ADVANCEMENTS IN LEUKEMIA DIAGNOSIS: A COMPREHENSIVE REVIEW OF DEEP LEARNING APPROACHES FOR WHITE BLOOD CELL IMAGE ANALYSIS

Abstract

Cancer of the blood is one of the most prevalent and perilous forms of the disease. The aberrant and unregulated synthesis of white blood cells (WBC) in the bone marrow is the root cause of leukaemia, a kind of blood cancer. Getting an early diagnosis of leukaemia increases the likelihood that cancer can be cured with the appropriate therapy. One method that can be used to identify leukaemia is the counting of the quantity of white and red blood cells (RBC). In the conventional approach, the counting of white blood cells and red blood cells is performed with the assistance of a piece of machinery called a haemocytometer. These examinations take a lot of time and are quite difficult to understand, which might lead to errors in categorization.

Image processing of microscopic pictures, which is very affordable due to the uncomplicated nature of the circuitry, is another method that may be used to improve the identification of cancer cells. Researchers have found cancer cells, but because they didn't look at all of the criteria at the same time, such as picture enhancement, noise reduction, image recognition, and so on, their findings aren't as accurate as they might be. The goal and purpose of this research is to create a new system that is capable of taking into account all of the aspects that have been discussed thus far. Image acquisition, image pre-processing, picture segmentation, edge detection, and feature extraction are some of the steps that are included in the process. These phases are used to identify the presence of cancer as well as the stages at which cancer progresses. The paper contributes to the

Authors

S. Dhiravidaselvi
PG Scholar
Roever Engineering College Perambalur, India.
nivethaselvi98@gmail.com

Dr. P. Deepan
Associate Professor & Head
Department of CSE
St. Peter’s Engineering College
Hyderabad.
deepanp87@gmail.com

Dr. R. Santhoshkumar
Associate Professor & Head
Department of CSE
St. Martin’s Engineering College
Hyderabad.
santhoshkumar.aucse@gmail.com

Dr. B. Rajalingam
Associate Professor & Head
Department of AI&DS
St. Martin’s Engineering College
Hyderabad.
rajalingam35@gmail.com

R. Vidhya
Research Scholar
Department of IT
B.S. Abdur Rahman Crescent Institute of Science and Technology
Chennai.
vidyarajesh23@gmail.com

process of measuring the amount of white blood cells and red blood cells, as well as their average cell sizes and whether or not they are regular or irregular. This information may be used to determine whether or not a person has cancer.

Keywords: Leukemia disease; disease diagnosis; machine learning; deep learning; white blood cell; image analysis;

I. INTRODUCTION

Blood is essential for life, and the normal functioning of a large number of organs in the body is dependent on it. It is possible to assess the state of health of the blood's cells by doing an analysis of the blood's contents. Blood typically consists of cells in addition to plasma, which is a component of the blood that is liquid [1]. Plasma accounts for the remaining 55% of the entire volume of blood, whereas blood cells make up around 45% of the total volume. There are three different kinds of blood cells: erythrocytes, leukocytes, and thrombocytes [2]. Red blood cells are called erythrocytes. White blood cells are called leukocytes. Platelets are called thrombocytes. While the white blood cells (WBC) account for around 1% of the blood, the red blood cells (RBC) make up approximately 40–45% of the total blood volume. Organs throughout the body require each of the three main kinds of blood cells for a specific reason. The bone marrow is responsible for the production of white blood cells, which are a vital component of the blood. White blood cells are in charge of the immune system, which serves as a defence mechanism in the body against outside materials, particularly disease-causing substances [3].

