International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Blood Cell Image Classification Using the Machine Learning Methods With Nature Inspired Optimization

[1]Kumod Kumar Gupta, [2]Kamalraj R, [3]Rupal Gupta, [4]Surendra Yadav

Submitted:19/04/2023 **Revised:**09/06/2023 **Accepted:**24/06/2023

Abstract: Blood cells are the biological components that make up our bodies' blood. They are essential in a variety of physiological activities like oxygen transport, immunological response, and coagulation. There is theoretical as well as practical interest in the issue of detecting and counting blood cells inside the blood smear. For the diagnosis and treatment of many disorders, the differentiating blood cell count offers pathologists vital information. In this study, we offer integrated of Enhanced Naive Bayes Ant Colony Optimization (ENACO) for efficiently classifying and identifying blood cell images. The most crucial steps in this automated procedure are segmentation and classification of blood cells. To detect and categorize the various kinds of blood cells, Digital microscopic pictures of the stained blood cells are segmented, and geometrical characteristics are extracted for each segment. The Enhanced Naive Bayes-Ant Colony Optimization (ENBACO) was suggested. The experimental findings are compared with the researchers' other current approaches, demonstrating the efficiency of the suggested strategy. Typical assessment metrics include F1-score, recall, accuracy, and precision. The suggested results were superior to other traditional procedures.

Keywords: *Blood cell, image classification, Enhanced Naive Bayes Ant Colony Optimization (ENBACO)*

1. Introduction

Blood cell evaluation and counting are particularly instructive in clinical practice. Patients with blood diseases should pay special attention to this information as they go through the phases of observation, development, and planning for their therapy [1]. Researchers must identify the blood cells in the blood samples and determine their relative abundance in order to provide an accurate diagnosis of the condition. Red blood cells (RBCs), white blood cells (WBCs), and platelets may be identified among blood cells in peripheral blood as erythrocytes, leukocytes, and thrombocytes [2]. RBCs are necessary for carrying oxygen throughout the body. Blood clotting depends heavily on platelets [3].

The immune system's initial line of defense is comprised of leukocytes. Numerous blood conditions, including leukemia, acute infections, and inflammation, cause aberrant white blood cell multiplication. When blood problems are present, differential leukocyte counts may provide important information for a precise illness diagnosis [4]. Leukocytes may be divided into five primary kinds based on the nucleus's size and structure, the cytoplasm's staining color, and the neutrophil, eosinophil, basophil, monocyte, and lymphocyte ratio of nucleus to cytoplasm [5].However, differential counting of blood cells requires a lot of time and effort if it is exclusively done by humans [6]. Additionally, the outcomes may vary based on each expert's own subjective viewpoint. Although the automated cell counters that are now on the market are based on the concepts of flow cytochemistry and laser light scattering, of all analyzed blood samples, 23% need expert microscopic analysis [7]. Consequently, several attempts to create automated cell analysis systems employing image processing have already been established. Hence In this paper we introduce Enhanced Naive Bayes-Ant Colony Optimization (ENBACO) for classified Blood cells.

2. Related Works

The study [8] suggested a novel Deep Convolutional Generative Adversarial Network (DC-GAN) and Residual Neural Network (ResNet)-based methodology for classifying blood cell images. The paper [9] presented a fused convolutional neural network (CNN) model in this study to categorize images of white blood cells (WBC). We use a fully connected network with one hidden layer, three max-pooling layers, five convolutional layers and these layers together. The research [10] proposed the CNN, W-Net based technique for classifying WBCs. They test W-Net using a large-scale actual data set that they got from The Catholic University of Korea. This dataset consists of

1Assistant Professor, Department of Artificial Intelligence (AI), Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India, Email id: kumodkumargupta@gmail.com
2Professor, Department of Computer Science and IT, Jain (Deemed-to-be University), Bangalore-27, India, Email Id: r.kamalraj@jainuniversity.ac.in
3Assistant Professor, College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, Email id: r4rupal@yahoo.com
4Professor, Department of Computer Science & Application, Vivekananda Global University, Jaipur, India, Email Id: surendra.yadav@vgu.ac.in

6,562 genuine photos of the five categories of WBC. The article [11] proposed novel method for classifying microscopic pictures of WBCs using an Optimized Feed Forward Neural Network (OFNN). To teach a Feed-Forward Neural Network (FFN) and find its weights such that it has a low error rate and a high classification rate, we use the Particle Swarm Optimization (PSO) and the Gravitational Search Algorithm (GSA). The article [12] employed the parallelism features of GPUs to speed up the separation of WBCs from microscopic images. The study [13] suggested a two-stage hybrid multi-level strategy for accurately separating four distinct types of cells: lymphocytes, monocytes, segmented neutrophils, and eosinophils. The study [14] proposed using bounding boxes to locate and identify white blood cells using the YOLOv3 object identification approach. The study [15] explained automatic WBC categorization based on morphological characteristics such form, color, size, and texture using conventional image processing techniques.

