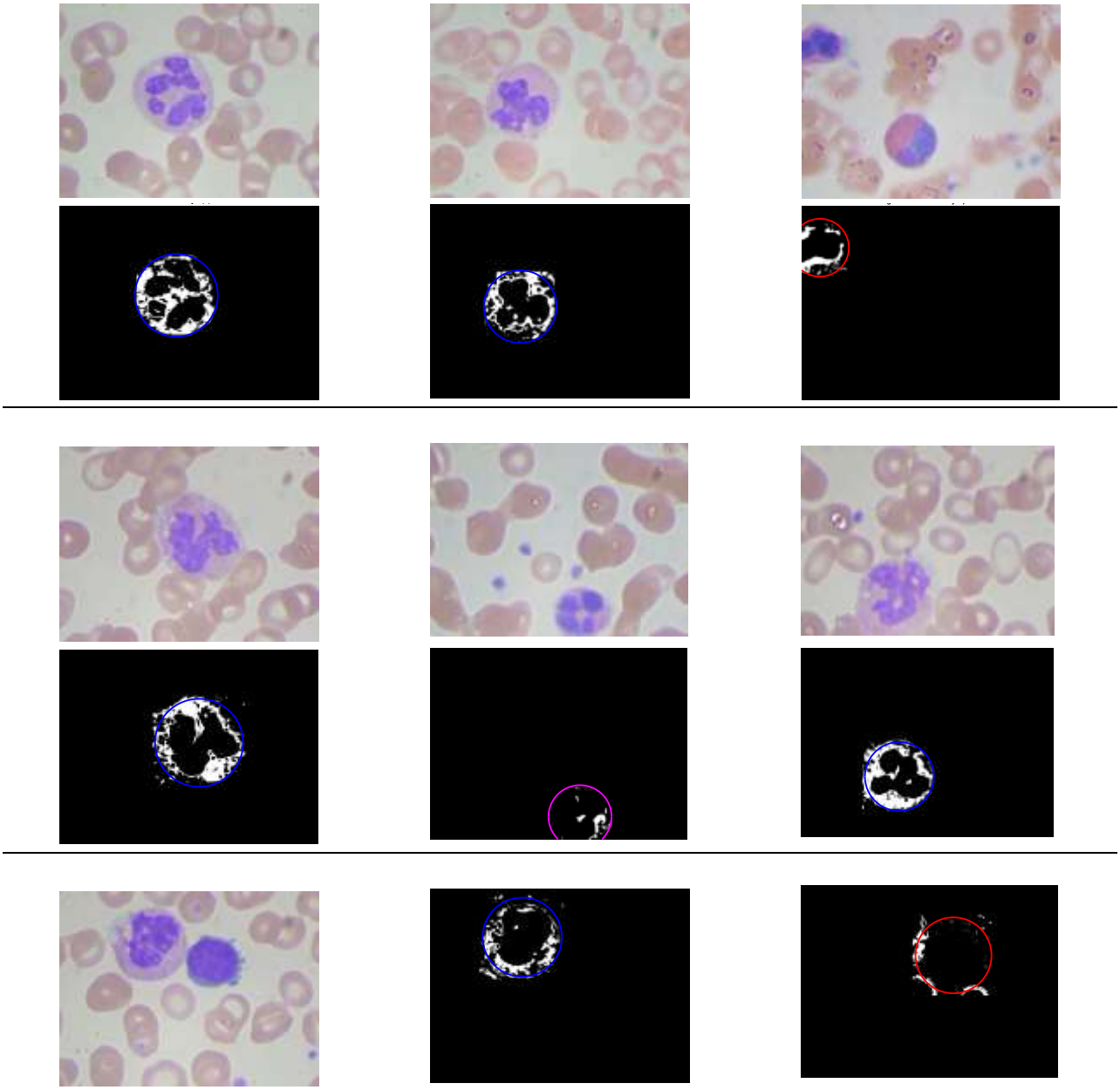

Figure 22 - Cytoplasm Extraction Results

Figure 22 contains a few cytoplasm extraction results using images from Mohamed's data set. Results indicate adequate performance, with difficulties on lighter cytoplasm shades and if there is an abundance of red blood cells in the surrounding ROI.