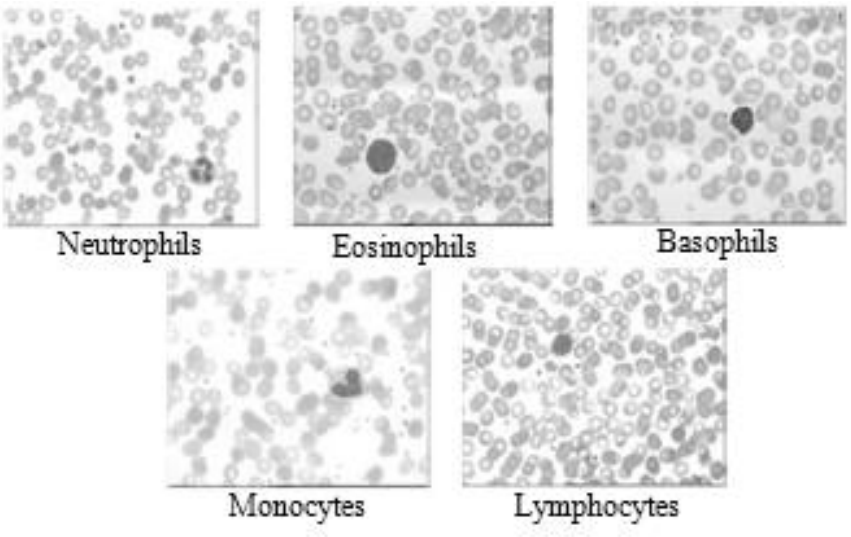

Figure 1: White Blood Cells Sample in Different Aspects

The immune system is a defence mechanism against foreign elements in the body. There are five different types of white blood cells, as seen in Figure 1. These white blood cells include neutrophils, eosinophils, lymphocytes, monocytes, and basophils. sometimes, granulocytes and agranulocytes, sometimes known as nongranulocytes, are two distinct kinds of blood cells that can be further subdivided (see Figure 2 for more information).

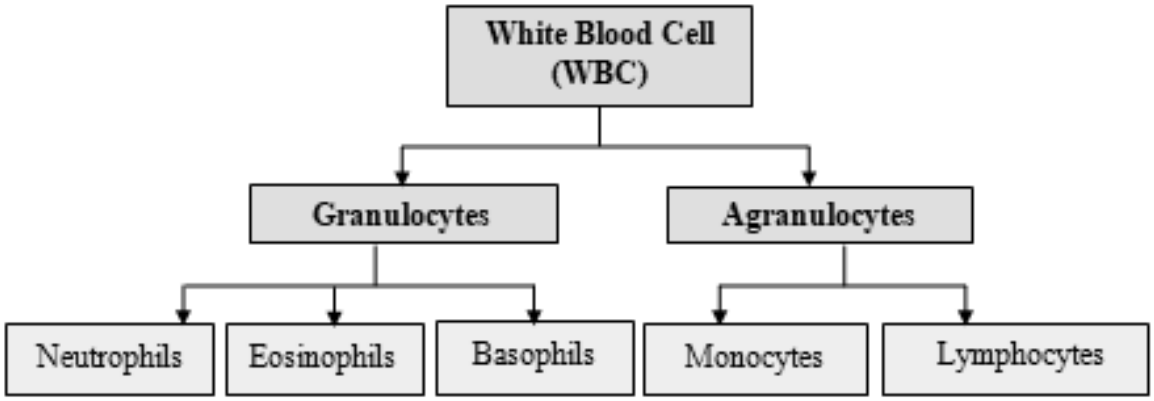

Figure 2: White Blood Cell and its Types

Granulocytes are white blood cells that have granules that are visible under a microscope, whereas agranulocytes are white blood cells that do not have any granules that are visible under a microscope [4]. The cells known as neutrophils, eosinophils, and basophils belong to the granulocytes class, while monocytes and lymphocytes are members of the agranulocytes class. There are between 40 and 60 percent neutrophils, 1 to 4 percent eosinophils, 20 to 40 percent lymphocytes, 2 to 8 percent monocytes, and 0.5 to 1 percent basophils in the blood [5].

Each of the five different kinds of white blood cells performs a unique set of duties and is associated with a distinct component of a patient's health (a topic). Specifically, precise identification makes it possible to count the many different white blood cells in order to evaluate whether or not they are present in the appropriate or anticipated proportions. After they have been recognised, the various white blood cells can then be isolated for the purpose of conducting a more in-depth examination for abnormalities [6]. Analysing a patient's white blood cells, both quantitatively and qualitatively, can reveal a significant amount of information on their overall health. It is feasible, for instance, to screen individuals for disorders such as leukaemia, anomalies in the immune system, and malignant cells [7]. Traditionally, identification has been carried out in a laboratory environment. First, pictures of blood cells have been obtained, then they have been stained with certain chemicals, and finally, an expert has examined the images using a microscope. This method, on the other hand, is delicate and calls for a human specialist to do the test without making any mistakes. Unfortunately, after many hours of testing, specialists might grow fatigued and end up making inaccurate identifications of white blood cells.