3. Proposed Method

The remaining portion of the manuscript is organized as follows: Concerning the aims or objectives of the research, segment 2 describes the preceding study and identifies any deficiencies or discrepancies. The Five different kinds of leukocytes, erythrocytes, and thrombocytes make up peripheral blood cells. Pathologists can diagnose and treat a wide range of disorders with the use of the differential counting of blood cells. Micrographs of blood cells are shown in Figure 1.

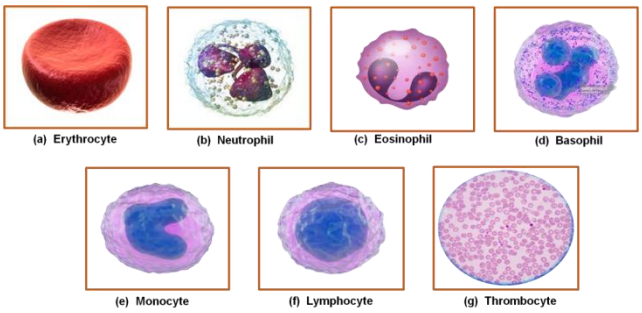

Fig.1. The blood cells' micrographs.

Blood cells may be categorized into seven categories on the basis of nuclear properties and cytoplasm: erythrocyte, neutrophil, eosinophil, basophil, monocytes, lymphocyte, and thrombocyte respectively shown in Figure 1.Therefore, the capacity to identify and characterize blood cells is significant for exact disease diagnosis. Preprocessing, image segmentation, feature extraction, and classification are the four significant procedures in our automated prediction of blood cells in microscopic images. Figure 2 depicts the automated recognition chart.

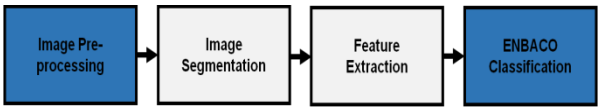

Fig.2. The Automatic identification of blood cells flowchart.

3.1. Image Pre-processing

Signal transmission noises of all kinds often impair images. De-noising a picture is crucial to image processing since it directly affects how well image processing operations like image segmentation, feature extraction, and image identification turn out. A common method for removing noise from images or other data is the median filter, a non-linear digital filtering technique. The basic concept behind the median filter is to replace each element in the signal with the median of its neighboring entries. The method is able to maintain the image's crisp edges and features while also removing noise. The picture of the blood cell is zoomed in Fig. 3(a). Fig. 3(b) provides a clear illustration of the relevant de-noising outcome.

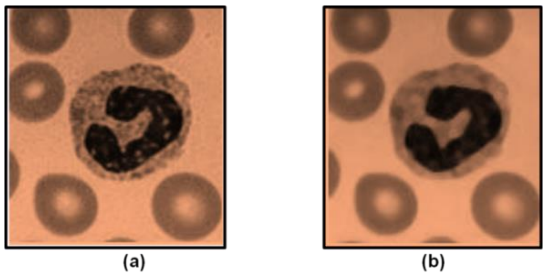

Fig.3. Median filter. (a) The blood cell picture that has been zoomed. (b) The outcome of (a) after de-noising.

3.2. Experimental setup for eye movement data dataset

Basically, image segmentation involves dividing an image into a number of distinct and homogenous sections that are believed to relate to picture items that are significant to a particular application. Thus, the blood slide is segmented by watershed to include each component in a separate space.

