INTERNATIONAL JOURNAL OF
ARTIFICIAL INTELLIGENCE AND EMERGING TECHNOLOGY

Print ISSN
2735-4792

VOLUME 2, ISSUE 1, 2023, 15 – 26.

Online ISSN
2735-4806

Sickle Cell Anaemia Detection using Deep Learning

Hussien, shereen A[1] , Amer, Mostafa A.[2] , Ibrahim, Doaa [3]

[1] Assistant Professor, computer science, faculty of computers and artificial intelligence, fayoum university, Email: sam26@fayoum.edu.eg

[2] BSc, Bioinformatics, faculty of computers and artificial intelligence, fayoum university

[3] BSc, Bioinformatics, faculty of computers and artificial intelligence, fayoum university

Abstract

Many disorders, including sickle cell anaemia which causes periodic bouts of pain and severe, pronounced anaemia, result in red blood cell (RBC) deformation. It takes longer to monitor patients with these disorders since peripheral blood samples must be examined under a microscope. The observation of isolated RBCs is subjective; hence the error rate is considerable and an expert is needed to perform this approach, SCD can be adequately managed and the death rate can be decreased with early detection. Therefore, this work proposes a deep learning method for sickle cell detection based on Convolutional neural networks (CNN). VGG model differentiate between three classes of red blood cells which are circular (normal), elongated (sickle cells), and other blood content. it is applied on ERYTHROCYTESIDB dataset for validation. A comparison of the results showed that the proposed model is superior for the diagnosis of sickle cell anaemia with 99.4% of overall accuracy.

Keywords: Sickle Cell Anaemia, Red Blood Cell, CNN, VGG, Transfer Learning

Introduction

Sickle cell anaemia is a serious disease in which the body makes sickle-shaped ("C"-shaped) red blood cells. Normal red blood cells are disk-shaped that

* *Corresponding author*: author@institute.xxx

Print ISSN
2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN
2735-4806

contain the protein haemoglobin and move easily through your blood vessels. Sickle cells contain abnormal haemoglobin that causes the cells to have a sickle shape, which don't move easily through the blood vessels. These sickle cells are stiff and sticky and tend to form clumps and get stuck in the blood vessels [1]. The clumps of sickle cells block blood flow in the blood vessels that lead to the limbs and organs. Blocked blood vessels can cause pain, serious infections, and organ damage. sickled cells that cause haemolysis of blood cells, anaemia, painful episodes, organ damage, and in some cases death.

As per WHO's report [2], Approximately 5% of the world's population carries trait genes for haemoglobin disorders, mainly sickle cell disease and thalassemia. The percentage of people who carry these genes is as high as 25% in some regions. Over 300,000 babies with severe haemoglobin disorders are born each year. Early detection of SCD can help to reduce the mortality and manage the disease effectively. Therefore, so may methods have been developed to detect the sickle cell disease and the carrier states with high sensitivity and specificity.

Related work

Most of the works reported in previous literatures are based on multi-stage workflow, including steps such as image pre-processing, cell segmentation, feature extraction and single cell classification using various machine learning models.

Moreira et al. [3] used a simple segmentation procedure based on global thresholding and morphological filtering which was enough to isolate the cell shape from the rest in blood smear microscopy images. Different classifiers such as Naive Bayes, K-NN, and SVM. applied on erythrocytesIDB dataset with randomized k-fold cross-validation and the best subset of dataset features, ranked using univariate ANOVA-based selection, and the best classifier for each of them.

Print ISSN
2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN
2735-4806

Fadhel et al. [4] focused on counting the numbers of normal and abnormal RBC cells using two image processing techniques; circular Hough transform and watershed segmentation based on such different shapes, such counting is practically important in assessing the level of anaemia disease danger.

Font et al. [5] used a Chan-Vese active contour model to segment objects in images due to its ability to obtain a wider range of convergence and handle topological changes in a natural manner. The obtained borders are analysed to determine whether the detected objects are cells, cell clusters, or other objects in the image.

