International Journal of Research in Engineering, Science and Management
Volume 4, Issue 12, December 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Blood Cell Segmentation and Classification by Machine Learning

Md. Haris Uddin Sharif[1*], K. Maroti Yamaguchi[2], Shaamim Udding Ahmed[3]
[1]*University of the Cumberland, Williamsburg, Kentucky, United States of America*
[2,3]*Strayer University, Maryland, United States of America*

***Abstract*: The immune system is the third defensive line of the human body which defends the body from viruses, bacteria and pathogens. This natural protection detects and kills defective cells like the tumor cells. The immune system contains immune organs, immune and immune cells. The essential constituent of immune cells is white blood cells (WBCs), and they are an essential part of our body immunity. The light disperses system theory is used by automatic machines to compute red blood cells (RBCs) and WBCs. WBCs can be classified into five distinct types: basophils, eosinophil, neutrophils, lymphocytes and monocytes. They are granular in the first three forms and non-granular in the second two. It was ineffective to differentiate these types through the use of the light dispersal process. In this paper we will provide data that will consist images and .xml file of each image then system will provide us the information of image. We will segment out the blood cells images based on .xml file information. Also, will extract the features of blood cells depending upon the nature of image. To extract the color features will use Grid Color Moment Algorithm, for texture features Local Binary Patterns Algorithm, for classification will use K Nearest Neighbors Algorithm. Finally, will calculate the accuracy from our given data.**

***Keywords*: White blood cell, Segmentation technique, Cell segmentation, Machine Learning algorithms, Image processing.**

1. Introduction

In this experiment we were provided with a data that consisted of 4889 images and also with .xml file of each image that gives us the information of that image. We had to segment out the blood cells images of different types (includes RBC, WBC and Platelets), from each image using that information from .xml file in the form of annotation and then we had to extract the features of those blood cells depending upon the nature of our image. So considering the nature of our images, we extracted the color features using “Grid Color Moment Algorithm”, and Texture features using “Local Binary Patterns Algorithm”. And then we did classification using KNN (K Nearest Neighbors) Algorithm. And then we concatenated all the features of Color Moment and LBP and applied KNN and then we calculated the accuracy and plotted the confusion matrix for all our given data.

The analyses are done using pictures taken in a hospital. MATLAB is used to implement this. The WBCs are classified for each image by the hematology specialist and registered for database build [7]. The procedure is done on separate blood cell images to test the effects of the techniques. The picture of blood cells comprises RBCs, WBCs and platelets [6]. The WBC alone is segmented from these and the number of WBCs found using different techniques is compared with the actual number in the manually collected image. The pathologist extracts the essential truth about the entire dataset.

We have used image recognition that can provide human evaluations. This approach uses high-resolution and scalable computer vision extracted features to maximize performance and precision in operation. Researchers now work on creating a device which is able to classify WBCs automatically with blood photographs [5]. However, correct WBC classification is also a concern prior to classification. Function extraction is a special variety of dimensional reduction in image segmentation [8]. If the information in an algorithm is too large to process, the input information is converted into a reduced representational set of characteristics. Input data is translated into a sequence of functionalities called extraction of functions [4].

Our report consists of methodology that we applied algorithms that we used for the calculation of image features flow diagram, tables, graphs and results.

Counting blood cells is a major field of biotechnology [9]. Many approaches are tested and used to achieve improved outcomes in cases involving human blood cell segmentation. [2] Introduces an approach during the blood smear test to count multiple blood cells. The most prevalent form of cancer in Canadian adults is chronic lymphatic leukemia (CLC). The aim of [3] is to decrease the total watershed algorithm section error by eliminating 1% of local minimum values.

2. Methodology-Dataset

A. *DataSet- JPEG File*

The data consisted of 364 annotation files (.xml format) and images of blood cells. This data set will be used to train and test our model.

B. *DataSet- XML File*

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note

*Corresponding author: haris.uddin.sharif@gmail.com peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

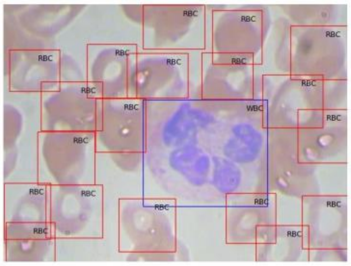

Fig. 1. JPEG file

3. Images Labeling Using Annotations

Using annotations, we labeled each image from 364 images file with different types of cell types as we had Xmin, Xmax, Ymin and Ymax for each blood cell type in annotation.

