

Machine Learning and Image Processing Approach of Automatic Identification and Counting of Blood Components

Aishwarya Jonganavar¹, Apeksha Avane², Apeksha Siddannavar³, Ashwini Patil⁴

¹Aishwarya Jonganavar, Computer Science & Engineering, Angadi Institute of Technology & Management, Belagavi, Karnataka, India

²Apeksha Avane, Computer Science & Engineering, Angadi Institute of Technology & Management, Belagavi, Karnataka, India

³Apeksha Siddannavar, Computer Science & Engineering, Angadi Institute of Technology & Management, Belagavi, Karnataka, India

⁴Ashwini Patil, Computer Science & Engineering, Angadi Institute of Technology & Management, Belagavi, Karnataka, India

⁵Prof. Priyanka Pujari, Department of Computer Science & Engineering, Angadi Institute of Technology & Management, Belagavi, Karnataka, India

Abstract - A complete blood cell count is an important test in medical diagnosis to evaluate overall health condition. Traditionally blood cells are counted manually using haemocytometer along with other laboratory equipment's and chemical compounds, which is a time-consuming and tedious task. In this work, we have used a machine learning approach for automatic identification and counting of three types of blood cells using Convolutional Neural Network (CNN). The primary role of this work is the accurate segmentation of cytological images of blood smears and subsequent feature extraction. Our proposed approach benefits from powerful image processing techniques to perform complete blood cell count without human intervention.

Key Words: CNN, haemocytometer, cytological image, blood smears, feature extraction

1. INTRODUCTION

A complete blood cell (CBC) count is an important test often requested by medical professionals to evaluate health condition. The main three types of cells that constitute blood are red blood cells (RBCs), white blood cells (WBCs), and platelets. RBCs also known as erythrocytes are the most common type of blood cell, which consists of 40–45% of blood cells. Platelets also known as thrombocytes are also huge number in blood. WBCs also known as leukocytes, are just 1% of total blood cells. RBCs carry oxygen to our body tissues and the amount of oxygen tissues received is affected by the number of RBCs. WBCs fight against infections and platelets help with blood clotting. As these blood cells are huge in number, traditional manual blood cell counting system using haemocytometer is highly time consuming and erroneous and most of the cases accuracy vastly depends on the skills of a clinical laboratory analyst. Therefore, an automated process to count different blood cells from a smear image will greatly facilitate the entire counting process.

In this work, we develop a system for automatic detection and classification of cytological images which plays an increasing important role in medical diagnosis. The primary aim of this work is the accurate segmentation of cytological images of blood smears and subsequent feature extraction. Our proposed approach benefits from powerful image processing techniques to perform complete blood count (CBC) without human intervention. The general framework in this blood smear analysis research is as follows. Firstly, a digital blood smear image is de-noised using optimized Bayesian non-local means filter to design a dependable cell counting system that may be used under different image capture conditions. Then an edge preservation technique with Kuwahara filter is used to recover degraded and blurred white, red blood cell boundaries in blood smear images while reducing the residual negative effect of noise in images. After denoising and edge enhancement, the next step is binarization using Otsu to separate the cells and stained background. Cells separation and counting is achieved by granulometry. Using three main types of features: shape, intensity and texture this work is examined using Convolutional Neural Networks in connection with white, red blood cell classification. These experimental results are also assessed with ground truth manually obtained from medical experts.

