

PAPER 111 – DESIGN OF A LEUKEMIA DETECTION SYSTEM USING DIGITAL BLOOD SMEAR IMAGES

S. Saminu^{1*}, I. O. Muniru¹, S. A. Yahaya¹, A. J. Oladimeji¹, M. T. Ajibola¹, M. O. Ibitoye¹, Y. K. Ahmed¹, L. J. Jilantikiri¹

¹Department of Biomedical Engineering, University of Ilorin, Ilorin, Nigeria.

*Email: saminu.s@unilorin.edu.ng

ABSTRACT

Leukaemia is a fatal blood cancer that occurs due to the formation of abnormal and excessive increases in white blood cells in the bone marrow or blood. The traditional approaches used to diagnose the disease involve the manual analysis of blood sample images obtained from a microscope. This approach is tedious, slow, time-consuming, and prone to errors. Therefore, automatic detection of leukaemia based on the counting of the two blood cells is paramount for diagnosis and increasing the patient's survival rate. This paper presents a system that can detect each of the two blood cells needed through image processing, segmentation, and classification. The detection, classification, and counts are only limited to two of the cells present in the digital blood smear which are the white blood cells (WBCs) and red blood cells (RBCs). The model was evaluated with a collection of confirmed cases and normal cases to test its effectiveness in predicting the presence of Leukaemia by computing the ratio of WBC to RBC. The suggested model exhibits good performance results and can be utilized to make a reliable computer-aided diagnosis detection of leukaemia cancer.

KEYWORDS: Leukaemia, White Blood Cells (WBCs), Red Blood Cells (RBCs), Detection, Machine Learning, Blood Smear.

1. INTRODUCTION

Patients suffering from cancer that are dying due to the delay in discovering the cancerous cells in a blood smear is alarming, especially in developing countries. Cancer has been one of the leading causes of death among humans of all ages and gender according to research conducted by the Centre for Disease Control and Prevention in the United States (Bray *et al.*, 2021). Leukaemia is a type of cancer that affects the White Blood Cells (WBCs) of the blood through malicious cloning of abnormal WBC; the bone marrow generates aberrant WBCs (Leukocytes). These aberrant cells should normally die after a while in normal circumstances, but, instead, their number increases. These aberrant WBCs prevent normal WBCs from carrying out their functions and as such cause Leukaemia disease. According to National Cancer Institute (USA), blood cancer is treatable if identified early (Sun, 2020). Leukaemia is detected by counting the corresponding blood cells in the suspected patient's blood sample under a microscope; however, this method can be tedious, time-consuming, and unsettling. One method of detecting blood cancer is by doing a complete cell or blood count to check for abnormalities in the blood which may indicate the presence of Leukaemia (Hegde, 2019). The complete blood count is performed by a haematologist who examines the blood sample under a microscope and checks for abnormalities as there is a normal range for both Red Blood Cell count and white blood cell count.

There are three types of cells in the human body; White blood cells, Red blood cells (RBCs) and, Platelets as depicted in Figure 1. The WBC are responsible for providing the necessary immune system required by the human body and it plays a pivotal role in protecting against the various diseases and infections by acting as a defence wall in the body (Shafique, 2018). The RBC is responsible for supplying oxygen to the body tissues from the heart (Fujita, 2021). According to Healthline, the normal RBC count range from 4.7 to 6.1 million cells per microliter (μL) for men and 4.2 to 5.4 million cells per microliter (μL) for women and the normal number of White Blood Cells in the blood is 4,500 to 11,000 WBCs per microliter (4.5 to $11.0 \times 10^3/\mu\text{L}$). (Kannarkat, 2022). The nature of the disease can be determined using the correct counting and classification of these blood cells. This count is often tedious, hence the reason for the design of this system, which does the complete blood count automatically using image processing segmentation of blood smears and counting.

The design of a system that counts white blood cells and red blood cells in a sample digital blood smear and check for the presence of excessive white blood cells would reduce the manual work of medical practitioners, thus aids in the early detection and cure.

