Leukemia Detection using an improved method with k-means

Debojyoti Dey\_19BCE1186, Avihrik Basak\_19BCE1202, Pranish Shrestha\_19BCE1758

Vellore Institute of Technology, Chennai.

*Abstract***— Leukemia, commonly known as “blood cancer” is a fatal and serious disease. It is caused by irregular WBC cell division. To provide proper treatment to the patient, we must detect the issue in its early stages. This can be done in a multitude of ways, both manual, like cell culturing, and digital methods like taking microscopic images of the person’s blood cells and passing them through some automated image processing method. Our algorithm uses k-means clustering algorithm to segment out WBCs, followed by a traditional method to detect the RBCs. Digital methods yield early results compared to manual methods and can allow the patients to take proper steps before the cancer advances to later stages. Our algorithm detects acute lymphocytic (or lymphoblastic) leukemia and has an accuracy rate of greater than 95% using this method.**

*Keywords*— **Leukemia detection, acute lymphocytic leukemia, Image processing, k-means, filtering, Image de-noising.**

1. **INTRODUCTION**

Image processing has been used in the recent times to employ multiple techniques in order to assist people at certain tasks. As such, it finds great usage in the medical field where time is a precious resource, and can result in the life or death of an individual.

Leukemia is the term used for the cancer of blood cells. It has four major types: Acute Lymphocytic Leukemia (ALL), Acute Myelogenous Leukemia (AML), Chronic Lymphocytic Leukemia (CLL) and Chronic Myelogenous Leukemia (CML). It can be manually detected by cell culturing but that process is time consuming and requires professionals to determine the outcome as the classification is done manually under a microscope.

Digital image processing does not require neither trained professional nor expensive lab equipment nor long amounts of time. Using this, we can detect leukemia by just inputing an microscopic image of the blood smear of the person and the algorithm will automatically determine the presence of leukemia.

We considered several algorithms to achieve our goal but finalized on using the k-means algorithm and other feature extraction processes to get our results.

1. **PROCESS OVERVIEW**

**Image Acquisition**

For our research, we collected multiple microscopic images of blood smears from both healthy individuals and leukemia affected individuals. This formed the reference for computing the results for the implementation. Then we ran several image segmentation and other image preprocessing methods to remove the background noise and extract out the WBCs and RBCs.

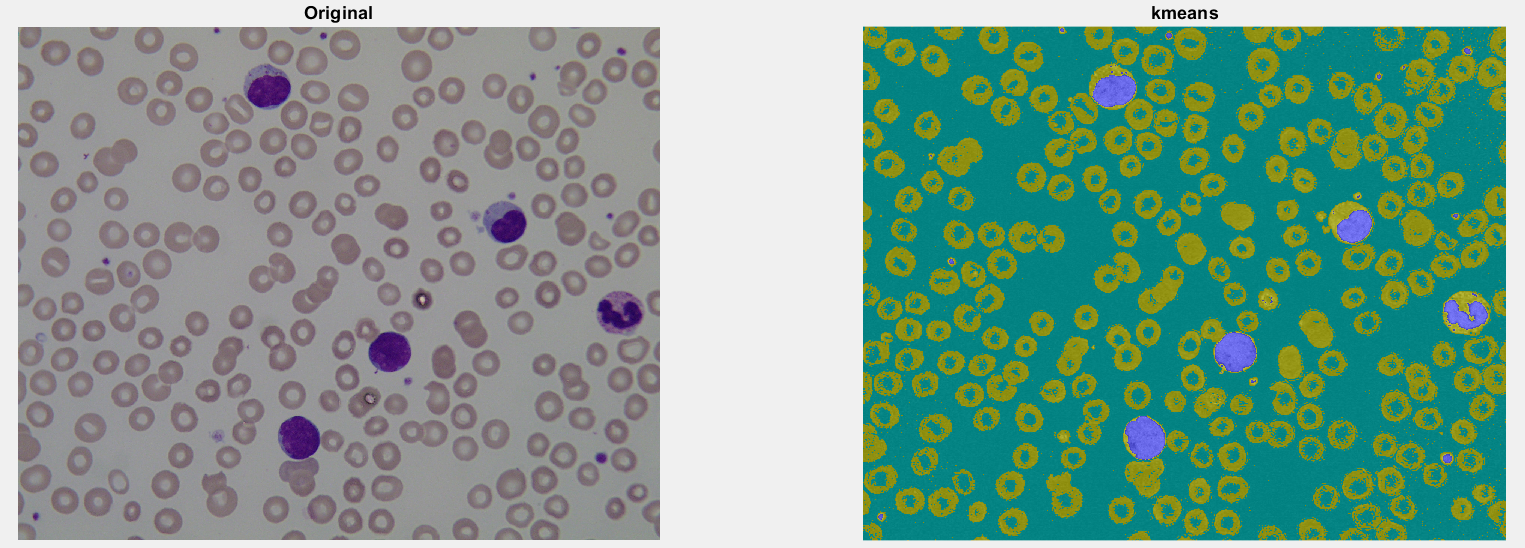
**Image Preprocessing and Feature Extraction**