2. SYSTEM ARCHITECTURE

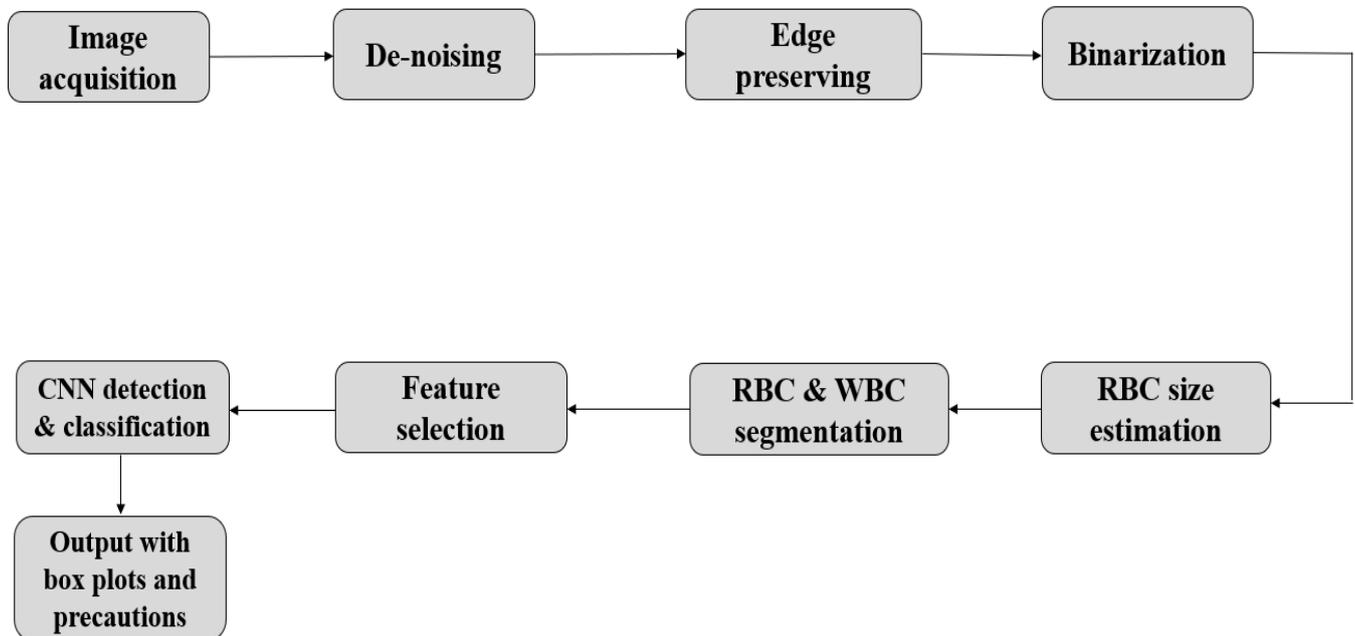


Fig -1: System Architecture

On continuing discussion concerning the methodologies used, the normal blood images are saved in JPEG format. Then a key step is to choose a proper gray scale channel to maintain the high and low frequency of components in each blood image. The method used for denoising is based on the Bayesian nonlocal means filter. In a comparative study with other state-of-the-artwork Bayesian nonlocal mean brings the highest PSNR value. To build better boundaries for white blood cells and also to replace white with red blood cell, Kuwahara filter is addressed. Then, a binarization technique is introduced by merging with the Otsu method. Area-Granulometry is used to estimate RBC size. Further in feature selection, three features such as shape, intensity and texture are selected for detection and classification. The objective of CNN research in this case-study is to determine whether CNNs can be good predictors in blood classification with few available sample data.

3. METHODOLOGY

Important steps to be performed on images:

Color Selection:

Some previous published work used the *green* channel of the RGB color encoding to analyze blood image data. Also, white blood cell granular cytoplasm pixels can be highlighted better in the image histogram of the green channel. Several other color spaces rather than RGB have been addressed in literature for different specific purposes. Several attempts have been made to use gray scale intensity of colorful JPEG blood smear images.

Noise Removal:

Many efforts have been devoted to reducing this undesired effect. In this work, we had used Bayesian non-local means filter for de-noising because it gets the highest PSNR value. Noise removal helps to stabilize the next steps to achieve accurate localizations or parametric estimations.

Blood Smear Image Enhancement

Image quality can interfere with the cell border tracking and local information. Therefore, image pre-processing is an important phase of the segmentation procedure. It includes steps to capture a digital image and then remove Gaussian noise of

blood smear. It also includes enhancement techniques of image smoothing, edge preserving and background subtraction, which allow more efficient data analysis.

Blood Image Pre-Processing

Image acquisition is the action of retrieving raw images from a capturing source, usually a digital camera. Storing raw files into computerized image format as we have all experienced, is an inseparable part of camera shots. Different electronic file formats are available for images. Each format stores the image in a specific way. The most common image file formats found are: Graphics Interchange Format(.GIF), Joint Photographic Experts Group (.JPG), Portable Network Graphics (.PNG).

BAYESIAN NON-LOCAL MEANS FILTER

Consider a gray-scale image $z = (z(x))_{x \in \Omega}$ defined over a bounded domain $\Omega \subset \mathbb{R}^2$, (which is usually a rectangle) and $z(x) \in \mathbb{R}^+$ is the noisy observed intensity at pixel $x \in \Omega$. The NL-means filter is defined as,

$$NL z(x) = \frac{1}{C(x)} \sum_{y \in \Omega} w(x, y) z(y)$$

where $NL z(x)$ at pixel x is the weighted average of all gray values in the image and $C(x)$ is a normalizing factor, i.e.

$$C(x) = \sum_{y \in \Omega} w(x, y).$$

The weights $w(x, y)$ defined as,

$$w(x, y) = \exp\left(-\frac{1}{h^2} \int_{\mathbb{R}^2} G_a(t) |z(x+t) - z(y+t)|^2 dt\right) := \exp\left(-\frac{\|z(x) - z(y)\|_{2,a}^2}{h^2}\right) \quad (2)$$

KUWAHARA FILTER

The Kuwahara filter is a non-linear smoothing filter used in image processing for adaptive noise reduction. Most filters that are used for image smoothing are linear low-pass filters that effectively reduce noise but also blur out the edges. However, the Kuwahara filter is able to apply smoothing on the image while preserving the edges.