Using the results from the cytoplasm extraction, the entropy was calculated in an effort to classify the different WBCs. It was believed that a smaller entropy would correspond to either lymphocytes or monocytes (agranulocytes) and high entropy values to basophil, eosinophil, or neutrophil (granulocytes). However, results indicated poor performance and further study is needed to classify WBC types. This is documented in the following section.

Results & Validation

In order to validate the performance of the proposed algorithm, a random batch of 112 images from Mohamed's data set were evaluated. The evaluation of the 112 images was done manually and not by a hematologist, thus there may be several inaccuracies in the assessment of the true WBC classification. In addition, several, images from the test are presented below. The images showcase situations where the algorithm worked well and where it did not.

Depending on the entropy value calculated, different colored circles indicate the estimated WBC type; purple for low entropy, red for medium levels, and blue for high entropy. Thus, it can be interpreted that purple will indicate lymphocytes, red identifies monocytes, and blue is used for eosinophil or neutrophil (basophil may be lumped in with blue indicators). Results for the WBC classification are also stored in a data structure if a user wishes to extract any information.

Results for the 112 image sample are shown below.

Overall Performance (112 Total)		
	Count WBC Identified Correct	**Count WBC Classification Correct**
Total	109	80
Percent	97.32%	71.43%

Table 1 - Overall Algorithm Performance

Lymphocyte Accuracy (Low Entropy – Type 2)	
Total Number of Indications:	10
Correct:	9
False Positives (3 type 1, 4 type 3):	7

Table 2 - Lymphocyte Accuracy (Low Entropy)

Monocyte Accuracy (Med Entropy – Type 3)	
Total Number of Indications:	13
Correct:	5
False Positives	17

Table 3 - Monocyte Accuracy (Medium Accuracy)

Neutrophil / Eosinophil Accuracy (High Entropy – Type 1)	
Total Number of Indications:	89
Correct:	66
False Positives	7
Missed due to no Identification:	1

Table 4 - Neutrophil / Eosinophil Accuracy

Tables 1 through 4 contain the performance evaluations for the sample set. The complete results are featured in Appendix A. Based on the obtained results, the WBC identifier works great and only incorrectly identified three blood smears, achieving 97% accuracy. The WBC classification on the other hand does not perform as well. Results indicate that further work needs to be done. It is noted that the majority of the blood smears used are either eosinophil or neutrophil, which contain the highest entropy levels, and may have skewed the accuracy to be higher than what it actually is. For the high entropy levels, 66 of the 89 neutrophils/ eosinophil were correctly classified, with 7 false positives, thus achieving an acceptable performance. 9 out of 10 of the featured lymphocytes were identified. However 7 additional WBCs were declared to be lymphocytes. The large number of false positives is unacceptable, relative to the total number of lymphocytes in the sample. The worst performance was exhibited by the medium entropy values, anticipated to be monocytes. Out of the 13 total monocytes, 5 were classified correctly, with a total of 17 false positives. The large number of false positives can be attributed to the medium entropy value acting as a transition between lymphocytes and neutrophils/ eosinophils. Several images from the sample are shown below.

Positive Results:

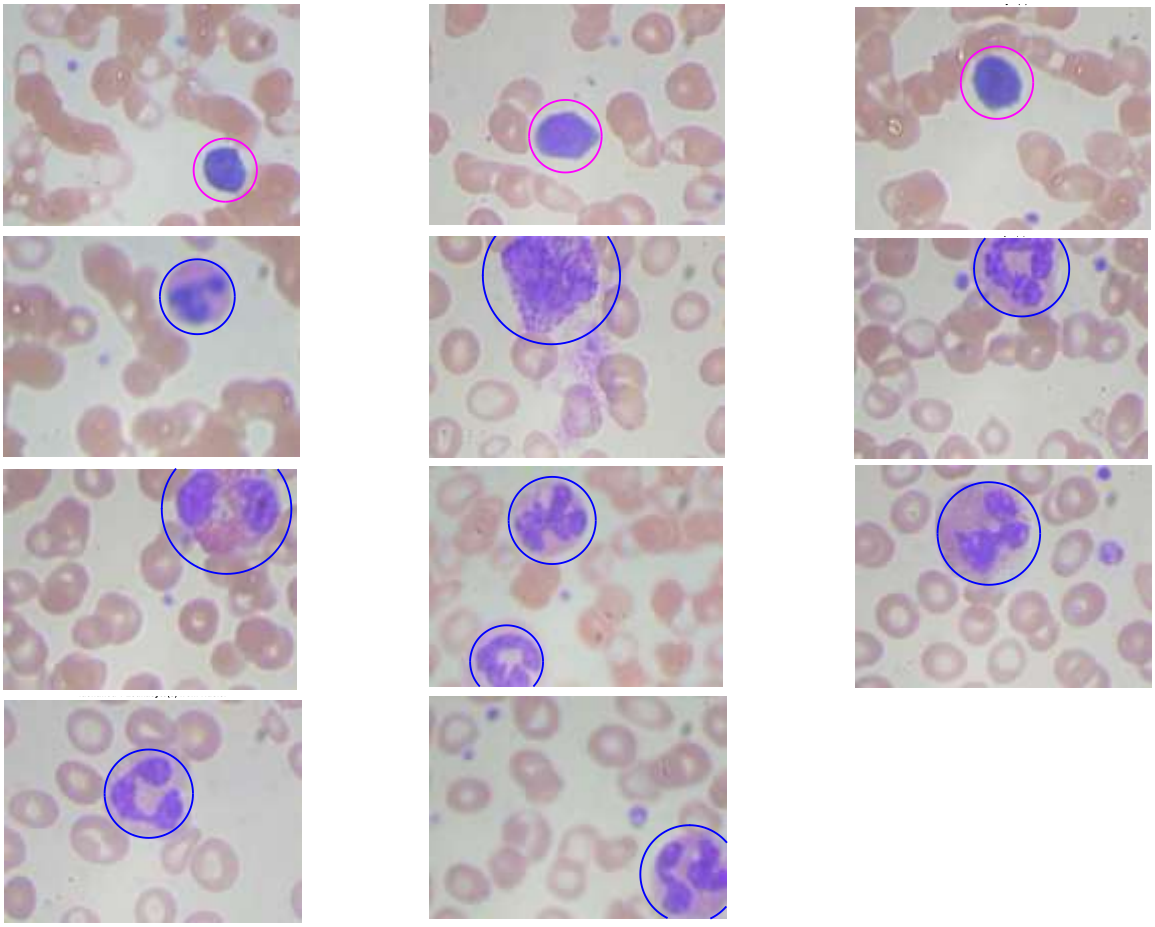

Figure 23 - Positive Results

Negative Results:

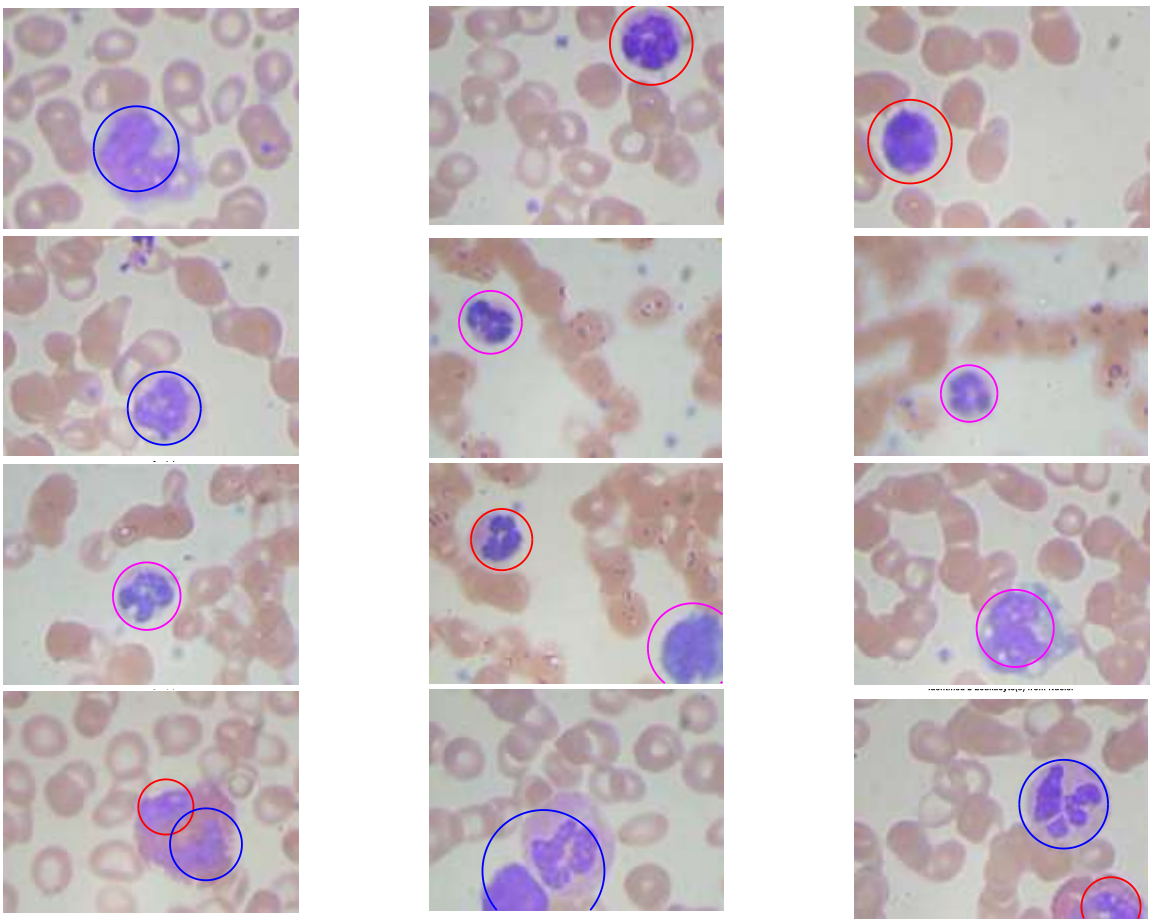

Figure 24 - Negative Results

Key: Purple = Lymphocytes, Red = Monocytes, Blue = Eosinophil / Neutrophil

Figure 23 and Figure 24 correspond to several examples where the proposed algorithm works well and not so well.

The proposed algorithm for leukocyte identification works very well. As observed in Figure 23, the algorithm is capable of identifying various types of leukocytes. Even if there multiple WBCs in the image, the algorithm is able to detect them. Where the algorithm fails (Figure 24) is when there are very large leukocytes, multiple leukocytes close together, or when the nucleus is of very low intensities. Possible fixes are addressed in the future works section.

The proposed algorithm for classification is less than satisfactory and is evident in both Figure 22 and Figure 24 . Although it is capable of classifying some WBCs, the results are too bad to consider the classification a success. There are two key components of the algorithm, the first is cytoplasm extraction and the second is entropy classification. The cytoplasm extraction is acceptable at the moment (Figure 22), but there are several issues present. As mentioned in the previous section, the cytoplasm extraction has difficulties on lighter cytoplasm shades and when

there are multiple red blood cells in the region of interest. Despite thresholding the hue value in the region of interest in order to combat these issues, the current methods are not enough. The second component to the proposed algorithm is the entropy classification. Based on the obtained results, it is not recommended to classify leukocytes on the entropy value alone. Even if cytoplasm extraction has been successful, the entropy can vary for each of the types. While the entropy may typically be lower for lymphocytes, there can also be neutrophils whose nucleus may take up a large area of the cytoplasm, resulting in low entropy values. In addition, the transition region from low to high entropy values is hazy and all types of WBCs occupy this region. Entropy of cytoplasm is not a good indicator for WBC classification.