Epifanio et al. [6] presented an elastic metric for the deformation in the closed planar curves is used, taking advantage of the stretching and bending features possibilities by means of the square root velocity function (SRVF).

Petrovi et al. [7] proposed a method to select the classification and features with best performance for diagnostic support through peripheral blood smear images of red blood cells. Randomized and Grid search are used to evaluate the parameters from the resulting data in the feature extraction from blood cells.

Zhang et al. [8] extended the U-Net architecture by adding deformable convolution to implement the deformable U-Net (dU-Net) model for RBC semantic segmentation, specifically to address the challenge of inter-cell variation and the needs for spatial invariance. Experimental results show that dU-Net achieves better performance than other segmentation models.

Dataset

The erythrocytesIDB [9] is a dataset of microscopic images of the blood smears. it contains a voluntary donated image samples of patients with sickle cell anaemia which is classified by a specialist from “Dr. Juan Bruno Zayas” Hospital General in Santiago de Cuba, Cuba. The data set consists of 1302 image, each cell is 80 × 80 pixels in size, distributed in three files

INTERNATIONAL JOURNAL OF
ARTIFICIAL INTELLIGENCE AND EMERGING TECHNOLOGY
Print ISSN 2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN 2735-4806

erythrocytesIDB1, erythrocytesIDB2 and erythrocytesIDB3 [5]. Some samples of these images are shown in figure 1.

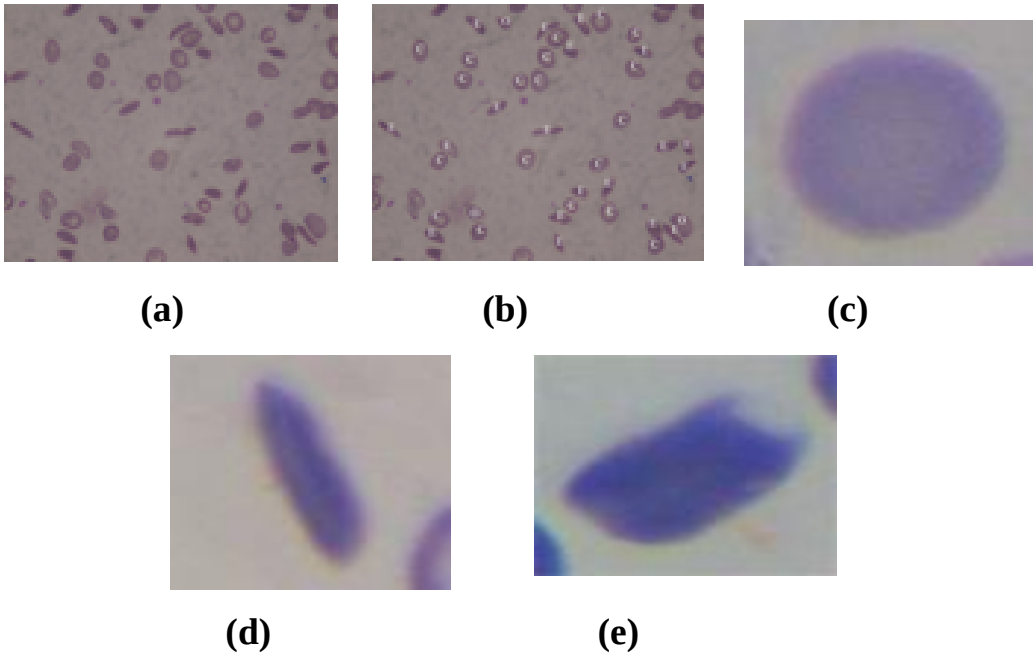

Figure 1 (a) original image. (b) image cells classified by specialist. (c) circular cell. (d) elongated cell. (e) other.

Proposed system

The overall architecture of the proposed model with main phases is presented in Figure. 2.

