STUDY OF MICROSCOPIC BLOOD SAMPLES FOR EARLY DETECTION OF WHITE BLOOD CELL DISEASES

Otabek Ismailov
Professor, International School of Finance Technology and
Science Institute (ISFT)

Ravshan Mallayev
Tashkent State Pedagogical University

Tulaganova F.K.
Tashkent University of Information Technologies named
after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan.

Miralisher Ismoilov
Minister of Agriculture of the Republic of Uzbekistan,
Tashkent, Uzbekistan

Abstract

The immune system is the natural guard of the human body against various diseases due to viruses and bacteria. White Blood Cells are centrally involved in the body's defense mechanism. WBC features extracted from blood samples can be used in diagnosing blood disorders. These features include chromatic, geometric, and textural features of WBC nucleus. Traditional manual diagnosis is error-prone, subjective, and time-consuming. Considering these challenges, this research introduces a more advanced automated algorithm in the classification of microscopic blood sample datasets. This proposed algorithm turned out to have a high accuracy in classification of 90.79%. The results obtained from the experiment showed that it had been trained even faster at higher accuracy of 94%. This algorithm, utilizing artificial intelligence techniques, significantly enhances WBC type classification efficiency and hence assists hematologists in screening several blood disorders at an early stage, thus significantly improving the limitations of manual diagnosis.

Keywords: Immune system, white blood cells (WBCs), chromatic, geometric and textural features, traditional manual diagnostics, automated algorithm.

Introduction

Blood is a mine of information, which can be isolated and learned to understand a person's health. Blood comprises up to 45% of red blood cells and 55% plasma, along with less than 1% white blood cells and platelets. Plasma carries nutrients, proteins, hormones, and minerals via the circulatory system not only to various organs but to different parts of the body, moving unwanted other elements out of the body as waste products. RBCs exist from four to six million per microliter. The most important constituent of RBCs, hemoglobin, carries oxygen to all parts of the body. Platelets help in blood clotting and range from 150,000 to 450,000 per microliter in a normal individual. The WBC rate exists between 4500 to 11,000 per microligiter in a normal person. WBCs are the key main components of immune cells, providing immunity to the body against the viruses, resistance to diseases, and protection from various infections like fungi, bacteria, and viruses. White blood cells are produced inside the bone marrow, lymphoid tissue, and some important glands. There are mainly five types of WBCs: eosinophils, lymphocytes, monocytes, neutrophils, and basophils. This WBC count, when increased or decreased, leads to a number of chronic and fatal diseases [5]. In addition, a large number of diseases, such as bacteria, leukemia, immunodeficiency syndrome, and infections, emerge as an outcome of an increase or a decrease in four WBCs: eosinophils, lymphocytes, monocytes, and neutrophils. The number of images available for basophils is few due to their low incidence (0–3%). Owing to the unavailability of such images and their low significance, we will address the other four types of WBCs in this paper. The rise in the neutrophil cells in blood can be attributed to bacteria, endotoxins, exotoxins, and fungi [6,7]. The cells of lymphocyte increase due to diseases like hepatitis, whooping cough, viruses, bordetella, leukemia, and brucellosis; other diseases like HIV, chickenpox, rubeola, and tuberculosis reduce lymphocyte cells. Malaria, listeriosis, and viral and bacterial infections raise monocyte cells [8]. Allergies, parasites, and atopic diseases increase eosinophil cells [9]. The WBC type and number are diagnosed with a blood test (hemogram) on peripheral smear. Blood examination is performed on a microscopic slide by spreading the blood on it and then viewing the WBCs under the microscope [10]. The WBCs can be further subdivided into polynuclear cells, including eosinophils, neutrophils, and basophils and granulocytic cells including lymphocytes and monocytes. The following are some simple ways to identify and count the types of WBCs on a stained blood smear for the purpose of diagnosis when conducting blood sample tests: one, identification of the type and location of micrographs, two, description of the shape and color for each cell. Therefore, it involves much effort, is pretty time-consuming, and is prone to different diagnostic opinions of blood experts due to errors. Therefore, the computer-aided automatic diagnostic system will help sort out automatic diagnostic issues, which will, in turn, reduce the errors involved in manual diagnosis and bring correct diagnostic accuracy. This dataset includes all image slides whose noises result from the mixture of blood samples with components like Wright's stain or a mixture of methylene blue or eosin, red dye.