The remaining parts of the paper may be broken down into the following categories: In the following section, we will conduct a study of the research literature in the subject of automated detection systems for leukaemia identification and classification. In addition, we will examine relevant work that has been done throughout the years, with a particular emphasis on picture segmentation and classification. In Section 3, we investigate the process of collecting datasets and apply several data augmentation strategies to enhance the quality of the dataset photos. In Section 4, we will discuss the ongoing research in deep learning techniques, and in Section 5, we will discuss the findings of the study as well as its implications for the future.

II. LITERATURE SURVEY

In the past few decades, a large number of researchers have created categorization and diagnostic systems for leukaemia disorders based on machine learning. These systems use pictures of white blood cell counts. For instance, Hiremath [8] utilised a number of different methods for the segmentation process, such as the k-mean clustering algorithm and the Expectation Maximisation algorithm. Active contour, the snake algorithm, and zack thresholding were utilised by Farnoosh Sadeghian and colleagues [9] in order to provide a framework for the extraction of the nucleus and cytoplasm area in white blood cells. Yujie, Yuhki, and Kohei [10] have presented morphological methods that may be used to detect cancer cells in a sample. Vinod and Kimbahune, neleh [11] put out the idea of counting and segmenting different types of blood cells. The counting and extraction of red blood cells is an idea that was put out by Poomeokrakl and Neatpisamvanit [12]. A technique known as histogram equalisation was used in Fatin A.Dawood's [13] unsupervised nucleus segmentation of microscopic white blood cell images.

Likewise, Navi D Jambhekar [14] suggested the Red Blood Cell Classification by Support Vector Machine method. Nasrul and Muhammad [15] are able to find red blood cells by using a method called the circular Hough transform. Heidi Berge [16] came up with the idea of segmenting red blood cells in a thin blood smear picture using the Zack's Method as her foundation. The Hough transform was suggested by Guitao et al. [17] for use in urine microscopy as a method for locating and collecting red blood cells.

The present state of the art in blood smear analysis automation was discussed in Kendall Preston's [18] article, along with the state of the art in other relevant topics such as multi-resolution microscopy. Otsu adaptive thresholding and the watershed transform were first described by Nicola Ritter [19] for the purposes of cell segmentation and the identification of cell borders in pictures of peripheral blood smear slides. Histogram thresholding was utilised by Roy Dimayuga et al. [20] in order to distinguish the nuclei of leukocytes, also known as white blood cells, from the rest of the cells present in the image.

Madhuri and Patil [21] proposed the otsu adaptive thresholding in segmenting and extracting the WBC in the blood cell image. This was done on the basis of thresholding. Sumeet [22] created a watershed segmentation technique for RBC counting and using the Circular Hough Transform methodology in order to determine the amount of red blood cells. Wavelet transformation was used by Nur Alom [23] to separate the various types of blood cells, and then a fuzzy inference system was used to draw a conclusive judgement on blood cancer based on the total number of cells that were counted. It was hypothesised by Fauziah Kasmin and colleagues [24] that leukaemia may be identified in a microscopic image. A technique for RBC segmentation based on the making and watershed algorithms for automated RBC counting was proposed by Sharif J.M. et al. [25]. This approach was developed for automated RBC counting.

The inability of over segmentation approaches and machine learning models to generate appropriate results in terms of accuracy is the most prevalent issue that is observed in a variety of surveys and research. In addition, a manual diagnosis of leukaemia illness requires a significant amount of time and may not give trustworthy results. In order to

overcome the issues, we will need to develop an automated system for the diagnosis, classification, and detection of leukaemia based on a deep learning model.