First, the image is resized to better suit our algorithm. Then we extract the “saturation” component of the image. This gives us a 2-D matrix to work with. When k-means segmentation is performed on this 2-D matrix, using 3 clusters, the WBCs are highlighted in a color that separates it from the RBCs and other background components. The resulting image is converted to HSV color model. The components of the resultant image, hue, saturation and intensity, are then subtracted from each other. Upon performing several permutations of this process, we found that the difference between the intensity and saturation components provides greater distinction of WBCs. The resultant image is then filtered out for noise and is used as a mask to extract and count the WBCs.

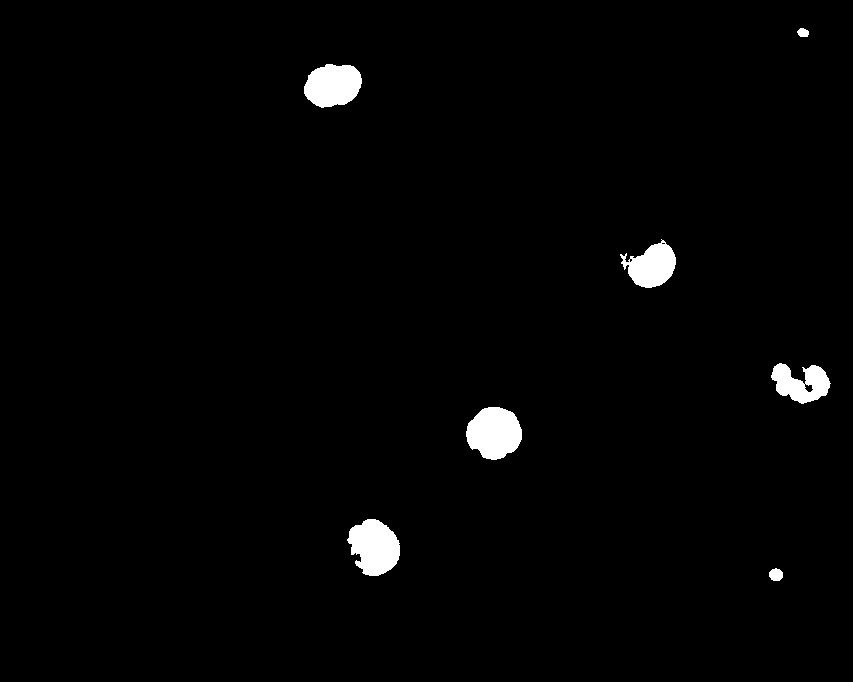
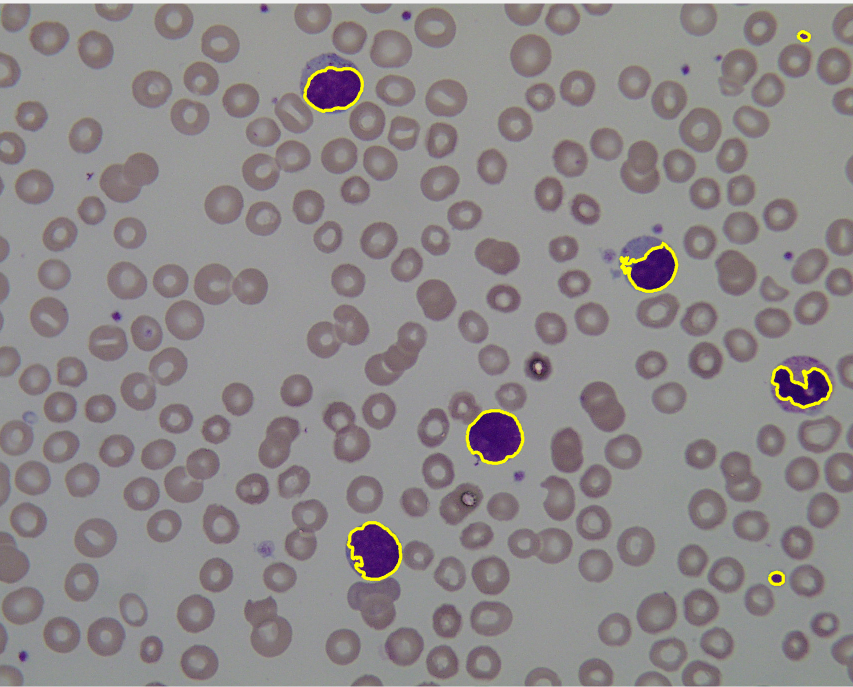
To extract the RBCs, a different approach was carried out as their overlapping nature and color prevented them from being picked up properly by k-means. To overcome this issue, we converted the image to a binary image and cleared up the background noise. The white portions were then dilated and any ring-shaped or deformed portions were filled up to turn them into disks. The image was passed through a circle finding algorithm to count the number of circular objects. This gives us an estimate of the number of the total cells in the blood smear. We can get the number of RBCs by subtracting the number of WBCs from the number of total cells.

We calculated the ratio of WBCs to RBCs of the images in our dataset to get an estimate of the threshold of healthy and leukemia affected cells. Using that data, we applied a check on the ratio of WBC to RBC of the sample to determine whether the sample shows signs of acute lymphocytic leukemia or not.