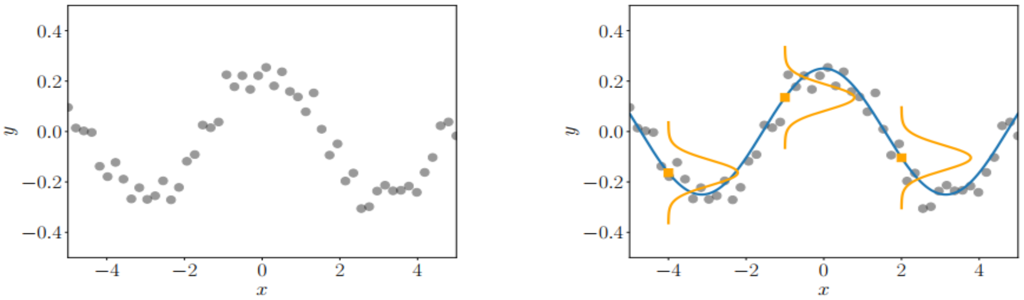
It also contains all the slide images with noise, which originates from the variety of microscope devices, from their accuracy and optical reflections, or from the techniques of storing the dataset. These noises hinder arriving at an ultra-accurate diagnosis on the input images. Biomedical image processing hence begins with pre-treatment. Two filters have been applied in the present work: median for noise removal and Laplacian for edge contrast of WBCs.

First, the images were optimized with an average filter of 4 × 4 pixels and passed sequentially to process all image pixels. The average filter smooths the image by eliminating disparities between nearby pixels and replacing each center (target) pixel with an average of 15 neighboring pixels. The process is carried out continuously until the whole image is processed. Equation (1) shows how the average filter works.

$$Average\ (M\frac{1}{L}\sum_{-t=0}^{M-1} y(M-1)) \qquad (1)$$

In this process, regression processes are implemented, which can be the basis of the mathematical Laplace filter. Yanni Regression is a fundamental problem in machine learning, and regression problems appear in a variety of research areas and applications, time series analysis (eg, system identification), control. It is part of regression algorithms. It requires different solutions to try, try:

In this process, regression processes are implemented, which can be the basis of the mathematical Laplace filter. Yanni Regression is a fundamental problem in machine learning, and regression problems appear in a variety of research areas and applications, time series analysis (eg, system identification), control. It is part of regression algorithms. It requires different solutions to try, try:



(a) (b)

Figure 1 (a) Dataset; (b) possible solution to the regression problem

1 Choice of the model (type) and the parametrization
2 Finding good parameters.
3 Overfitting and model selection.
4 Relationship between loss functions and parameter priors.
5 Uncertainty modeling
The following formula is the basis for isolating the main blood cells.

$$p(y \mid \mathbf{x}) = N\, y \mid f(\mathbf{x}), \sigma 2) \qquad (2)$$

where $\epsilon \sim N\ 0, \sigma 2$ is independent, identically distributed (i.i.d.) Gaussian measurement noise with mean 0 and variance $\sigma 2$. Our objective is to find a function that is close (similar) to the unknown function f that generated the data and that generalizes well. In this chapter, we focus on parametric models, i.e., we choose a parametrized function and find parameters $\boldsymbol{\theta}$ that “work well” for modeling the data. For the time being, we assume that the noise variance $\sigma 2$ is known and focus on learning the model parameters $\boldsymbol{\theta}$. In linear regression, we consider the special case that the parameters $\boldsymbol{\theta}$ appear linearly in our model. An example of linear regression is given by

$$p(y \mid \mathbf{x}, \boldsymbol{\theta}) = N\, y \mid \mathbf{x}\, T\boldsymbol{\theta}, \sigma 2 \Longleftrightarrow y = \mathbf{x}\, T\boldsymbol{\theta} + \epsilon\ , \epsilon \sim N\ 0, \sigma 2 \qquad (3)$$

2. Parameter Estimation

Consider the linear regression setting (9.4) and assume we are given a *training set* $D := \{(\mathbf{x}1,$

*y*1), . . . , (***x****N*, *yN*)} consisting of *N* inputs ***xn*** ∈R*D* and corresponding observations/targets *yn* ∈R, *n* = 1, . . . , *N*. The corresponding graphical model is given in Figure 1. Note that *yi* and *yj* are conditionally independent given their respective inputs ***xi***, ***xj*** so that the likelihood factorizes according to

p(*Y* | *X*, ***θ***) = *p*(*y*1, . . . , *yN* | ***x***1, . . . , ***x****N*, ***θ***)

$$= \prod_{n=1}^{N} p(y_n|x_n, \theta) = \prod_{n=1}^{1^v} N(y_n|x_R\theta, \sigma^2) \qquad (4)$$

To find the optimal parameters ***θ***ML of our linear regression problem, we minimize the negative log-likelihood