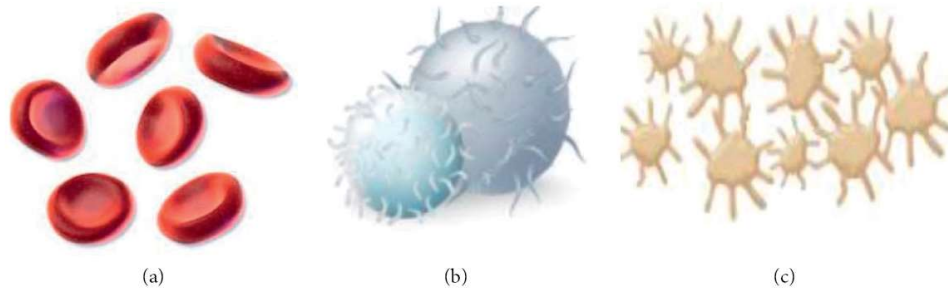


Figure 1. (a) Red blood cells (b) White blood cells (c) Platelets (Bukhari, 2022)

Several works have reported the segmentation of lymphocyte images as a region of interest for laboratory image analysis using computer-aided diagnostic and machine-learning techniques to improve the diagnosis and early detection of the disease (Jagadev, 2017). Cells in different types of leukaemia have been counted and differentiated based on the analysis of blood smear images (Wen, 2018; Xin, 2016). (Ghaderzadeh, 2021) proposed a colour-based segmentation method that segment the lymphoblast and lymphocyte details of WBCs and eliminates extra elements contained in the blood. Generally, the segmentation of leukaemic disease is categorized into shape-based, region-based, and pixel-based approaches (Al-jaboriy, 2019). Edge-based and k-means were applied by (Sarder, 2020) to segment several smears of blood. The hybrid technique based on morphological and thresholding were proposed to segment the smears of blood with high accuracy (Bukhari, 2022). Most of the reported works in the literature employed complex machine learning and deep learning approaches that require sophisticated, higher memory processing units for algorithms execution. Therefore, this work aims to design a system that aids the detection of leukaemia (excessive presence of white blood cells) using image processing and Hough transform to count red blood cells and white blood cells in a digital blood smear based on low complexity and low-cost resources. This work only encompasses the complete blood count using digital image processing to detect whether there is a presence of leukaemia or not, the design does not tell which type of Leukaemia is present, its stage or the method of treatment.

The objectives of this work are to:

1. Analyze and extract both red blood cells and white blood cells using colour segmentation in MATLAB environment.
2. Design a system that can count the extracted entities i.e. (WBCs and RBCs count)
3. Ensure the accuracy of design results by testing against both confirmed Leukaemia blood smear samples and normal blood smear samples.
4. Deploy on a portable hardware that interfaces the user with the model.

The rest of this paper is organized as follows: section two discusses the methodology and materials used in this study. section three provides the results and a discussion of the results obtained. Finally, section four presents the conclusion of the study.

2. MATERIALS AND METHODS

2.1 Dataset

The dataset used in this work was obtained from the American Society of Hematology (ASH) image bank, a web-based image library that offers a comprehensive collection of images relating to a wide range of hematologic topics. It is a publicly available database that can be accessed via "<https://imagebank.hematology.org/>".

2.2 METHODOLOGY

A MATLAB-based system was developed to ease the task of automatic classification of blood smear samples as having leukaemia or not. The system is a graphical user interface designed using GUIDE, an interactive design tool in MATLAB. The implementation process follows a typical image analysis pipeline. The block diagram

depicting the stages followed to achieve the proposed system is shown in Figure 2. It consists of image acquisition, image pre-processing, image segmentation, cell counting, and performance analysis. The main idea followed to achieve the desired result is that for any blood smear sample, the normal ratio of white blood cells to red blood cells should not be greater than 10%. The system is designed to generate an output “POTENTIAL LEUKEMIA DETECTED” if the ratio is more than 10%.

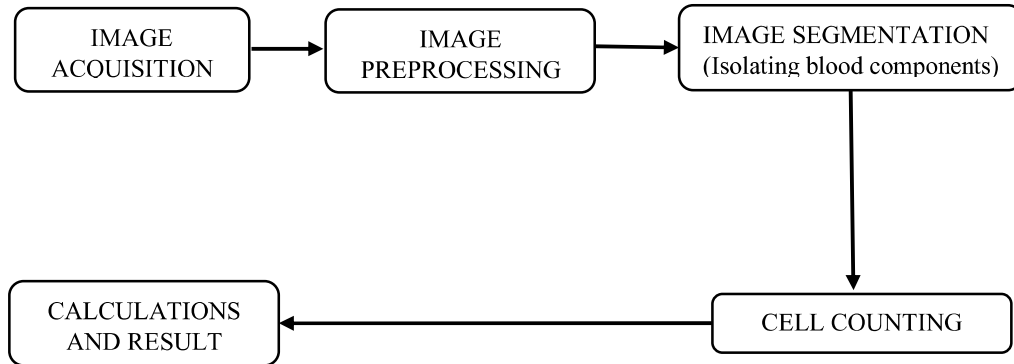


Figure 2. Block diagram of the proposed system

2.2.1 IMAGE ACQUISITION: In this study, the actual capturing of the microscopic images of the blood smear was not carried out. Rather, images downloaded from the ASH database were used. The image analysis process commences by loading an image into the MATLAB application built for this study.