Pre-processing Phase

Data is rescaled by normalization scaler (1. /255) to limit range of values between 0 and 1. This rescaling technique [10] is done to reduce the impact of outliers, and ensure that all features contribute equally to the model and to prevent features with larger values from dominating the model. Followed by resizing of image input in 224×224 to be fed to VGG network.

INTERNATIONAL JOURNAL OF
ARTIFICIAL INTELLIGENCE AND EMERGING TECHNOLOGY
Print ISSN
2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN
2735-4806

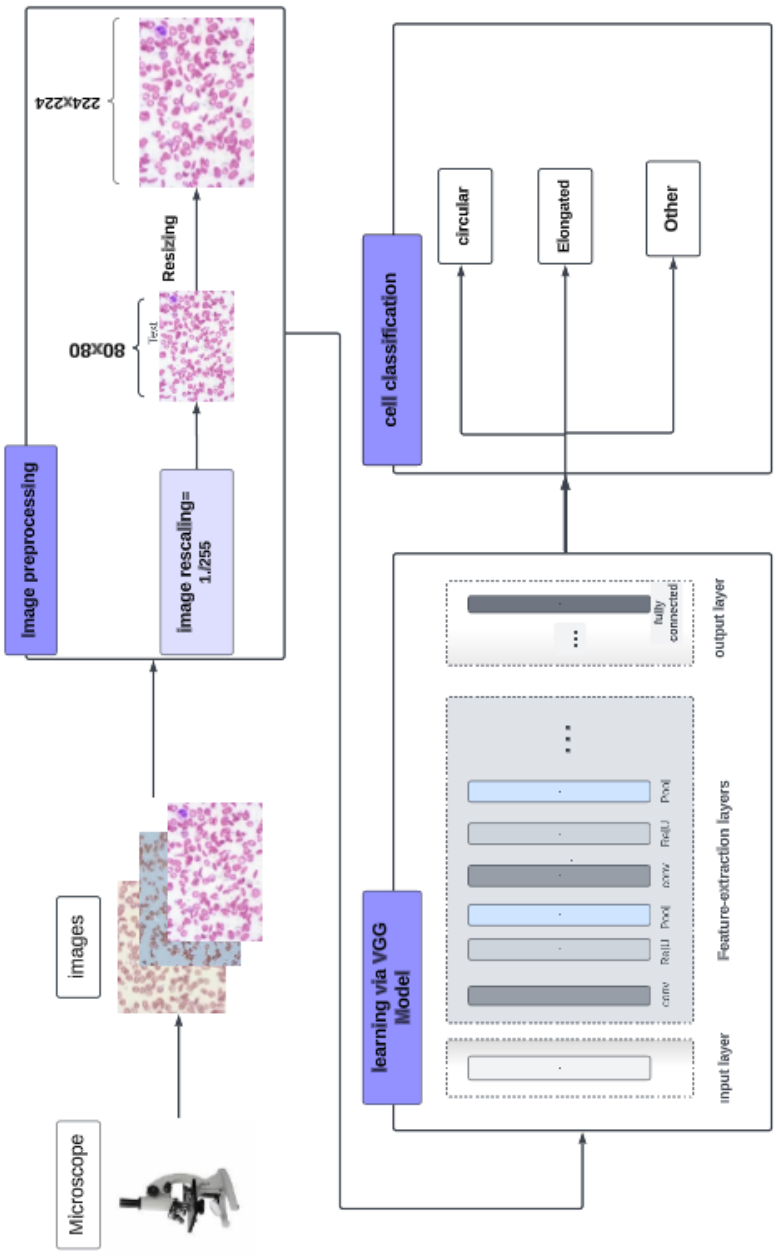

Figure 2 Proposed Model Architecture

Model Building Phase

The VGG model [11] is a classical convolutional neural network architecture used in pattern recognition and achieved top results for image classification. This work proposed a transfer learning VGG model with a pretrained network with ImageNet without final classification layer. these best weights then are concatenated the with deep learning to enhance the classification of blood cells.