III. CURRENT RESEARCH IN AI TECHNIQUES

This section presents a complete review of the approaches that are used for advancements in leukemia illness prediction. This analysis includes a comparison of the pre-processing stages, feature extraction stages, and classification stages. As can be seen in Figure 3, we are able to divide those methods into two distinct categories: machine learning techniques and deep learning approaches.

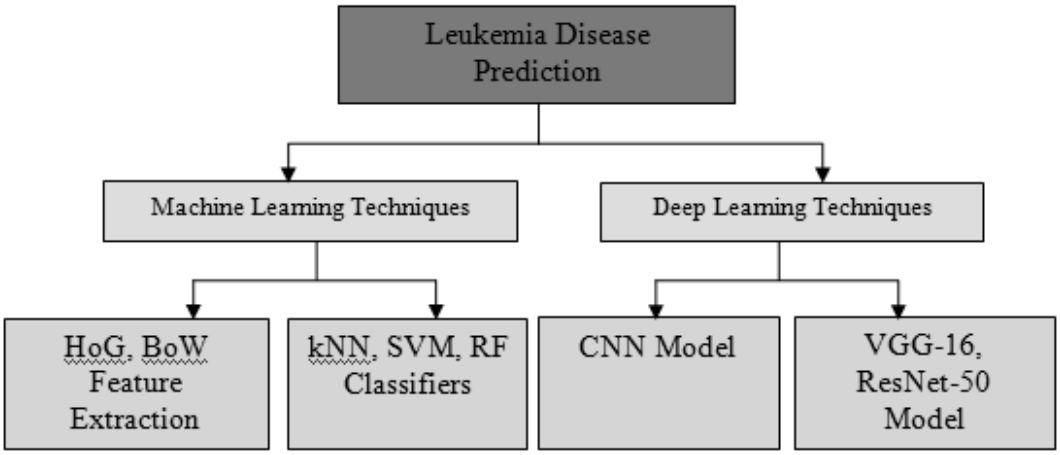

Figure 3: Leukemia Disease Prediction Techniques

1. Machine Learning Techniques

Machine learning is one of the developing approaches that may be used to categorise the leukemia diagnosis that are found in the dataset for white blood cell image analysis. Figure 4 illustrates the order of operations involved in the categorization of leukemia disease diagnosis using machine learning. In a classification system for leukemia diseases that is based on machine learning, there are three phases that need to be completed [26].

- Pre- Processing Phase
- Feature Extraction and
- Classification

Pre- Processing Stage: During the pre-processing stage, distortion is reduced to improve image quality and simplify further processing. Many images are processed in advance through steps including smoothing, enhancing, cropping, and colour space conversion [27]. Depending on the image quality, this module's functionality will change. Improved contrast and brightness, among other effects, may be achieved by applying the appropriate filters to the image. A Laplacian filter is also used to further refine the picture. Additionally, Gabor wavelets and histogram equalisation are used to filter and control the impacts of various lighting arrangements [28].