2.2.2 IMAGE PREPROCESSING: The main goal of image preprocessing is to enhance the quality of the image by removing noise, correcting distortions, improving contrast, adjusting brightness and colour balance, and removing unwanted artefacts. Doing this can help to improve the accuracy of the subsequent image analysis steps. The final task then involves the conversion of the image to three common colour models used in image processing namely RGB, CYMK, and YCbCr. RGB is an additive colour model used in digital imaging systems that combines red, green, and blue colour channels to describe colours, with a range of intensity levels for each channel between 0 and 255. CMYK, on the other hand, is a subtractive colour model used in printing that describes colours by the amount of cyan, magenta, yellow, and black needed to produce a colour. Meanwhile, YCbCr is a colour space used in digital video and image processing that separates colour information from brightness information, with three channels: Y for brightness, and Cb and Cr for colour. YCbCr is often used in digital video and image formats such as JPEG and MPEG due to its efficient compression.

The reason for this is to determine which of the colour scales gives the best performance.

2.2.3 IMAGE SEGMENTATION:

This function allows the application to segment the image into its constituent cells using image processing techniques such as thresholding and morphological operations. Whereas thresholding is used to separate two or more images by setting an intensity threshold, a morphological operation is used to manipulate the shape and size of objects in an image. This step separates the white blood cells, red blood cells, and platelets from each other. The outcome of this function is that rectangular bounding boxes are placed around the WBCs and RBCs to enhance visual counting

2.2.4 CELL COUNTING: In this part, the bounded cells per group are counted automatically using the Hough transform. The ratio of the WBCs to the RBCs is then computed. An output is generated based on the value of the ratio.

3 RESULTS AND DISCUSSION

The range of WBCs and RBCs for a healthy adult is between 4,500 μL -11,000 μL and 4,500 μL – 5,900 μL respectively (Sun, 2020). μL pronounced as microliter is one-millionth of a litre, and is commonly used as a unit of measurement for blood components. In a sample of the portion of a regular arrangement of red blood cells, it is considered that if the ratio of WBCs to RBCs is 8% or more (Li, 2020), the outcome of such is regarded as “POTENTIAL LEUKEMIA DETECTED”.

3.1 MODEL PERFORMANCE

The system is meant for counting components in blood smear samples. To test for performance, the model is supplied with various images of confirmed cases of leukaemia in blood smears and normal blood smears without any presence of excessive white blood cells.

3.1.1 Performance of the system with T-cell prolymphocytic leukaemia digital blood smear

T-cell prolymphocytic leukaemia (T-PLL) is an extremely rare and typically aggressive malignancy (cancer) that is characterized by the out-of-control growth of mature T-cells (T-lymphocytes). T-cells are a type of white blood cell that protects the body from infections. T-PLL affects older adults with a median age at diagnosis of 61 years, and it is more common in men than in women. (Castorvy et al, 2012). A typical T-Cell prolymphocytic blood smear sample used for the test is shown in Figure 3.

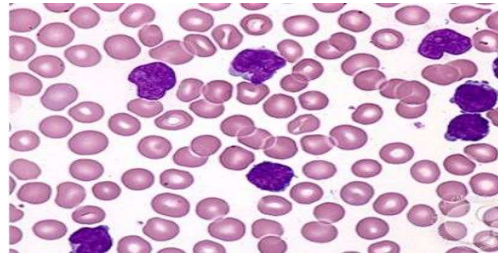


Figure 3. T-cell prolymphocytic leukaemia blood smear

The output from input the above image into the design is shown as follows:

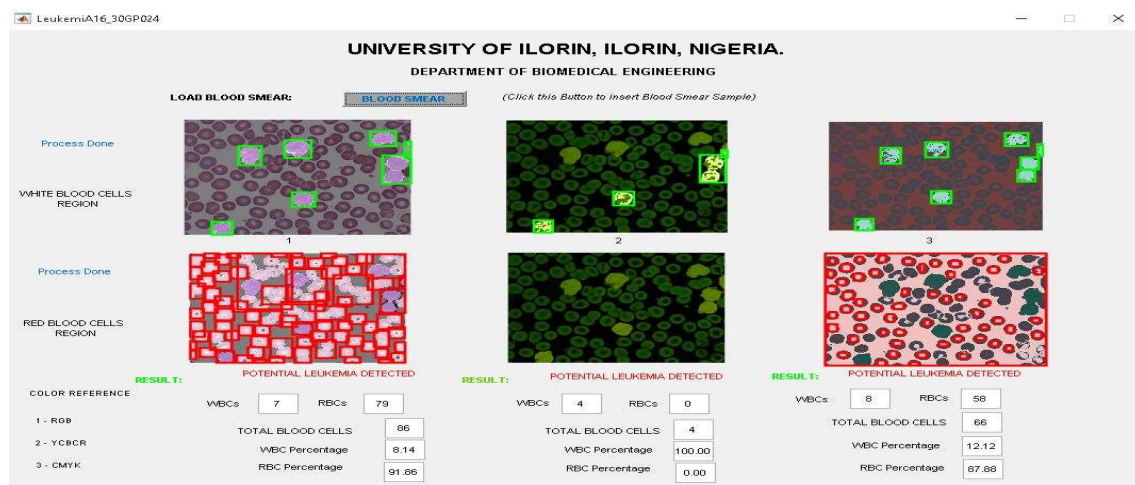


Figure 4. Result output for T-cell prolymphocytic leukaemia blood smear

The processed T-cell Prolymphocytic Leukaemia blood smear image shown in Figure 3 is transformed within 3 colour spaces to isolate overlapping cells into individual cells. There is a presence overlapping of white blood cells in the RGB regions of Figure 3 which alters the output of the result for processing in Figure 4. For the YCbCr colour space conversion of Figure 4, the overlapping cells are well isolated and the end user decides on

which colour space result to use. The CYMK colour region fails to identify red blood cells and can therefore be ignored in this result.

3.1.2 Performance of the system with normal cells from digital blood smear

A sample image of a normal blood smear with no Leukaemia shown in Figure 5 is tested against the design to ensure accurate performance of the design.

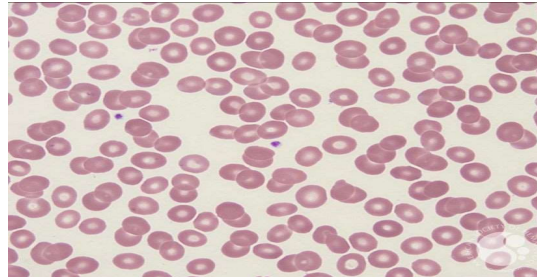


Figure 5. Normal blood smear

The output of the above image supplied to the designed system gives an output shown in Figure 6 which corresponds to the expected result for a normal blood smear

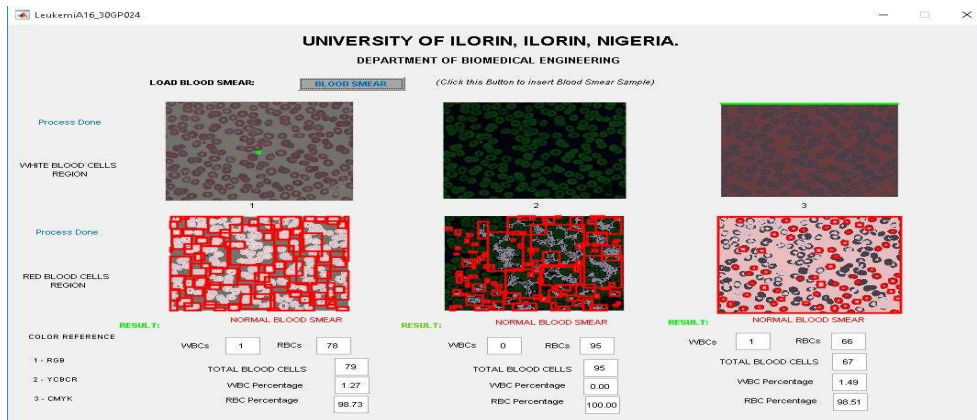


Figure 6. Result of normal blood smear from design output

The design runs the supplied image across three colour spaces, as sources and pixels of images differ, the best output which best selects the needed component should be used for result. It is discovered that the result correctly predicts the presence of a normal blood sample.

3.1.3 Performance of the system based on accuracy

To check if the design predicts the required result, series of tests were carried out by running the design against confirmed blood smear samples of both normal and leukemic blood smears. Two samples relating the actual expected result, also known as MANUAL COUNT and the results acquired from the design “DESIGN COUNT” for both normal and leukemic blood smear are presented in Tables 1.1 and 1.2.