We have three kinds of labels as shows on below image.

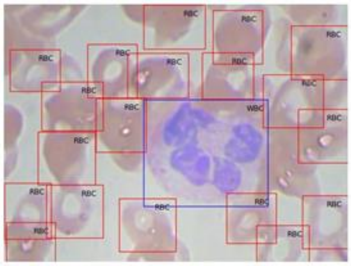

Fig. 2. Three kinds of labels

- RBC (Red Blood Cell)
- WBC (White Blood Cell)
- Platelets

A. *Segmentation*

Just crop the cells from each image of 364 images file using annotations information and forward those segmented images towards feature extraction for deep learning and classification, for the reason to fulfill the requirements of our machine learning model.

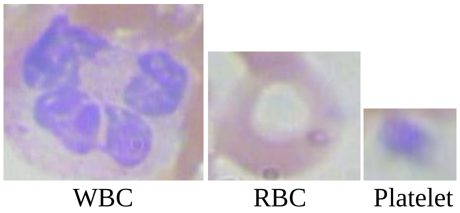

Fig. 3. Segmentation

After segmentation of 364 images now we have 4889 images of these above labels.

4. Features Extraction

Feature extraction is a dimensionality method that an original set of raw data is reduced to other flexible combinations for processing. A characteristic of these vast data sets is multiple variables that need many computing sources to process. Feature extraction is the title for systems that select and consolidate variables into features, effectively decreasing the amount of data that can be processed while still wholly and accurately representing the primary data set. However below have few steps of extraction.

A. *Color Features (Step 01)*

To extract the color features we used Grid Color Moment Algorithm:

- Convert the image from RGB to HSV color space
- Uniformly divide the image into 3x3 blocks

$$x' = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- For each of these nine blocks
- Computed its mean color (H/S/V)

Where N is the number of pixels within each block, xi is the pixel intensity in H/S/V channels.

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{N}(x_i - x')^3}{\left(\frac{1}{n}\sum_{i=1}^{N}(x_i - x')^2\right)^{3/2}}$$

- Computed its variance (H/S/V)
- Computed its skewness (H/S/V)
- Each block had 3+3+3=9 features, and thus the entire image had 9x9=81 features. Before we use KNN to train the classifier, we first needed to normalize the 81 features to be within the same range, in order to achieve good numerical behavior. To do the normalization, for each of the 81 features.
- Computed the mean and standard deviation from the training dataset.

$$\mu = \frac{1}{M}\sum_{i=1}^{M} f_i$$

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=0}^{M}(f_i - \mu)^2}$$

Where M is the number of images in the training dataset and fi is the feature of the ith training sample.

- Perform the "whitening" transform for all the data (including both the training data and the testing data), and get the normalized feature value.

B. *Local Binary Patterns- (Step 02)*

To extract the texture features we used Local Binary Patterns Algorithm- Local Binary Pattern (LBP) is an algorithm or a type of operator that labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number

- Read image in grey scale
- Padding of size one
- Sliced 3x3 portion of image
- Applied threshold by considering the center value of that 3x3 portion as a threshold value.

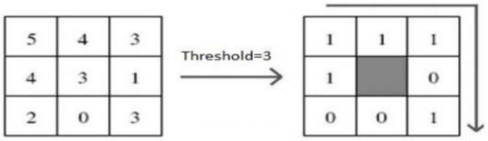

Fig. 4.

- And then append these 8 neighbors to an array
- Converted these neighbors binary to decimal
- Computed the sum of this array
- Replace the center value with this computed value
- And then slide throughout the image with stride = 1
- Compute the histogram of the texture image after LBP
- So we got a feature vector of 256 feature values
- Normalized these 256 features
- Computed the mean and standard deviation from the training dataset.

$$\mu = \frac{1}{M}\sum_{i=1}^{M} f_i$$

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=0}^{M}(f_i - \mu)^2}$$

Where M is the number of images in the training dataset and fi is the feature of the ith training sample.