INTERNATIONAL JOURNAL OF
ARTIFICIAL INTELLIGENCE AND EMERGING TECHNOLOGY
Print ISSN 2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN 2735-4806

The transfer learning allows training a given dataset using a pretrained CNN model to overcome the difficulty and time consumption problems [12-13].

The structure of proposed VGG model contains four convolutional layers, interleaved with max-pooling, normalization layers, with activation function layers [14]. More specifically, after the input layer, there are convolution layer with 32 kernels of size 3×3. The structure repeats but changes in the number of convolution layers and number of kernels. Then, a flattening layer turns the matrix into one vector. After all the convolution layers, two dense layers are added, each with 512 hidden nodes. Adding these dense layers to the convolutional base, which has its weights frozen while additional dense layers are trained. The final dense layer with soft-max of 3 nodes is to generate the output cells. The proposed model compiled with the Adam optimizer [15] is a stochastic gradient descent method with default learning rate and parameters. The proposed model is fitted at epochs where value equals 100, train batch size equals 20, and validate batch size equals 20 and it have been early stopped at epoch 68 and dataset was split into training, validating and testing sets with ratio 50/30/20 to avoid the overfitting problem. The best-saved weights were loaded after those testing images were loaded.

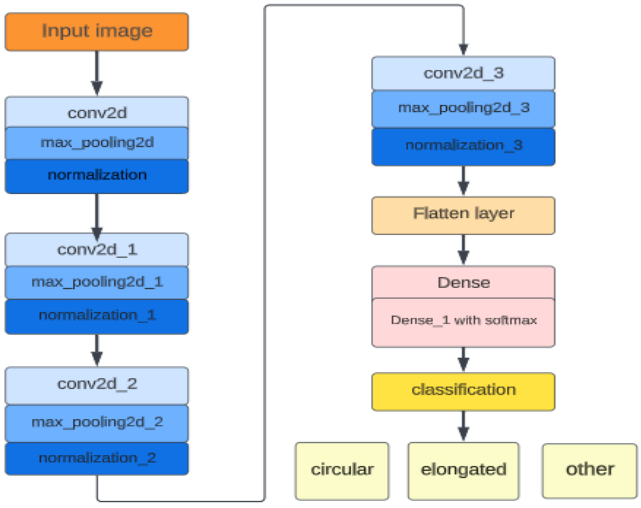

Figure 3 the structure of the VGG model

INTERNATIONAL JOURNAL OF
ARTIFICIAL INTELLIGENCE AND EMERGING TECHNOLOGY
Print ISSN 2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN 2735-4806

Discussion and Experimental Results

This work demonstrates that the highest accuracy is achieved with erythrocytesIDB dataset using VGG model of transfer learning. The dataset is divided into a ratio of 8:2. All experiments are implemented using Colab [16] that is provided by Google, the Keras framework [17] that can run on top of TensorFlow, and the Python programming language. All experiments were conducted on a 12 GB NVIDIA Tesla K80 GPU (Graphical Processing Unit) and 12 GB of RAM.

The performance of the model is evaluated by precision (1) which is the ratio between TP and the total of positives classification, recall (2) which is the ratio between TP and the total of positive class values, F-score (3) which is the harmonic average of the precision and recalls and accuracy (4) which is the ratio between the true classifications and the total number of samples. These performance equations values are shown in Tables 1 and Table 2.

Table 1 performance Equations

Assessment Equation	Equation No.
Precision = TP / (TP + FP)	(1)
Recall = TP / (TP + FN)	(2)
F-score = 2 * (Precision*Recall / Precision + Recall)	(3)
Accuracy = (TP + TN) / (TP + TN + FP + FN)	(4)

Where:

True-positives (TP), the model predicts the positive class correctly.

True-negatives (TN), the model correctly classifies the negative class.

False-positives (FP), the model predicts the positive class incorrectly.

INTERNATIONAL JOURNAL OF
ARTIFICIAL INTELLIGENCE AND EMERGING TECHNOLOGY
Print ISSN 2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN 2735-4806

False-negative (FN) classifications, the model predicts the negative class incorrectly.