Disscussion & Future Work

The next steps for leukocyte identification and classification would be to explore different methods of classification, as well as fix current issues with the detection algorithm. Possible fixes for leukocyte detection include exploring the watershed algorithm for when there are multiple WBCs close together, and to investigate HSV color space with the hope that it would aid in nuclei visualization. HSV color space was not applied until later on the procedure and the strong appearance of the nucleus in Figure 19 is of particular interest. It is also desired to test the identification algorithm on more blood smear images due to how well it performs.

More research would be done for the classification algorithm since current methods do not produce desirable results. Based on literature, such as in [6], common methods for identification utilize optimization and machine learning techniques, such as principle component analysis. However, there are other techniques that would be interesting to explore such as feature matching and comparing the cytoplasm regions against one another to identify differences. Since there are visible differences between the WBCs, there exists some algorithm which should be able to perform the classification.

The attempted problem of both identification and classification proved to be more difficult than anticipated. In a perfect image, each of the leukocytes can be easily identified and classified. However, blood smears vary substantially in appearance and the leukocytes are rarely clear. There is a reason why hematologists and proprietary solutions exist, in addition to identification/classification being a current research topic.

If the same project were to be redone, greater emphasis would be placed on identifying each leukocyte type, one at a time, rather than trying to first identify all possible types of leukocytes. Focus on creating the best possible algorithm for detecting type A, before trying to detect type B, and so on. At the end, a master algorithm could be applied which combines all the other detection schemes in order to perform classification. Other approaches include applying techniques such as principle component analysis.

In summary, leukocyte identification and classification was performed. The proposed algorithm for classification is less than satisfactory; however the proposed identification algorithm produces exceptional results.

References

[1] "CellaVision," 2017. [Online]. Available: http://www.cellavision.com/en/.

[2] [Online]. Available: http://library.med.utah.edu/WebPath/HEMEHTML/HEME100.html.

[3] "Identify Normal Leukocytes," 19 December 2011. [Online]. Available: http://www.pathologystudent.com/?p=4776.

[4] M. Mohamed, B. Far and A. Guaily, "An Efficient Technique for White Blood Cells Nuclei," *2012 IEEE International Conference on Systems, Man, and Cybernetics,* pp. 220-225, 14-17 October 2012.

[5] F. Sadeghian, Z. Seman, A. R. Ramli, B. H. A. Kahar and M.-I. Saripan, "A Framework for White Blood Cell Segmentation," *Biological Procedures Online,* pp. 196-206, 2009.

[6] J. Prinyakupt and C. Pluempitiwiriyawej, "Segmentation of white blood cells and comparision of cell morphology by linear and naïve Bayes classifiers," *BioMed Central,* 2015.

[7] M. Mohamed, "An Efficient Technique for White Blood Cells Nuclei Automatic Segmentation," [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/36634-an-efficient-technique-for-white-blood-cells-nuclei-automatic-segmentation.

[8] "ASH Image Bank," [Online]. Available: https://imagebank.hematology.org/atlas-images/list/.

[9] "Image Atlas Categorical Index - White Blood Cells," [Online]. Available: http://www.bloodline.net/imageatlas/white-blood-cells/.

[10] "Contrast Enhancement Techniques," [Online]. Available: https://www.mathworks.com/help/images/examples/contrast-enhancement-techniques.html.

[11] "Lab Color," Mathworks, [Online]. Available: https://www.mathworks.com/discovery/lab-color.html.