- log *p*(*Y* | *X* , ***θ***) = - log NY *n* =1 *p*(*yn* | ***xn***, ***θ***) = - NX *n* =1 log *p*(*yn* | ***xn***, ***θ***) (5)

where we exploited that the likelihood (5) factorizes over the number of data points due to our independence assumption on the training set. In the linear regression model (5), the likelihood is Gaussian (due to the Gaussian additive noise term), such that we arrive at

log *p*(*yn* | ***xn***, ***θ***) = - 1 2*σ*2 (*yn* - ***x*** *T n* ***θ***)2 + const (6)

Using the results from Chapter 5, we compute the gradient of *L* with respect to the parameters as

$$\frac{dl}{d\theta} = (\frac{1}{2Q^2}(y - xQ)t(y - x\theta)) \qquad (7)$$

$$= \frac{1}{2Q^2} - \frac{d}{d\theta}(y^T y - 2y^T XQ + Q^T X^T XQ) \qquad (8)$$

$$= \frac{1}{Q^2}(-y^T X + Q^T X^T X)\epsilon R^{1*D} \qquad (9)$$

The maximum likelihood estimator ***θ***ML solves dd*L**θ*** = **0** *T* (necessary opti mality condition) and we obtain

dL d***θ*** = **0** *T* (⇐⇒ 9.11c) ***θ*** *T* ML***X*** *T****X*** = ***y*** *T****X*** (10)

⇐⇒ ***θ*** *T* ML = ***y*** *T****X***(***X*** *T****X***)-1 (11)

⇐⇒ ***θ***ML = (***X*** *T****X***)-1***X*** *T****y*** . (12)

We could right-multiply the first equation by (***X*** *T****X***)-1 because ***X*** *T****X*** is positive definite if rk(***X***) = *D*, where rk(***X***) denotes the rank of ***X***. *Remark.* Setting the gradient to **0** is a necessary and sufficient condition, and we obtain a global minimum since the Hessian 2 ***θ****L*(***θ***) = ***X X***

We are concerned with a regression problem *y* = ***ϕ*** *T*(*x*)***θ***+*ϵ*, where *x* ∈ R and ***θ*** ∈ R*K*. A transformation that is often used in this context is

$$\boldsymbol{\Phi}\mathbf{(x)} = \begin{array}{ll} \boldsymbol{\phi}o & (x) \\ \boldsymbol{\phi}1 & (x) \\ \boldsymbol{\phi}k-1 & (x) \end{array} = \begin{array}{c} 1 \\ X \\ X\ k-1 \end{array} \epsilon R^k \qquad (13)$$

This means that we "lift" the original one-dimensional input space into a *K*-dimensional feature space consisting of all monomials *xk* for *k* = 0, . . . , *K* - 1. With these features, we can model polynomials of degree ⩽ *K*-1 within the framework of linear regression: A polynomial of degree

$$\boldsymbol{\Phi} := \begin{bmatrix} \boldsymbol{\phi}^\top(\boldsymbol{x}_1) \\ \vdots \\ \boldsymbol{\phi}^\top(\boldsymbol{x}_N) \end{bmatrix} = \begin{bmatrix} \phi_0(\boldsymbol{x}_1) & \cdots & \phi_{K-1}(\boldsymbol{x}_1) \\ \phi_0(\boldsymbol{x}_2) & \cdots & \phi_{K-1}(\boldsymbol{x}_2) \\ \vdots & & \vdots \\ \phi_0(\boldsymbol{x}_N) & \cdots & \phi_{K-1}(\boldsymbol{x}_N) \end{bmatrix} \in \mathbb{R}^{N\times K}, \qquad (14)$$

where *Average* (*M*) denotes the optimized image (output), *y*(*M* − 1) denotes the previous input, and *M* denotes the image's pixel count. Because of the blurred edges between WBCs and other cells, the Laplacian filter was used, which detects the edges of WBCs, shows them clearly, and distinguishes them from other blood cells. The Laplacian filter's action mechanism on the region of interest (WBCs) is described in Equation (2)

$$\nabla^2 f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y} \qquad (15)$$

where $\nabla^2 f$ refers to a second order differential equation, and *x* and *y* refer to the coordinates in 2D matrices.