Table 2 performance Measure Values

Cell shape	Precision	Recall	F-score
circular	0.96	1.00	0.98
Elongated	1.00	1.00	1.00
Other	1.00	0.96	0.98
Accuracy			0.99

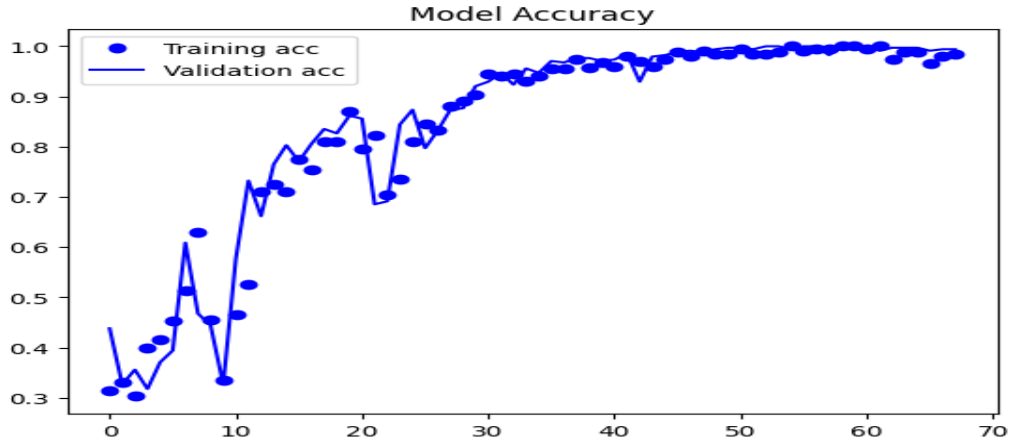

Figure 4 Accuracy of the model

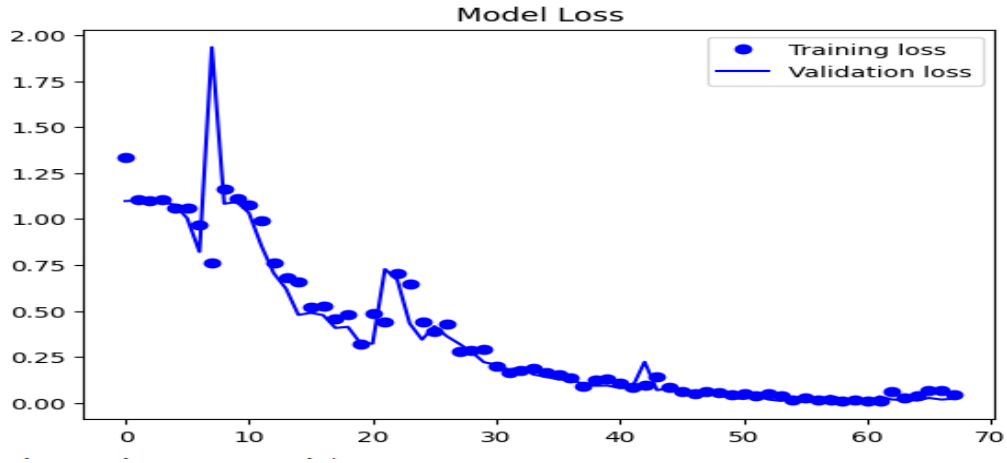

Figure 5 Model loss

INTERNATIONAL JOURNAL OF
ARTIFICIAL INTELLIGENCE AND EMERGING TECHNOLOGY
Print ISSN
2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN
2735-4806