OTSU'S METHOD

In computer vision and image processing, Otsu's method is used to perform automatic image thresholding. In the simplest form, the algorithm returns a single intensity threshold that separate pixels into two classes, foreground and original image background. This threshold is determined by minimizing intra-class intensity variance, or equivalently, by maximizing inter-class variance. Otsu's method is a one-dimensional discrete analog of Fisher's Discriminant Analysis, is related to Jenks optimization method, and is equivalent to a globally optimal k-means performed on the intensity histogram.

1. Compute histogram and probabilities of each intensity level
2. Set up initial $\omega_i(0)$ and $\mu_i(0)$
3. Step through all possible thresholds $t = 1 \dots$ maximum intensity
 1. Update ω_i and μ_i
 2. Compute $\sigma_b^2(t)$
4. Desired threshold corresponds to the maximum $\sigma_b^2(t)$
5. You can compute two maxima (and two corresponding thresholds). $\sigma_{b1}^2(t)$ is the greater max and $\sigma_{b2}^2(t)$ is the greater or equal maximum
6. Desired threshold =
$$\frac{\text{threshold}_1 + \text{threshold}_2}{2}$$

Fig -2: Otsu's Algorithm

CONVOLUTIONAL NEURAL NETWORKS (CNN)

A Convolutional Neural Network is a Deep Learning algorithm which can take in an input image, assign importance (weights) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.

In a regular neural network, there are three types of layers:

1. **Input layers:** It's the layer in which we give input to our model. The number of neurons in this layer is equal to total number of pixels in our image.
2. **Hidden layers:** The input from Input layer is then fed into hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layer can have different number of neurons which are generally greater than the number of features.
3. **Output layers:** The output from the hidden layer is then fed into a logistic function like sigmoid which converts the output of each class into probability score of each class.

CNN is composed of two major parts:

- **Feature Extraction:** In this part, the network will perform a series of convolutions and pooling operations during which the features are detected.
- **Classification:** Here, the fully connected layers will serve as a classifier on top of these extracted features. They will assign a probability for the object on the image being what the algorithm predicts it is.

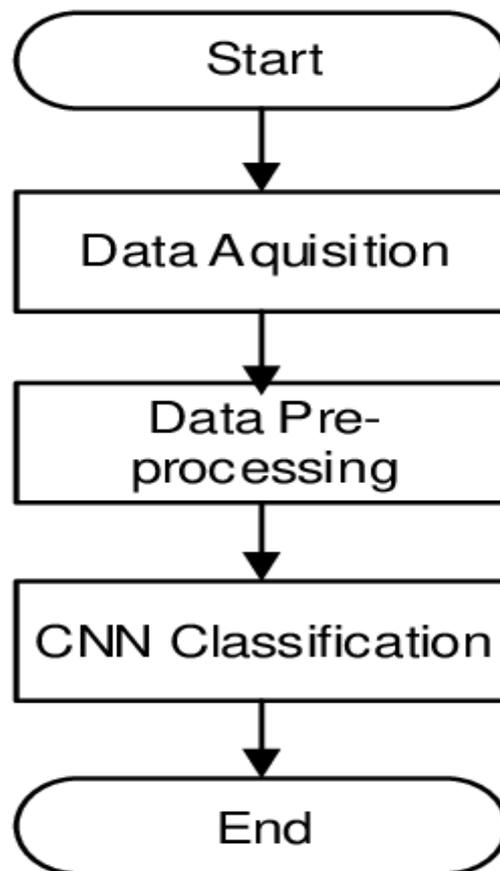


Fig -3: Dataflow diagram for CNN classification

4. RESULTS

With the proposed method, we automatically identify and count RBC's, WBC's and Platelets. The method is fast and accurate as CNN algorithm has been used which has higher efficiency. We tested our model using a test dataset. Noise removal helps to stabilize the next steps to achieve accurate positions of the cells. Pre-processing is continued by edge enhancement which maintains better white blood cell boundaries appearance. It also facilitates to get an effective visual appearance and helps in feature selection. The boxplots will help the users to easily count the cells. Also, the precautions mentioned will help the users to maintain their health accordingly.