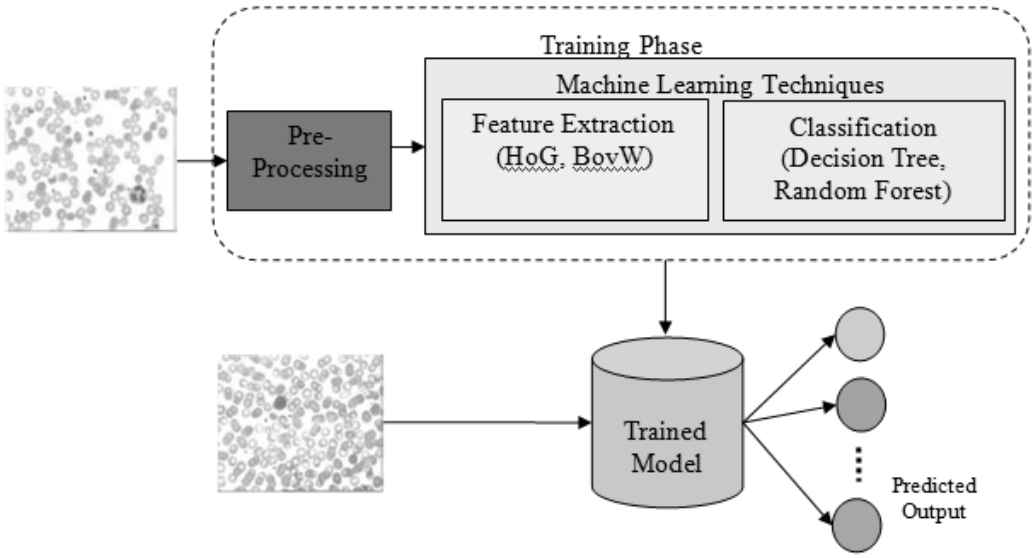

Figure 4: Machine Learning Based Leukemia Disease Recognition System

Feature Extraction Stage: In order to extract the features from the input images, the second and crucial stage is feature extraction. Low-level and high-level features can both be included [29]. The feature's granularity lies in its ability to extract the edges or boundaries of objects. Like the high-level feature, the interior shape of objects may be extracted. Features can be extracted using a number of different techniques; some of the most common include the Histogram of oriented Gradients (HoG) feature, the Bag-of-Words (BoW) feature, the Scale Invariant Feature Transform (SIFT) feature, the Speed- Up Robust Feature (SURF) feature, and the Haar-like feature [30].

Classification Stage: In order to determine which group the new information belongs to, classification is used [31]. The classification method may be broken down into two independent subfields: supervised and unsupervised learning. Supervised classification goes down to training a classifier using a pre-existing labelled dataset. The Support Vector Machine (SVM), Artificial Neural Network (ANN), k-Nearest Neighbour (kNN), Logistic Regression, Decision Tree, and Random Forest are all supervised learning methods [32].

2. Deep Learning Techniques

Deep learning is a machine learning methodology employed for constructing models that acquire knowledge from many types of data, including images, videos, audio, and text, with the purpose of executing tasks such as classification and detection. The process of scene categorization involves the utilisation of several models, such as Convolutional Neural Network (CNN) [26] and Deep CNN models [27], to categorise situations. Deep learning methodologies encompass several techniques such as Convolutional Neural Networks (CNN), transfer learning utilising VGG-16 and VGG-19 architectures, Recurrent Neural Networks (RNN), Auto Encoders, and Generative Adversarial Networks (GAN) [34]. In the field of image classification, Convolutional Neural Networks (CNNs) and transfer learning models are widely employed methodologies.

Convolutional Neural Networks (CNNs) are widely employed in the field of computer vision for the purposes of picture segmentation and classification. Although Convolutional Neural Networks (CNNs) were initially developed in 1989, their remarkable achievement in the ImageNet Competition in 2012 attracted significant further interest. The computational complexity of Convolutional Neural Network (CNN) architecture exhibits an upward trend due to the escalating number of layers, neurons with a multitude of weights, and interconnections among diverse neurons [35]. The fundamental block structure of a Convolutional Neural Network (CNN) is seen in Figure 5. This diagram illustrates the many components, including convolutional layers, pooling layers, activation functions, and fully connected layers, each serving distinct functions within the network.