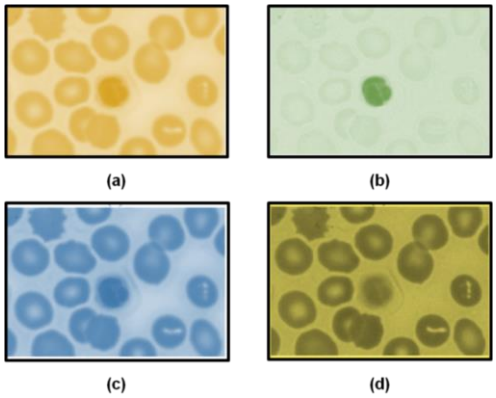

Fig.4. (a) The original picture of a blood cell (b) Red channel c) The green channel (d) The blue channel

3.3. Watershed

Finding all the lines with the greatest gray levels, or watershed lines, is the goal of watershed segmentation. The "immersion approach" is the most straightforward method of elucidating watershed segmentation. Let's pretend we've bored a hole in every surface minimum and are going to use that hole to fill several catch basins. Dams are constructed when additional immersion would cause water from separate catchment basins to mix. The water level will rise to a point when just the dam's peak is visible. To solve the problem of overlapping cells, the space transform of the binary mask of the cells with the largest area is applied to the watershed transform. The outcome of the watershed segmentation of the blood cell picture is shown in Figure 5.

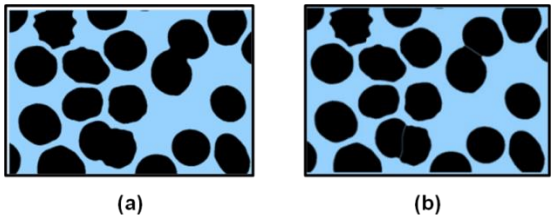

Fig.5. Representation dividing up an image (a) A binary representation of Figure 4(c) (b) The splitting of the watershed that follows from (a)

A composite association between eye movements and cognitive activities reveals deficiencies associated with each underlying process via the oculomotor controls for accomplished tasks. As a result, it shows signs of many complicated behaviors in those with mental problems. For research objectives, several gaze-related motions have been examined to get quantitative data on the categorization process. Data on eye movements include a range of oculomotor controls, including fixations, saccadic movement trajectories, duration, and velocity. Saccades are quick shifts in the direction of the gaze. During activities where maintaining fixation is required, ADHD participants have been seen to engage in a high number of intrusive saccades.

3.4. Feature extraction using Histogram

It should be noted that a normal human erythrocyte lacks a nucleus and has the appearance of a biconcave lens. As a result, we separate the nucleus using the red channel. Figure 6 shows the histogram from Figure 4(b).Since the nuclei are situated at less intensity level, erythrocytes may be distinguished with accuracy. The histogram enables us to discriminate between erythrocytes and leukocytes, as illustrated in Figure 6.

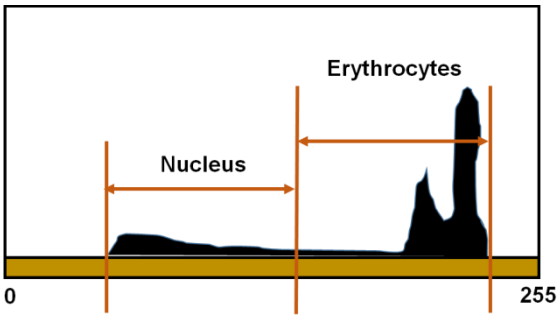

Fig.6. The histogram of figure 4(b).

3.5. Enhanced Naive Bayes-Ant Colony Optimization (ENBACO)

Using the Bayes rule and a strong assumption of independence, a Naive Bayes classifier is a straightforward probabilistic model. The conditional independence assumption is simplified in the Naive Bayes model. The words are conditionally independent of one another after they have been assigned a class (positive or negative). This presumption allows for the use of very quick classification algorithms without significantly affecting text classification accuracy. In their 2003 publication, the researchers explore the effectiveness of Naive Bayes on text categorization problems. In this instance, the expression: gives the greatest likelihood probability of a word belonging to a certain class.

There is much iteration in the ACO algorithm. A number of ants build comprehensive solutions on each iteration utilizing heuristic data and the accumulated knowledge of earlier populations of ants. The pheromone trail, which is left behind on the components of a solution, is used to symbolize these accumulated experiences. Depending on the issue being solved, the pheromone may be applied to the connections or components of a solution. Following is a description of the pheromone update rule process.

$$B(y_j|v) = \frac{Count\ of\ y_j in\ documents\ of\ class\ v}{Total\ no\ of\ words\ in\ documents\ of\ class\ v} \quad (1)$$

Throughout the training stage, hash tables are used to hold the word frequency counts. The Bayes Rule states that the likelihood that a certain document belongs to class v_j is given by,

$$B(v_j|t) = \frac{B(t|v_j)*B(v_j)}{B(t)} \quad (2)$$

The terms are conditionally independent of one another given a class (positive or negative), according to the simplified conditional independence assumption. This simplification assumption causes the model to be labeled as "naive."

$$B(v_j|t) = \frac{(\prod B(y_j|v_i))*B(v_i)}{B(t)} \quad (3)$$

The words of the text are represented by the y_j in this case.