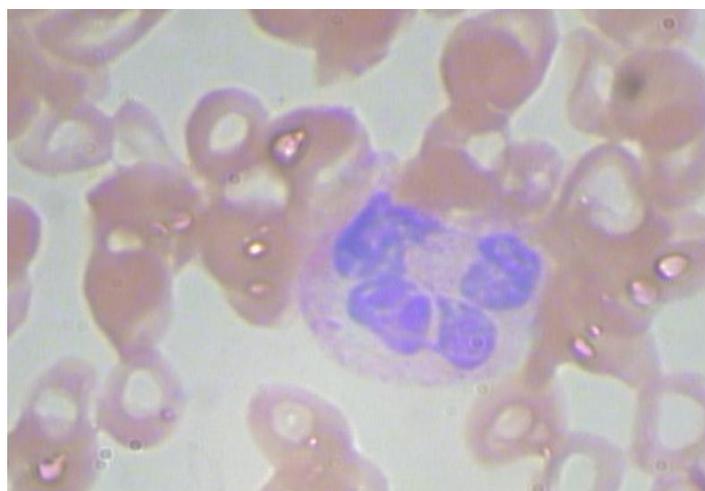


Fig -4: Image collected from Kaggle website

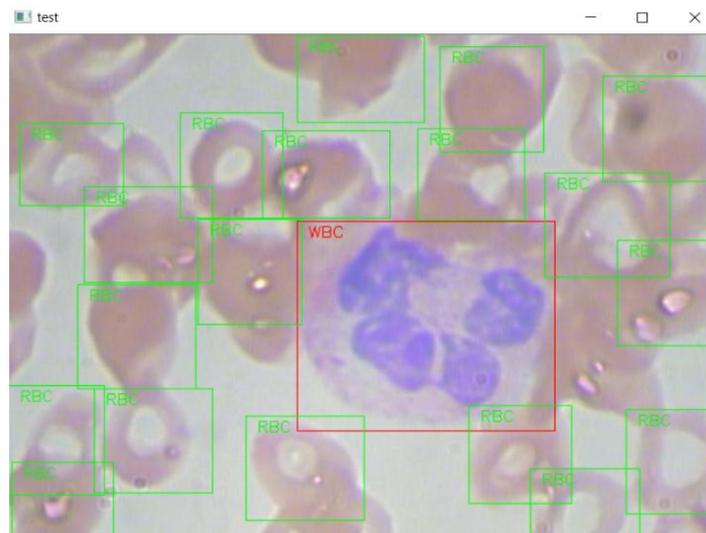


Fig -5: Boxplots to differentiate between RBC's, WBC's, and Platelets

```
(base) C:\Users\User>c:
(base) C:\Users\User>cd/
(base) C:\>cd/
(base) C:\>cd Project
(base) C:\Project>cd Blood Cell
(base) C:\Project\Blood Cell>cd scripts
(base) C:\Project\Blood Cell\scripts>python visualize.py
Precautions Need to be taken

Take your temperature by mouth four times each day. Call your oncologist if your oral temperature is above 100.5o F.

Eliminate uncooked foods, which may contain germs, including:
1)raw meats or fish salads
2)natural cheeses
3)uncooked eggs
4)fresh fruits/vegetables
5)if you can peel it, you can eat it sushi and sashimi
(base) C:\Project\Blood Cell\scripts>
```

Fig -6: Displaying some precautions for the users to maintain a normal blood cell count

5. CONCLUSION

In this work, a machine learning approach to automatically identify and count blood cells from a smear image based on CNN algorithm is presented. Our proposed method is evaluated on publicly available dataset. It is observed that our method accurately identifies RBC's, WBC's, and Platelets. The proposed method has also been tested on a dataset of smear images, where it has performed satisfactorily. With the accuracy and the detection performance of the proposed method, it can be said that, the method has the potential to ease up the manual blood cell identification and counting process.

REFERENCES

- [1] Sarrafzadeh O., Dehnavi A.M., Rabbani H., ET AL.: 'Circler based framework for red blood cells segmentation and counting', December 2015
- [2] Lou J., Zhou M., Li Q., ET AL.: 'An automatic red blood cell counting method based on spectral images', October 2016
- [3] Acharjee S., Chakrabartty S., Alam M.I., ET AL.: 'A semiautomated approach using GUI for the detection of red blood cells', 2016

- [4] Cruz D., Jennifer C., Valiente L.C., ET AL.: 'Determination of blood components (WBCs, RBCs, and platelets) count in microscopic images using image processing and analysis', December 2017
- [5] Acharya V., Kumar P.: 'Identification and red blood cell automated counting from blood smear images using computer-aided system', 2018
- [6] <https://www.kaggle.com>