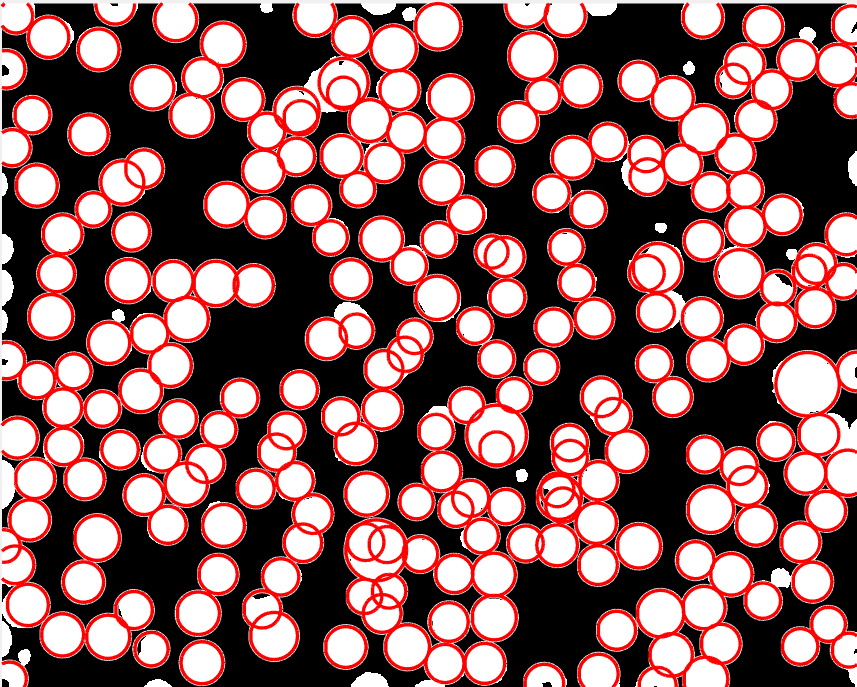




*Figure 1 –* Original image vs k-means segmented image

*Figure 2 –* Mask for WBCs *Figure 3 –* WBCs extracted out



*Figure 4* – Total cells

1. **LIMITATIONS**

* Better image pre-processing methods are needed to properly differentiate out WBCs that are in close proximity to each other as this algorithm often considers a group of WBCs as a single cell.
* Can only be used to detect acute lymphocytic leukemia.
* Getting an accurate count of the RBCs is problematic due to their overlapping positions in the smear sample.

1. **FUTURE WORK**

* More efficient method for approximating the actual count of the total cells in sample needed.
* A shape detection algorithm, with a database detailing the shapes of deformed WBCs, can be used in conjunction with this algorithm to detect other forms of leukemia.

1. **CONCLUSION**

This paper suggests a relatively simple algorithm to detect the presence of leukemia using blood smear samples. The proposed system has been tested with over 108 images and has been able to detect leukemia with great efficiency. Images are converted to HSV model to optimize the efficiency of the algorithm. The algorithm can be further refined using techniques mentioned in the “Future Work” section.

**REFERENCES**

1. Prasidhi G. Fal Desai, Geeta Shet, (*Department of Electronics and Telecommunication, Goa University,(India))* “DETECTION OF LEUKEMIA USING IMAGE PROCESSING”, *International Journal of Advance Research in Science and Engineering,* Vol.07, Special Issue No.03, April 2018, pp. 149-156
2. V. Shankar, M. M. Deshpande, N. Chaitra and S. Aditi, "Automatic detection of acute lymphoblasitc leukemia using image processing," 2016 IEEE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 186-189
3. Dr. R. Janaki, “Detection of leukemia in microscopic white blood cell images using gaussian feature convolutional visual recognition algorithm”, Journal of Critical Reviews, Vol. 07, Issue 3, January 2020, pp 173-180
4. Yazan M. Alomari, Siti Norul Huda Sheikh Abdullah, Raja Zaharatul Azma, and Khairuddin Oma, “Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm”, *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 979302, 17 pages, 2014.
5. Lorenzo Putzu, and Cecilia Di Ruberto, “White Blood Cells Identification and Counting from Microscopic Blood Image,” World Academy of Science, Engineering and Technology International Journal of Medical, Health, Pharmaceutical and Biomedical Engineering Vol:7 No:1, 2013.
6. Understanding Leukemia -- Diagnosis and Treatment, *Jennifer Robinson, MD*, *https://www.webmd.com/cancer/lymphoma/understanding-leukemia-treatment*
7. DBScan and Gaussian mixtures with Gene Expression data for Acute Myeloid Leukemia, [*https://www.leukemiaairesearch.com/blog/detection-early-detection/article/dbscan-gaussian-mixtures-gene-expression-acute-myeloid-leukemia*](https://www.leukemiaairesearch.com/blog/detection-early-detection/article/dbscan-gaussian-mixtures-gene-expression-acute-myeloid-leukemia)
8. “The 5 Clustering Algorithms Data Scientists Need to Know”, Towards Data Science, [*https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68*](https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68)
9. Categories of Leukemia, *https://www.cancer.org/cancer/chronic-myeloid-leukemia/about/what-is-cml.html*