The output of the classifier is the group with the highest posterior probability.

3.5.1. The transition rule

In the ACO algorithm, an ant is a straightforward computational agent. It builds a solution to the current issue iteratively. Every ant advances from a state r to a state s, which corresponds to a more thorough intermediate solution, during each iteration of the process. Based on the formula below, the r^{th}ant from state r to state s is chosen from the remembered unvisited states in I_k^r.

$$g = \arg\max_{u \in I_k^r} [\tau_j(k,w)^{\alpha}.\eta(k,w)^{\beta}] if\ o \le o0(Exploitation) \quad (4)$$

An a posteriori evaluation of the effectiveness of the maneuver is provided by the trail level. The trails are often updated after all ants have completed providing responses, with the level of the trails altering to indicate movements that were components of "good" or "bad" replies, respectively. The r^{th}ant typically transitions from state k to state g with probability$Br(k,g)$,

$$Br(k,g) = \begin{cases} \frac{\tau(k,g)^{\alpha}.\eta(k,g)^{\beta}}{u\epsilon I_k^r \tau(k,w)^{\alpha}.\eta(k,g)^{\beta}} if\ g\epsilon I_k^r \\ 0 \qquad otherwise \end{cases} \quad (5)$$

Here, the Possibility of a State Change is denoted by $Br(k,g)$, the pheromone concentration between states k and w in the j^{th} group is denoted by(k,g), also, the distance between states trail's length k and w is denoted by(k,g). I_k^r is the a group of uncharted states for the r^{th}ant in the j^{th} population; and are controlling factors; and o is a standardized probability between [0, 1].

3.5.2. The pheromone update rule

The pheromone trails need to be modified to raise the standard of the solution. Local and global updating is both included in trail updating. The following is a description of the local trail update formula:

$$\tau(k,w) = (1-\rho)\tau(k,g) + \sum_{r=1}^{n} \Delta\tau_r(k,g) \quad (6)$$

The pace at which the pheromone experiment evaporates is represented by (0, 1) in formula (6). The quantity of pheromone trail that was added to the edge (k,w) by ant r between time t and time t + 1t in the tour is known as $\tau(k,w)$. The description is as follows:

$$\Delta\tau_r(k,g) = \begin{cases} \frac{O}{F_r}(k,g)\epsilon\pi_r \\ 0\ otherwise \end{cases} \quad (7)$$

F_r is the distance traveled by the sequence t by the ant in one time unit, where O is a constant parameter.

4. Result and Discussion

Classifying blood cells is a crucial step in medical diagnosis since it aids in recognizing and comprehending a variety of illnesses and conditions. In this study, we proposed an enhanced Naive Bayes ant colony optimization algorithm for classifying blood cell images. The standard parameters for assessment include F1-score, recall, accuracy, and precision. We also compared our suggested technique ENBACO with other current methods like CNN, C-RNN, and RD-Net.

In classification tasks, accuracy is a popular assessment metric used to quantify the overall correctness of a model's predictions. It denotes the proportion of successfully predicted occurrences in relation to the total number of examples in the dataset. The accuracy of existing and suggested techniques is compared in Figure 7 and Table 1. The graph illustrates that the suggested approach is more precise than previous ones.

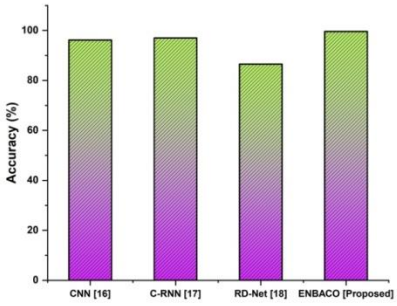

Fig.7. Accuracy comparisons between the suggested and current approaches.

Table 1. Comparison of Accuracy

Methods	Accuracy (%)
CNN [16]	96.2
C-RNN [17]	97
RD-Net [18]	86.53
ENBACO [Proposed]	99.6

precision compares the number of actual achievements to the number of projected successes to determine the precision of the suggested Procedure. It calculates the percentage of accurately predicted positive cases out of all positive instances predicted by the model. Figure 8 and Table 2 compare the precision of the existing and proposed approaches. As the graph shows, the suggested technique is more exact than prior ones.