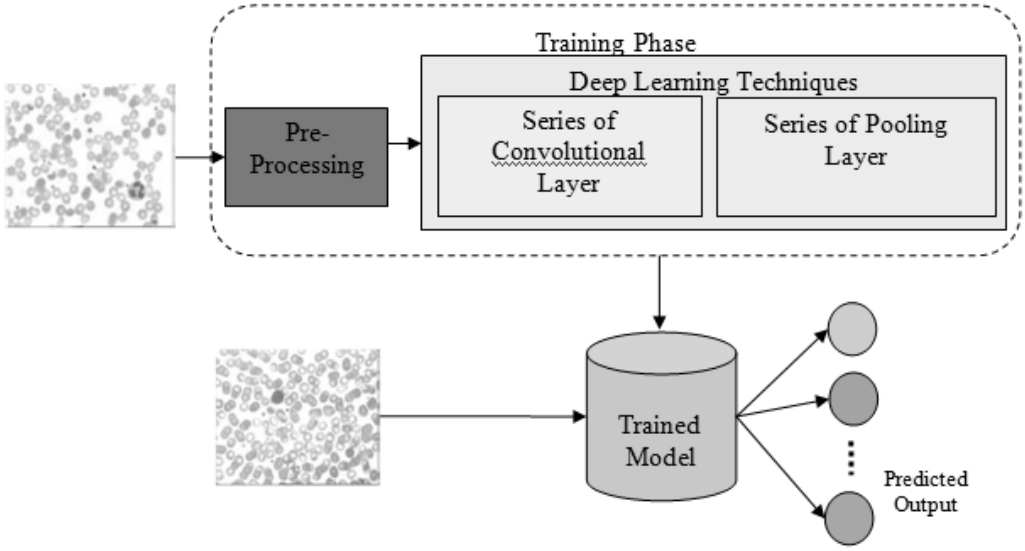

Figure 5: Deep Learning Based Leukemia Disease Recognition System

The generation of feature maps is accomplished by convolving input pictures with a kernel in the convolutional layer. at the process of information propagation to subsequent layers, the preceding convolutional layers' outcomes undergo downsampling at the pooling layer. This downsampling is achieved by either selecting the maximum or computing the average value within the given neighbourhood. The loss function is integrated with the output of the remaining layers of the convolutional neural network (CNN) model in order to provide a prediction of the input data. The determination of network parameters involves the minimization of the loss function, which measures the discrepancy between predicted values and the actual ground truth labels. This process is carried out while also adhering to the regularisation constraints. Furthermore, the backpropagation algorithm is employed to iteratively adjust the network weights until convergence is achieved.

IV. DATASET USED

The dataset plays a crucial role in the deep learning algorithm, since it is imperative to gather accurately labelled data. Nevertheless, white blood cells (WBCs), specifically bone marrow smears, suffer from a significant absence of an extensive accessible dataset. Consequently, a dataset was gathered for the purpose of this study. The bone marrow smear

samples were produced by the Department of Laboratory Medicine at Seoul National University Hospital. The bone marrow aspiration underwent the Wright-Giemsa staining technique. The prepared slides were imaged using a light microscope at a magnification of 1000x. A dataset including 200 pictures was obtained for analysis. Each image had a pixel resolution of 1080 x 1330. The photos were acquired from 30 slides, each belonging to one of 10 human participants. For each slide, a random selection of five to ten nonoverlapping collection sites was made. The process of picture preparation was conducted in an anonymous manner, utilising pre-existing data. The study protocol was granted exempt status by the institutional review board of Seoul National University Hospital.

The complete set of photos underwent hand cropping to generate individual patch images, each measuring 96 by 96 pixels in resolution. A total of 2,174 cropped photos were collected and subsequently classified into appropriate categories by two haematologists who had expertise in the field. The classification process was confirmed by these experts. The dataset comprised ten distinct classes of white blood cells (WBCs) at various stages of maturation. These classes included four consecutive stages of the erythroid series, namely pronormoblast (C1), basophilic normoblast (C2), polychromatic normoblast (C3), and orthochromatic normoblast (C4). Additionally, the dataset encompassed six consecutive stages of the myeloid series, specifically myeloblast (C5), promyelocyte (C6), myelocyte (C7), metamyelocyte (C8), band neutrophil (C9), and segmented neutrophil (C10).

The dataset exhibited an imbalanced distribution of classes. The unequal distribution of white blood cells, as seen in the White Blood Cell Differential Count, posed an inevitable challenge during the data collection process. However, the unbalanced dataset might pose a challenge during network training due to the limited quantity of data instances from certain classes.

In order to address the issue of unbalanced data resulting from the diverse distribution of white blood cells, the technique of oversampling was employed for classes that had a relatively small number of data items. During the data preparation process for classes with limited data, we manually cropped numerous photos of the same cell at slightly varied centres. Consequently, there was an increase in the quantity and diversity of data points gathered.