Appendices

A – 112 Mohamed Image Sample Results

File	Count	Type_1	Type_2	Cnt Correct (1=y / 0 =n)	Type(s) Correct	Actual Cnt	Actual Type_1	Actual Type_2
BloodImage_00003.jpg	1	3		1	0	1	1	
BloodImage_00014.jpg	1	2		1	0	1	1	
BloodImage_00015.jpg	1	2		1	0	1	3	
BloodImage_00018.jpg	1	3		1	1	1	3	
BloodImage_00026.jpg	1	3		1	0	1	1	
BloodImage_00028.jpg	1	1		1	1	1	1	
BloodImage_00037.jpg	1	2		1	1	1	2	
BloodImage_00038.jpg	1	2		1	1	1	2	
BloodImage_00039.jpg	1	3		1	0	1	1	
BloodImage_00043.jpg	2	3	2	1	0	2	1	3
BloodImage_00045.jpg	1	3		1	1	1	3	
BloodImage_00049.jpg	1	3		1	0	1	1	
BloodImage_00057.jpg	1	3		1	0	1	1	
BloodImage_00063.jpg	1	3		1	0	1	1	
BloodImage_00072.jpg	1	2		1	0	1	1	
BloodImage_00077.jpg	1	3		1	0	1	1	
BloodImage_00079.jpg	1	3		1	0	1	1	
BloodImage_00086.jpg	1	2		1	0	1	1	
BloodImage_00090.jpg	1	3		1	0	1	1	
BloodImage_00092.jpg	1	3		1	0	1	1	
BloodImage_00097.jpg	1	1		1	1	1	1	
BloodImage_00099.jpg	1	2		1	1	1	2	
BloodImage_00100.jpg	1	1		1	1	1	1	
BloodImage_00104.jpg	1	2		1	0	1	3	
BloodImage_00107.jpg	1	3		1	1	1	3	
BloodImage_00109.jpg	1	1		1	1	1	1	
BloodImage_00110.jpg	1	1		1	0	1	3	
BloodImage_00115.jpg	1	3		1	0	1	1	
BloodImage_00130.jpg	1	1		1	1	1	1	
BloodImage_00132.jpg	1	3		1	0	1	2	
BloodImage_00133.jpg	1	3		1	1	1	3	
BloodImage_00140.jpg	1	2		1	1	1	2	
BloodImage_00141.jpg	1	1		1	1	1	1	
BloodImage_00143.jpg	1	1		1	1	1	1	
BloodImage_00144.jpg	1	3		1	0	1	1	

BloodImage_00147.jpg	1	1		1	1	1	1	
BloodImage_00150.jpg	1	2		1	0	1	3	
BloodImage_00159.jpg	1	3		1	0	1	1	
BloodImage_00161.jpg	1	1		1	1	1	1	
BloodImage_00166.jpg	1	1		1	1	1	1	
BloodImage_00167.jpg	1	1		1	1	1	1	
BloodImage_00171.jpg	1	1		1	1	1	1	
BloodImage_00174.jpg	1	1		1	1	1	1	
BloodImage_00175.jpg	1	1		1	1	1	1	
BloodImage_00180.jpg	1	1		1	1	1	1	
BloodImage_00184.jpg	1	1		1	1	1	1	
BloodImage_00189.jpg	1	1		1	1	1	1	
BloodImage_00190.jpg	1	1		1	1	1	1	
BloodImage_00193.jpg	2	1	3	1	0	2	1	1
BloodImage_00197.jpg	1	3		1	0	1	1	
BloodImage_00199.jpg	1	1		1	1	1	1	
BloodImage_00212.jpg	1	1		1	1	1	1	
BloodImage_00214.jpg	1	1		1	1	1	1	
BloodImage_00215.jpg	1	1		1	1	1	1	
BloodImage_00216.jpg	1	1		1	1	1	1	
BloodImage_00217.jpg	1	2		1	1	1	2	
BloodImage_00218.jpg	1	1		1	1	1	1	
BloodImage_00222.jpg	1	1		1	1	1	1	
BloodImage_00223.jpg	1	1		1	1	1	1	
BloodImage_00224.jpg	1	2		1	1	1	2	
BloodImage_00236.jpg	1	1		1	1	1	1	
BloodImage_00239.jpg	1	1		1	1	1	1	
BloodImage_00240.jpg	1	1		1	1	1	1	
BloodImage_00245.jpg	1	1		1	1	1	1	
BloodImage_00247.jpg	1	1		1	0	1	3	
BloodImage_00249.jpg	2	1	1	1	1	2	1	1
BloodImage_00250.jpg	1	1		1	1	1	1	
BloodImage_00252.jpg	1	1		1	1	1	1	
BloodImage_00257.jpg	1	1		1	1	1	1	
BloodImage_00261.jpg	1	1		1	1	1	1	
BloodImage_00262.jpg	1	2		1	0	1	3	
BloodImage_00264.jpg	1	2		1	1	1	2	
BloodImage_00265.jpg	1	1		1	1	1	1	
BloodImage_00266.jpg	1	1		1	1	1	1	
BloodImage_00268.jpg	1	1		1	1	1	1	