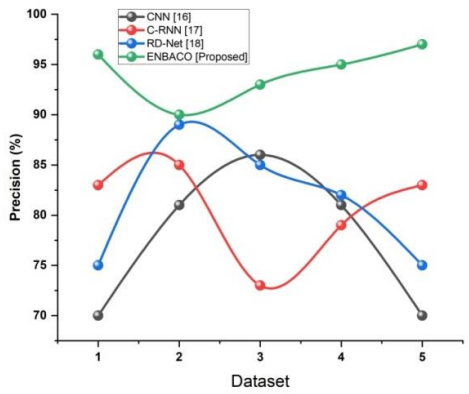

Fig.8. Comparison of precision of the existing and proposed methodologies.

Table 2. Comparison of Precision

Dataset	Precision (%)			
	CNN [16]	**C-RNN [17]**	**RD-Net [18]**	**ENBACO [Proposed]**
1	70	83	75	96
2	81	85	89	90
3	86	73	85	93
4	81	79	82	95
5	70	83	75	97

The recall is the probability of a positive test assuming it is positive. This is referred to as the true positive rate. It demonstrates that the new technique outperforms the previous methods in terms of recall. In Figure 9 and Table 3, the recall of the proposed strategy with existing techniques is compared to established ways.

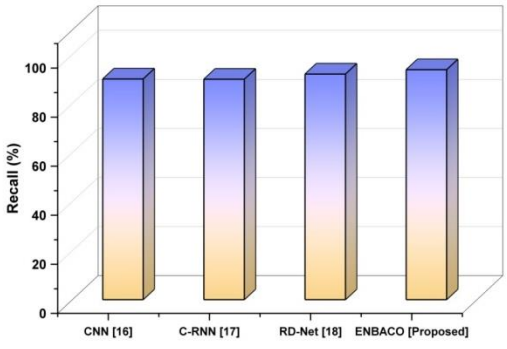

Fig.9. The comparison of the recommended method with the more traditional one.

Table 3. Comparison of recall

	Recall (%)
CNN [16]	90
C-RNN [17]	89.9
RD-Net [18]	92.01
ENBACO [Proposed]	93.79

The F1-Score is often utilized while evaluating material. It is possible to change the F1-score such that accuracy takes precedence over recall, or vice versa. When compared to the presently employed approaches, the proposed methodology has a higher f1-score. The f1-score of the suggested strategy is compared to that of the conventional techniques in Figure 10 and Table 4.

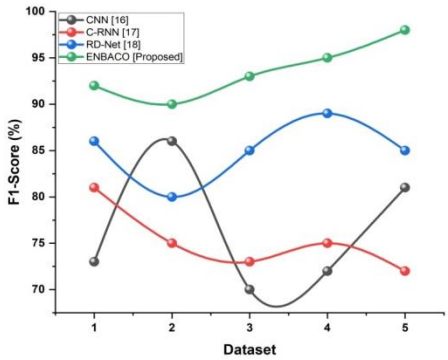

Fig.10. The Comparison of F1-score with Traditional and suggested techniques

Table 4. Comparison of F1-score

Dataset	F1-Score (%)			
	CNN [16]	C-RNN [17]	RD-Net [18]	ENBACO [Proposed]
1	73	81	86	92
2	86	75	80	90
3	70	73	85	93
4	72	75	89	95
5	81	72	85	98

5. Conclusion

The classification of blood cells is a crucial process in medical research and diagnosis. The automation and accuracy of blood cell classification have significantly improved because of developments in imaging and machine learning approaches. In this study ENBACO

classification scheme and geometric characteristics from the cytoplasm and nucleus were used in this work to develop an effective hierarchical blood cells classification approach. Because some leukocytes' cytoplasm shows just a very little variation from the background and contacts nearby cells, utilizing the recommended hierarchical approach for classification performed better than the single-stage classification method ENBACO. Furthermore, the experimental findings demonstrated that the ENBACO classification using hierarchical features might in fact enhance classification performance in comparison to the ENBACO technique.

References

[1] Choi, J.W., Ku, Y., Yoo, B.W., Kim, J.A., Lee, D.S., Chai, Y.J., Kong, H.J. and Kim, H.C., 2017. White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. PloS one, 12(12), p.e0189259.

[2] Osuka, A., Ishihara, T., Shimizu, K., Shintani, A., Ogura, H. and Ueyama, M., 2019. Natural kinetics of blood cells following major burn: impact of early decreases in white blood cells and platelets as prognostic markers of mortality. Burns, 45(8), pp.1901-1907.