- Perform the "whitening" transform for all the data (including both the training data and the testing data), and get the normalized feature value, as shown below,

$$f'_i = \frac{f_i - \mu}{\sigma}$$

C. *Concatenated the Features (Step # 03)*

In step three, concatenate the date and make a column for label with respect to each image.

$$f'_i = \frac{f_i - \mu}{\sigma}$$

D. *F-Nearest Neighbors (KNN) Algorithm*

The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. With respect to this project, we have to do classification on images and predict the result using the test data.

- Feature extraction will be done
- Read features data file with labels
- Randomize the data
- Split the data into test and train data
- And find the distance of each test file features from train features of all images
- Sorted the distances for each test image file
- Set the value of k and the took k distances and their respective labels from sorted distances list and labels list respectively
- Set the predicted label equal to maximum repeated label in that array of k selected labels
- And then find the accuracy by using the Predicted and True Labels.

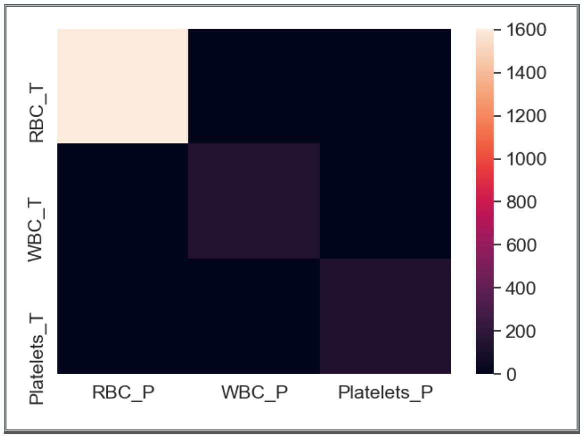

Fig. 5. Confusion matrix plot for color features

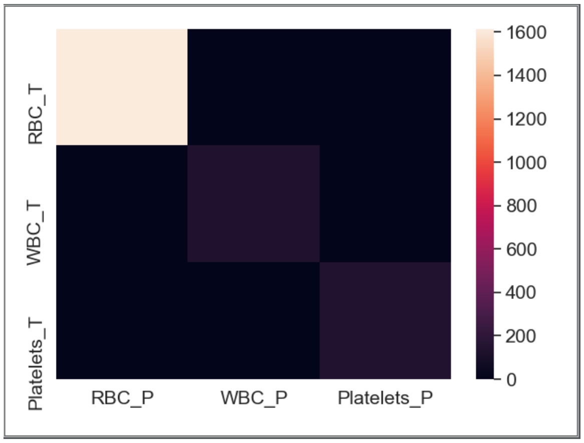

Fig. 6. Confusion matrix plot for LBP features

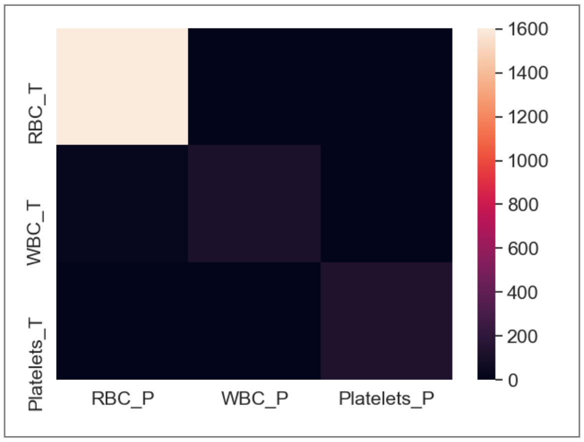

Fig. 7. Confusion matrix plot for all concatenated features of LBP & grid color moment