BloodImage_00271.jpg	1	1		1	1	1	1	
BloodImage_00274.jpg	1	1		1	1	1	1	
BloodImage_00275.jpg	1	2		1	1	1	2	
BloodImage_00277.jpg	1	1		1	0	1		
BloodImage_00278.jpg	1	2		1	1	1	2	
BloodImage_00280.jpg	1	1		1	1	1	1	
BloodImage_00282.jpg	1	3		1	1	1	3	
BloodImage_00285.jpg	1	1		1	1	1	1	
BloodImage_00287.jpg	1	1		1	1	1	1	
BloodImage_00291.jpg	1	1		1	0	1	3	
BloodImage_00295.jpg	1	1		1	1	1	1	
BloodImage_00296.jpg	1	1		1	1	1	1	
BloodImage_00303.jpg	1	1		1	1	1	1	
BloodImage_00304.jpg	1	1		1	1	1	1	
BloodImage_00310.jpg	1	1		1	0	1	3	
BloodImage_00314.jpg	1	1		1	1	1	1	
BloodImage_00317.jpg	1	1		1	1	1	1	
BloodImage_00319.jpg	1	1		1	1	1	1	
BloodImage_00320.jpg	1	1		1	1	1	1	
BloodImage_00322.jpg	1	1		1	1	1	1	
BloodImage_00336.jpg	1	1		1	1	1	1	
BloodImage_00341.jpg	1	1		1	1	1	1	
BloodImage_00345.jpg	1	1		1	1	1	1	
BloodImage_00346.jpg	1	1		1	1	1	1	
BloodImage_00357.jpg	1	1		1	1	1	1	
BloodImage_00359.jpg	1	1		1	1	1	1	
BloodImage_00376.jpg	1	1		1	1	1	1	
BloodImage_00382.jpg	1	1		1	1	1	1	
BloodImage_00385.jpg	1	1		1	1	1	1	
BloodImage_00387.jpg	1	1		1	1	1	1	
BloodImage_00388.jpg	1	1		1	1	1	1	
BloodImage_00393.jpg	1	1		1	1	1	1	
BloodImage_00397.jpg	1	1		1	1	1	1	
BloodImage_00407.jpg	1	1		1	1	1	1	
BloodImage_00000.jpg	0	0		0	0	1	1	
BloodImage_00313.jpg	1	1		0	0	2	2	1
BloodImage_00353.jpg	2	3	1	0	0	1	1	