Table 1.1. Test of design result against manual count for a sample Leukaemia blood smear

		WBCs	RBCs	PREDICTION	
MANUAL COUNT		7	82	LEUKEMIA Present.	
DESIGN COUNT		COLOUR SPACES	WBCs	RBCs	PREDICTION
		RGB	7	79	Potential Leukaemia Detected
		CMYK	4	0	Potential Leukaemia Detected
		YCbCr	8	58	Potential Leukaemia Detected

Table 1.2. Test of design result against manual count for a sample Normal blood smear

		WBCs	RBCs	PREDICTION	
MANUAL COUNT		1	98	Normal	
DESIGN COUNT		COLOUR SPACES	WBCs	RBCs	PREDICTION
		RGB	1	78	Normal Blood Smear
		CMYK	0	95	Normal Blood Smear
		YCbCr	1	66	Normal Blood Smear

4 CONCLUSION

A system that can count both RBCs and WBCs using image processing and segmentation will go a long way in relieving the stress put on the haematologist during the analysis of digital blood smear for leukaemia presence. The design proposed in this paper is suitable for quick counting of digital blood smear rather than having haematologist do it manually which can be stressful and contribute to errors. Although a haematologists still have to select the best suitable segmentation out of the 3 colour spaces used in this design, which are either RGB, CYMK or YCbCr, the performance results presented is still promising. A design that can classify the different types of Leukaemia detected by means of machine learning algorithm is required, as after using this design to count the various blood cells to determines excesses in the presence of white blood cells. There is still a need to study the shape of the corresponding cells to diagnose the type of Leukaemia present.

REFERENCES

- Al-jaboriy, S.S. Sjarif, N.N. Chuprat, S. and Abdulllah, W.M. (2019), Acute lymphoblastic leukaemia segmentation using local pixel information, *Pattern Recognition Letters*, 125, 85–90.
- Bray, F., Laversanne, M., Weiderpass, E., & Soerjomataram, I. (2021), The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*, 127(16), 3029-3030.
- Bukhari, M. Sadaf, Y. Saima, S. and Ahmed, A. (2022), A Deep Learning Framework for Leukaemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning, *Mathematical Problems in Engineering*, 2022, Article ID 2801227, 18 pages.
- Catovsky, D. Dearden, C. (2012), Treatment of T cell prolymphocytic leukaemia. In *UpToDate*. <https://www.uptodate.com/contents/treatment-of-t-cell-prolymphocytic-l> Dearden C. How I treat prolymphocytic leukaemia. *Blood*, 120(3), 538-551.
- Fujita, T.C. Sousa-Pereira, N. Amarante, M.K. and Watanabe, M.A. (2021), Acute lymphoid leukaemia etiopathogenesis, *Molecular Biology Reports*, 48(1), 817–822.

Ghaderzadeh, M. F. Asadi, A. Hosseini, D. Bashash, H. Abolghasemi, and A. Roshanpour, (2021), Machine learning in detection and classification of leukaemia using smear blood images: a systematic review,” Scientific Programming, 2021, Article ID 9933481.

Hegde, R.B. Prasad, K. Hebbar, H. Singh, B.M.K. and Sandhya, I. (2019), Automated decision support system for detection of leukaemia from peripheral blood smear images, *Journal of Digital Imaging*, 33, 361–374.

Jagadev, P. and Virani, H. (2017), Detection of leukaemia and its types using image processing and machine learning, in *Proceedings of the 2017 International Conference on Trends in Electronics and Informatics (ICEI)*, IEEE, Tirunelveli, India, 522–526.

Kannarkat, G. T., Darrow, J., & Moghekar, A. (2022). Reassessing accuracy of blood cell correction factor for traumatic lumbar puncture. *Journal of the Neurological Sciences*, 432, 120097.

Li, L. and Wang, Y. (2020), Recent updates for antibody therapy for acute lymphoblastic leukaemia, *Experimental Hematology & Oncology*, 9, 33–11.

Sarder, M.A. Maniruzzaman, M. and Ahammed, B. (2020), Feature selection and classification of leukaemia cancer using machine learning techniques, *Machine Learning Research*, 5(2), p. 18.

Shafique, S. and Tehsin, S. (2018), Acute lymphoblastic leukaemia detection and classification of its subtypes using pretrained deep convolutional neural networks, *Technology in Cancer Research & Treatment*, 17, Article ID 1533033818802789.

Sun, L., Yu, Y., Niu, B., & Wang, D. (2020). Red blood cells as potential repositories of microRNAs in the circulatory system. *Frontiers in genetics*, 11, 442.

Wen, J. Xu, Y. Li, Z. Ma, Z. and Xu, Y. (2018), Inter-class sparsity based discriminative least square regression Neural Networks, 102, 36–47.

Xing, F. and Yang, L. (2016), Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review