[3] Mahdi, A., Cortese-Krott, M.M., Kelm, M., Li, N. and Pernow, J., 2021. Novel perspectives on redox signaling in red blood cells and platelets in cardiovascular disease. Free Radical Biology and Medicine, 168, pp.95-109.

[4] Singh, C. ., Gangwar, M. ., & Kumar, U. . (2023). Improving Accuracy of Integrated Neuro-Fuzzy Classifier with FCM based Clustering for Diagnosis of Psychiatric Disorder. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2s), 244–248. https://doi.org/10.17762/ijritcc.v11i2s.6143

[5] Luo, J., Chen, C. and Li, Q., 2020. White blood cell counting at point-of-care testing: A review. Electrophoresis, 41(16-17), pp.1450-1468.

[6] Yang, Z., Zhang, Z., Lin, F., Ren, Y., Liu, D., Zhong, R. and Liang, Y., 2017. Comparisons of neutrophil-, monocyte-, eosinophil-, and basophil-lymphocyte ratios among various systemic autoimmune rheumatic diseases. Apmis, 125(10), pp.863-871.

[7] Choi, J.W., Ku, Y., Yoo, B.W., Kim, J.A., Lee, D.S., Chai, Y.J., Kong, H.J. and Kim, H.C., 2017. White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. PloS one, 12(12), p.e0189259.

[8] Vembadi, A., Menachery, A. and Qasaimeh, M.A., 2019. Cell cytometry: Review and perspective on biotechnological advances. Frontiers in bioengineering and biotechnology, 7, p.147.

[9] Ma, L., Shuai, R., Ran, X., Liu, W. and Ye, C., 2020. Combining DC-GAN with ResNet for blood cell image classification. Medical & biological engineering & computing, 58, pp.1251-1264.

[10] Banik, P.P., Saha, R. and Kim, K.D., 2019, February. Fused convolutional neural network for white blood cell image classification. In 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 238-240). IEEE.

[11] Jung, C., Abuhamad, M., Alikhanov, J., Mohaisen, A., Han, K. and Nyang, D., 2019. W-net: a CNN-based architecture for white blood cells image classification. arXiv preprint arXiv:1910.01091.

[12] Mohamed, S.T., Ebeid, H.M., Hassanien, A.E. and Tolba, M.F., 2020. Optimized feed forward neural network for microscopic white blood cell images classification. In The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019) 4 (pp. 758-767). Springer International Publishing.

[13] Baker, Q.B. and Balhaf, K., 2017, April. Exploiting GPUs to accelerate white blood cells segmentation in microscopic blood images. In 2017 8th International Conference on Information and Communication Systems (ICICS) (pp. 136-140). IEEE.

[14] Cheuque, C., Querales, M., León, R., Salas, R. and Torres, R., 2022. An efficient multi-level convolutional neural network approach for white blood cells classification. Diagnostics, 12(2), p.248.

[15] Gyawali, M. Y. P. ., Angurala, D. M. ., & Bala, D. M. . (2020). Cloud Blockchain Based Data Sharing by Secure Key Cryptographic Techniques with Internet of Things. Research Journal of Computer Systems and Engineering, 1(2), 07:12. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/5

[16] Pasupa, K., Tungjitnob, S. and Vatathanavaro, S., 2020. Semi-supervised learning with deep convolutional generative adversarial networks for canine red blood cells morphology classification. Multimedia Tools and Applications, 79, pp.34209-34226.

[17] Hegde, R.B., Prasad, K., Hebbar, H. and Singh, B.M.K., 2019. Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. Biocybernetics and Biomedical Engineering, 39(2), pp.382-392.

[18] Acevedo, A., Alférez, S., Merino, A., Puigví, L. and Rodellar, J., 2019. Recognition of peripheral blood cell images using convolutional neural networks. Computer methods and programs in biomedicine, 180, p.105020.

[19] Darrin, M., Samudre, A., Sahun, M., Atwell, S., Badens, C., Charrier, A., Helfer, E., Viallat, A., Cohen-Addad, V. and Giffard-Roisin, S., 2023. Classification of red cell dynamics with convolutional and recurrent neural networks: a sickle cell disease case study. Scientific Reports, 13(1), p.745.

[20] Zhu, Ziquan, Zeyu Ren, Shui-Hua Wang, Juan M. Górriz, and Yu-Dong Zhang. "RDNet: ResNet-18 with dropout for blood cell classification." In Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications: 9th International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2022, Puerto de la Cruz, Tenerife, Spain, May 31–June 3, 2022, Proceedings, Part I, pp. 136-144. Cham: Springer International Publishing, 2022.