V. CONCLUSIONS

This literature review examines the range of traditional and machine learning methods employed for quantifying white blood cell production in the bone marrow. One of the techniques employed in the identification of leukaemia involves quantifying the quantity of white and red blood cells. The conventional method involves the enumeration of white blood cells and red blood cells through the utilisation of specialised equipment known as a haemocytometer. The aforementioned tests are characterised by their demanding nature in terms of time and complexity, thereby resulting in the potential for misclassification.

The literature suggests a pressing demand for more research aimed at the development of automated techniques that can aid pathologists in the identification and classification of leukaemia. There has been a notable rise in the utilisation of automated methods for evaluating microscopic smear pictures in order to diagnose cases of leukaemia illness. Deep

learning techniques, including the use of Convolutional Neural Networks (CNNs) and Deep Convolutional Neural Networks (DCNNs), are expected to see a surge in demand for image classification tasks in the near future.

REFERENCES

[1] V. Piuri and F. Scotti, "Morphological classification of blood leucocytes by microscope images," in Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA '04), pp. 103–108, Boston, Mass, USA, July 2004.

[2] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA '05), pp. 96–101, Messian, Italy, July 2005.

[3] S. Liu, Z. Deng, J. Li, J. Wang, N. Huang, Q. Ye, "Measurement of the refractive index of whole blood and its components for a continuous spectral region," Journal of Biomedical Optics, Vol. 24(3), Article ID 035003, pp. 15-21, 2019.

[4] P. Zhou, Z. Meng, M. Liu et al., "The associations between leukocyte, erythrocyte or platelet, and metabolic syndrome in different genders of Chinese," Medicine, Vol. 95(44), Article ID e5189, pp. 12-17, 2016.

[5] K. E. Badior and J. R. Casey, "Molecular mechanism for the red blood cell senescence clock," IUBMB Life, Vol. 70(1), pp. 32–40, 2018.

[6] C. Herron, "Know your WBCs," Nursing Made Incredibly Easy!, Vol. 10(1), pp. 11–15, 2012.

[7] A. Mathur, A. S. Tripathi, and M. Kuse, "Scalable system for classification of white blood cells from Leishman stained blood stain images," Journal of Pathology Informatics, Vol. 4(2), pp. 1–15, 2013.

[8] P.S.Hiremath, Parashuram Bannigidad, Sai Geeta, "Automated identification and Classification of White Blood Cells (Leukocytes) in Digital Microscopic Images" IJCA Special Issue on "Recent Trends in Image Processing and Pattern Recognition"Vol. 1, pp. 1-12, 2010.

[9] Famosh Sadeghian, Zainina Seman, Abdul Rahman Ramli, Badrul Hiham Abdul Kahar, and M-iqbal aripan "A Framework for white Blood cell Segmentation Microscopic Blood Images using Digital Image Processing", Vol. 5(1), pp. 121-128, 2012.

[10] Yujie L, Yuhki Kitazono, Kohei Miyata, Shota Nakashima Shiyuan Yang, enchi erikawa, "An improved Detection Algorithm Baed on Morphology Methods for Blood Cancer Cell Detection", Pediatric Radiology, Vol. 2(5), pp. 643–650, 2014.

[11] Vinod V.Kimbahune, Mr. Nelesh j Uke, "Blood Cell Image segmentation and Counting", Int. J. Adv. Res. Sci. Eng. Technol., Vol. 3, pp. 136–142, 2013.

[12] J. Poomeokrakl and C. Neatpisamvanit, "Red Blood cell Extraction and Counting", Procedia Comput. Sci. Vol. 5(8), pp. 535–542, 2015.

[13] Fatin A.Dawood, "Unsupervised Segmentation of Microscopic White Blood Cell Image Using Histogram Equalization Technique", Journal of Algorithms Comput. Technol, Vol. 1(3), pp. 1-10, 2014.

[14] Navi D Jambhekar, "Red Blood Cells Classification using Image Processing", Bioengineering, Vol. 4(1), pp. 10-20, 2013.

[15] Nasrul Humaimi Mahmood and Mohammad Araf Manor,"Red Blood cells Estimation using Hough Transform Technique", IPIJ, Vol.3, No.2. April 2012.