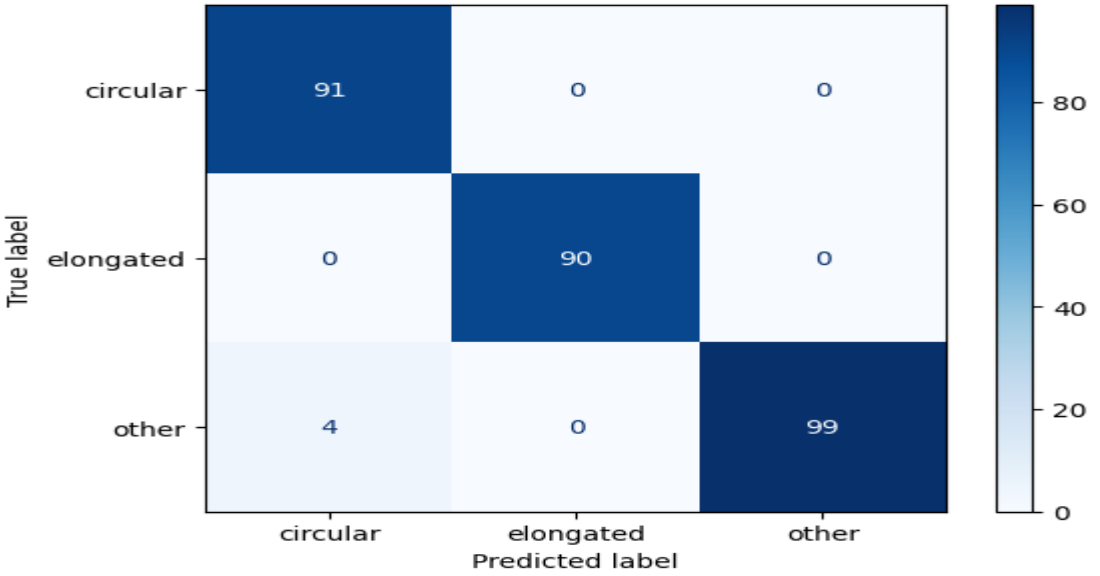

Figure 6 Model confusion matrix

Figures 4,5 show loss and accuracy for the model that are calculated during different epochs in the training phase. Figure 6 shows the confusion matrix that allows visualization of the performance of the model. Also, the proposed model performance is compared with other models that work on the same dataset as shown in Table 3.

Conclusion

Sickle Cell Anaemia is a haematological condition that obstructs blood vessels and can result in painful episodes and even death. The main function of red blood cells (RBCs) is to carry oxygen throughout the entire body. When sickle cell anaemia affects RBCs, they form the shape of a sickle that makes it difficult to pass through the bloodstream, which reduces the flow of oxygen. The first stage to a proper diagnosis, which helps determine the danger level of sickle cell anaemia, is the accurate classification of RBCs. This paper proposed a VGG-16 model to classify three classes of red blood cells which are circular (normal), elongated (sickle cells), and other blood content in blood smear microscopy images. This model was tested with a dataset consisting of 626 images, each image contains only one cell in evidence, appropriately labelled. Different

Print ISSN
2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN
2735-4806

optimizers and learning rate schedulers are explored, and the best testing accuracy achieved 99.4%, surpassing all network accuracies previously reported to detect sickle cell anaemia.

Table 3 classification accuracies summary

Paper name	Dataset	Method	Accuracy
Morphological analysis and classification of erythrocytes in microscopy images [3]	Erythrocytes IDB	Naïve Bayes, KNN, SVM	93% 93% 94%
Diagnosis support of sickle cell anaemia by classifying red blood cell shape in peripheral blood images [5]	Erythrocytes IDB	CSF and ESF	0.80
Red Blood Cells Detecting Depending on Binary Conversion at Multi Threshold Values [18]	Erythrocytes IDB	binary conversion at multi-threshold	88.45%
Machine learning based Diagnosis and Classification of Sickle Cell Anaemia in Human RBC [19]	Erythrocytes IDB	Random forest, Naïve Bayes, SVM, Logistic regression	92% 88% 90% 90%
Deep Convolutional Neural Network Model for Detection of Sickle Cell Anaemia in Peripheral Blood Images [20]	Erythrocytes IDB	PCNN-48, VGG19	78.67% 97.48%
The Proposed Model	Erythrocytes IDB	VGGNet	99.4%

Reference

1. “Symptoms and diagnosis of blood disorders”. MSD Manual Consumer Version. Available at: https://www.msdmanuals.com/home/blood-disorders/symptoms-and-diagnosis-of-blood-disorders.
2. "World Health Organization Sickle-cell anaemia". Available at: https://www.afro.who.int/health-topics/sickle-cell-disease.