B – wbcNucleiIdentification.m

```
function [nuclei, wbcType] = wbcNucleiIdentification(im, OP, CL, tune, ty)
%
% wbcNuclei: Identify Leukocytes based on nulcei.
%
% INPUT:    im - image
%           OP - opening structuring element size
%           CL - closing structuring element size
%           tune - parameter related to area limit and resolution
% OUTPUT:   [] - figures
%           nuclei - number of WBC's identified in image
%           wbcType - vector returning type of each WBC identified
%

srgb2lab = makecform('srgb2lab');
lab2srgb = makecform('lab2srgb');
imlab = applycform(im, srgb2lab); % convert to L*a*b*
max_luminosity = 100;
L = imlab(:,:,1)/max_luminosity;

% contrast stretching with luminance
imlab_adjust = imlab;
imlab_adjust(:,:,1) = imadjust(L)*max_luminosity;
imlab_adjust = applycform(imlab_adjust, lab2srgb);
R = imlab_adjust(:,:,1); % red channel
G = imlab_adjust(:,:,2); % green channel
B = imlab_adjust(:,:,3); % blue channel
nuclBW = ((1*B)-(0.75*R))./(G); % result will be capped at 0 or 1 (255)
level = 150;
nuclBW = nuclBW > level; % preventative measure

sqOpen = strel('disk',OP);
nuclMorph = imopen(nuclBW,sqOpen);
sqClose = strel('disk', CL);
bw = imclose(nuclMorph,sqClose);
r = regionprops(logical(bw)); % obtain properties of the resulting image
% display the original image with the thresheld cross-corelation marked
figure()
imshow(im,[])
[m, n] = size(bw);
cnt = 0; % number of WBCs
type = 0; % classification for one WBC
wbcType = 0; % classifcation vector to be returned to user

hold on
for i = 1:length(r)
    if r(i).Area>((m*n)/tune) % relative area threshold for WBC
        z = 3.*sqrt(r(i).Area/pi()); % identification circle
        col = 'g';
        % region of interest bounds
        A = round(r(i).Centroid(2)-z/2);
        B = round((r(i).Centroid(2)-z/2)+z);
        C = round(r(i).Centroid(1)-z/2);
        D = round((r(i).Centroid(1)-z/2)+z);
        if ty == 1 % if classification is turned on
            % WBC classification based on entropy
            type = cytoplasmClassification(im, A, B, C, D);
            if type == 1
                col = 'b'; % either Neutrophil or Eosinophils
            elseif type == 2
                col = 'm'; % Lymphocyte
            elseif type == 3
                col = 'r'; % Monocyte
            end
        end
        rectangle('position', [C,A,z,z],...
            'Curvature',[1 1],'EdgeColor',col,'LineWidth',2.5)
        cnt = cnt + 1;
        if type > 0
            wbcType = horzcat(wbcType,type);
        end
    end
end
title(['Identified ',num2str(cnt), ' Leukocyte(s) from Nucleus'])
% return information to user / output
nuclei = cnt;
if length(wbcType) > 1
    wbcType(1) = []; % remove 0 if classification is on
end
end
```

C – cytoplasmClassification.m

```
function type = cytoplasmClassification(im, A, B, C, D)
%
% cytoplasmClassification: Identify WBC type based on cytoplasm.
% Note: not robust. Works well for Mohamed data set.
% May uncomment figures if desired.
%
% INPUT:    im - image (note: must be color, NOT grayscale)
%           A,B,C,D - region of interest bounds
% OUTPUT:   type - type of wbc identified
%           [] - figures (if desired)
%

[m, n, o] = size(im);

% region of interest mask
mask = uint8(zeros(m,n,o));
bnd = [A+5,B-5,C+5,D-5]; % adjust window if needed.
bnd(bnd<1) = 1;

% consrtuct region of interst
mask(bnd(1):bnd(2),bnd(3):bnd(4),:) = 1;
mask = mask(1:m,1:n,:);
ROI = mask.*im; % region of interest
% figure; imshow(ROI); title('Region of Interest')

% convert to HSV, threshold hue to get cytoplasm
imgh = rgb2hsv(ROI);
% figure; imshow(imgh); title('Region of Interest - HSV Color Space')
imgh = imgh(:,:,1);
% figure; imshow(imgh); title('Region of Interest - Hue Value')
% figure(); imhist(imgh); title('Region of Interest - Hue Histogram')
cyto = (imgh>0.74 & imgh<0.81); % edit these for different data sets
% figure; imshow(cyto); title('Region of Interest - Cytoplasm')

E = entropy(cyto); % entropy of cytoplasm
type = 0;

% classify type of WBC based off of cyoplasm entropy
if E<=0.047
    type = 2; % Lymphocyte or Monocyte
elseif 0.047<E && E<=0.1
    type = 3; % Possible Monocyte or Neutrophil
elseif 0.06<E
    type = 1; % either Neutrophil or Eosinophils
end

% uncomment these to return entropy information to user
% disp('Entropy')
% disp(E)
% disp(type)
% fprintf('\n')

end
```