[16] Heidi Berge, Dale Taylor, riram Krishnan, ad Tania. Dougla. Improved Red Blood Cell Counting in thin Blood smears. Proceedings of IBI, pp.201-207, 2011.

[17] Guitao Cao, Cai Zhong, Ling Li and Jung Dong, "Detection of Red Blood Cell in Urine Micrograph". The 3rd International Conference on Bioinformatics and Biomedical Egineering (ICBBE). 2009.

[18] KendallPreton Jr., "Image Processing In Medical Micrscopy', IEEE transaction 1986.

[19] Nicola Ritter, Jame Cooper "Segmentation and Border Identification of cells in Images of Peripheral Blood Smear Slides", Journal of Neural Network, Vol. 61, pp. 85–117, 2015.

[20] Roy A.Dimayuga, Gerwin T. Ong, Rainier Carlo S.Perez, Gefferson O. Siy, Saman C. Sohrabi Langroudi and Miguel O.Gutierrez, "Leukemia Detection Using Digital Image Processing in mat lab". ECE Student from, De La Salle University, Manila. March 26,2010.

[21] Madhuri G.Bnamaro: "Automatic Blood cell Analysis by using digital Image processing": A preliminary study. IJERT Vol.2. sep.2013.

[22] Sumeet Chourasiya: "Automatic Red blood cell counting using watershed segmentation", IJCSIT, Vol. 5, 2014.

[23] .Nur Alom Taalukdar: "Automatic Blood cancer detection using Image processing based on fuzzy system", IJARCSSE, Vol. 4(2), Aug.2014.

[24] Fauziah Kasmin, Anton satria prabuwono, Azizi Abdullah: "Detection of leukemia in human blood sample based on microscopic images: A study", Journal of Theoretical and applied information technology , Vol. 46(2), pp. 1-10, 2013.

[25] J. M. Sharif, M. F. Miswan, M. A. Ngadi, Md ah Hj alam, Muhammad Mahadi bin abdul Jamil, "Red blood cell segmentation uing masking and watershed algorithm": A preliminary study, ICOBE, penang Malaysia, 2012.

[26] Shijie, J.; Ping, W.; Peiyi, J.; Siping, H. Research on data augmentation for image classification based on convolution neural networks. In Proceedings of the Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 4165–4170.

[27] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv 2014, arXiv:1409.1556.

[28] W. King, K. Toler, and J. Woodell-May, "Role of white blood cells in blood-and bone marrow-based autologous therapies," BioMed Research International, Vol. 2(8), Article ID 6510842, pp. 1-8, 2018.

[29] S. Fathima and K. Farhath, "Blood cells and leukocyte culture—a short review," Blood Research and Transfusion Journal, Vol. 1(2), Article ID 555559, pp. 12-20, 2017.

[30] S. Arslan, E. Ozyurek, and C. Gunduz-Demir, "A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images," *Cytometry A*, vol. 85, no. 6, pp. 480–490, 2014.

[31] M. Saraswat and K. V. Arya, "Automated microscopic image analysis for leukocytes identification: a survey," *Micron*, vol. 65, pp. 20–33, 2014.

[32] V. Singhal and P. Singh, "Correlation based feature selection for diagnosis of acute lymphoblastic leukemia," in *Proceedings of the 3rd ACM International Symposium on Women in Computing and Informatics (WCI '15)*, pp. 5–9, Kochi, India, August 2015.

[33] Nguyen, L. D., Lin, D., Lin, Z. & Cao, J. Deep cnns for microscopic image classification by exploiting transfer learning and feature concatenation. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5 (IEEE, 2018).

[34] Duggal, R., Gupta, A., Gupta, R., Wadhwa, M. & Ahuja, C. Overlapping cell nuclei segmentation in microscopic images using deep belief networks. In *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, ICVGIP '16, 82:1–82:8, https://doi.org/10.1145/3009977.3010043

[35] Marzahl, C., Aubreville, M., Voigt, J. & Maier, A. Classification of leukemic b-lymphoblast cells from blood smear microscopic images with an attention-based deep learning method and advanced augmentation techniques. In Gupta, A. & Gupta, R. (eds.) *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, 13–22