Print ISSN 2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN 2735-4806

3. Rodrigues, L. F., Naldi, M. C., & Mari, J. F. "Morphological analysis and classification of erythrocytes in microscopy images". In Proceedings of the 2016 Workshop de Visão Computacional, Campo Grande, Brazil, 2016.
4. Fadhel, M. A., Humaidi, A. J., & Oleiwi, S. R. "Image processing-based diagnosis of sickle cell anemia in erythrocytes". In Annual conference on new trends in information & communications technology applications (NTICT). IEEE, 2017.
5. Delgado-Font, W., Escobedo-Nicot, M., González-Hidalgo, M., Herold-Garcia, S., Jaume-i-Capó, A., & Mir, A. "Diagnosis support of sickle cell anemia by classifying red blood cell shape in peripheral blood images". Medical & biological engineering & computing, vol. 58, no. 6, 2020.
6. Epifanio, I., Gual-Arnau, X., & Herold-Garcia, S. "Morphological analysis of cells by means of an elastic metric in the shape space". International Society for Stereology, 2020.
7. Petrović, N., Moyà-Alcover, G., Jaume-i-Capó, A., & González-Hidalgo, M. "Sickle-cell disease diagnosis support selecting the most appropriate machine learning method: Towards a general and interpretable approach for cell morphology analysis from microscopy images". Computers in Biology and Medicine, 2020.
8. Zhang, M., Li, X., Xu, M., & Li, Q. "Automated semantic segmentation of red blood cells for sickle cell disease". IEEE journal of biomedical and health informatics, vol. 24, no. 11, 2020.
9. ErythrocytesIDB Dataset, version 2. Available at: http://erythrocytesidb.uib.es/
10. Danon, D., Arar, M., Cohen-Or, D., & Shamir, A. "Image resizing by reconstruction from deep features". Computational Visual Media, vol. 7, no.4, 2021.
11. Simonyan, K., & Zisserman, A. "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.
12. Barros, B., Lacerda, P., Albuquerque, C., & Conci, A. (2021). "Pulmonary COVID-19: learning spatiotemporal features combining CNN and LSTM networks for lung ultrasound video classification". Sensors, vol. 21, no. 16, 2021.
13. Shen, S., Sadoughi, M., Li, M., Wang, Z., & Hu, C. "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries". Applied Energy, 2020.
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. "Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

Print ISSN 2735-4792
VOLUME 2, ISSUE 1, 2023, 15 – 26.
Online ISSN 2735-4806

15. Kingma, D. P., & Ba, J. L. "Adam: a method for stochastic optimization". arXiv preprint arXiv:1412.6980, 2015.
16. Bisong, E. Google colaboratory. "Building machine learning and deep learning models on Google cloud platform". Berkeley, CA: Apress, 2019.
17. Keras Team. Getting started. Keras.io
18. Lefta, A. S., Daway, H. G., & Jouda, J. "Red Blood Cells Detecting Depending on Binary Conversion at Multi Threshold Values". Al-Mustansiriyah Journal of Science, vol. 33, no. 1, 2022.
19. Sen, B., Ganesh, A., Bhan, A., Dixit, S., & Goyal, A. "Machine learning based Diagnosis and classification of Sickle Cell Anemia in Human RBC". In Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021.
20. Dada, E. G., Oyewola, D. O., & Joseph, S. B. "Deep convolutional neural network model for detection of sickle cell anemia in peripheral blood images". Communication in Physical Sciences, vol. 8, no. 1, 2022.