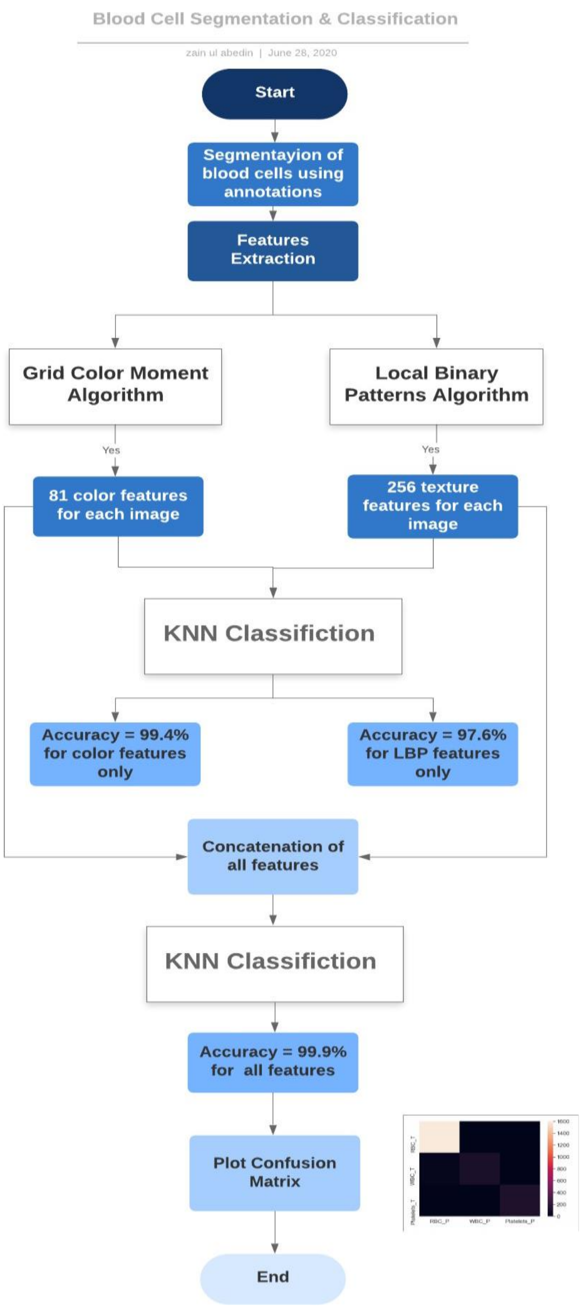
Blood Cell Segmentation & Classification
zain ul abedin | June 28, 2020
Start
Segmentayion of blood cells using annotations
Features Extraction
Grid Color Moment Algorithm
Local Binary Patterns Algorithm
Yes
Yes
81 color features for each image
256 texture features for each image
KNN Classifiction
Accuracy = 99.4% for color features only
Accuracy = 97.6% for LBP features only
Concatenation of all features
KNN Classifiction
Accuracy = 99.9% for all features
Plot Confusion Matrix
RBC_T
WBC_T
Platelets_T
RBC_P
WBC_P
Platelets_P
1600
1400
1200
1000
800
600
400
200
0
End

Fig. 8. Flow diagram

5. Result

Table 1
Result

	Color Features	LBP Features	ALL
Accuracy	99.4%	97.6%	99.9%

6. Conclusion

As we can see the conclusion in results that the feature extraction depends upon the nature of images we used for classification. And also, Color Features are showing a great role in this deep learning and we can easily assign a label to an unseen sample of blood cell by calculating the color features and LBP features as by using these our accuracy is 99.9%. So, in the end we will conclude that you should find features according to the nature of your image.

References

[1] Mandloi, G (2014), "A Survey on Feature Extraction Techniques for Color Images."

[2] Tulsani H. 2013. Segmentation using morphological watershed transformation for counting blood cells. IJCAIT. 2:28–36.

[3] Mohammed EA, Mohamed MMA, Naugler C, Far BH. 2013. Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding. In Electrical and Computer Engineering (CCECE), 2013 26th Annual IEEE Canadian Conference on, pp. 1–5.

[4] Algamili, A. S. (2016). Red blood cell segmentation and classification method using Matlab.

[5] Ortuño, F., & Rojas, I. (2015). *undefined*. Springer.

[6] Ramoser, H., Laurain, V., Bischof, H., & Ecker, R. (2005). Leukocyte segmentation and classification in blood-smear images. *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*.

[7] Safuan, S. N., Tomari, R., Zakaria, W. N., & Othman, N. (2017). White blood cell counting analysis of blood smear images using various segmentation strategies. https://doi.org/10.1063/1.5002036

[8] Sholeh, F. I. (2013). White blood cell segmentation for fresh blood smear images. *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*.

[9] Y G., & Ashour, A. S. (n.d.). Neutrosophic set in medical image analysis. Infinite Study.