Lastly, to obtain an enhanced and clear image, the image produced using the Laplacian filter is subtracted from the image produced using the averaging filter, as described in Equation (3)

Imege optimized $=Average(M)-\nabla 2\ f$

$$LBp_{(x_i,y_C)R,p} = \sum_{p=0}^{1-1} s\,((8_1 - 8_c) * 2^p \qquad (16)$$

where g_c denotes the central pixel, g_p denotes the neighboring pixel, *R* denotes the radius around the central pixel, and *P* denotes the number of neighbors. The binary threshold function *x* is defined according to Equation (5)

$$S_{(x)} = \begin{cases} 0x < 0 \\ 1, x \geq 0 \end{cases} \qquad (17)$$

Secondly, the GLCM algorithm is an algorithm used to extract texture features from the ROI (WBCs). The algorithm distinguishes between a smooth texture and a rough texture using spatial information; the texture is smooth when the pixels are of similar values. In contrast, the texture is coarse when the pixels are of different values. The pairwise correlations between pixels are determined by the distance d and directions θ between 1 pixel, where θ represents four directions: 0°, 45°, 90°, and 135°. The relationship between distance and directions is when d = 1, then θ between pixels is θ = 0 or θ = 90; when d = √2, then θ between pixels is θ = 45 or θ = 135. This algorithm extracted 13 representative features.

3. Part of the software

Creating a deep learning based blood cell analysis program using Convolutional Neural Networks (CNN) can be challenging, we have provided example code in Python using TensorFlow and Keras library. This example uses the opencv library for image processing. To use the program, you must install all the necessary libraries:

Sample code:

Python

```
import os
import cv2
import numpy as np
import tensorflow as tf
from tensorflow.keras import layers, models
from sklearn.model_selection import train_test_split
# Loading data (assuming you have a dataset of blood cell images)
def load_data(data_dir):
images = []
labels = []

for label in os.listdir(data_dir):
label_path = os.path.join(data_dir, label)
for filename in os.listdir(label_path):
img_path = os.path.join(label_path, filename)
img = cv2.imread(img_path)
img = cv2.resize(img, (128, 128))  # Image size can be changed
images.append(img)
labels.append(int(label))  #Labels are assumed to be integers
return np.array(images), np.array(labels)
data_directory = "/path/to/your/images "
images, labels = load_data(data_directory)
# Dividing data into training and test set
X_train, X_test, y_train, y_test = train_test_split(images, labels, test_size=0.2,
random_state=42)
# Creating a CNN Model
model = models.Sequential()
model.add(layers.Conv2D(32, (3, 3), activation='relu', input_shape=(128, 128, 3)))
model.add(layers.MaxPooling2D((2, 2)))
model.add(layers.Conv2D(64, (3, 3), activation='relu'))
model.add(layers.MaxPooling2D((2, 2)))
model.add(layers.Conv2D(64, (3, 3), activation='relu'))
model.add(layers.Flatten())
model.add(layers.Dense(64, activation='relu'))
model.add(layers.Dense(1, activation='sigmoid'))
# Compiling the model
model.compile(optimizer='adam',
loss='binary_crossentropy',
metrics=['accuracy'])
# Model training
model.fit(X_train, y_train, epochs=10, validation_data=(X_test, y_test))
```

```
# Accuracy assessment on test data set
test_loss, test_acc = model.evaluate(X_test, y_test)
print(f'Test accuracy: {test_acc}')
```

Please note that this code is a basic example and should be adapted to this task. When applying the program to other tasks, it is required in accordance with the requirements and data must be adapted to the specific task. It also requires a sufficient amount of data for machine learning and correct image labeling.

3. Visual analysis of the obtained results.

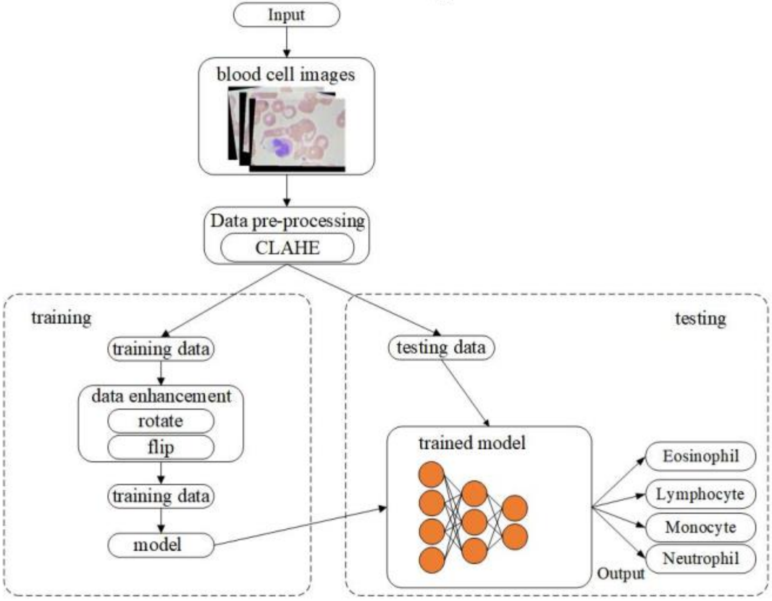
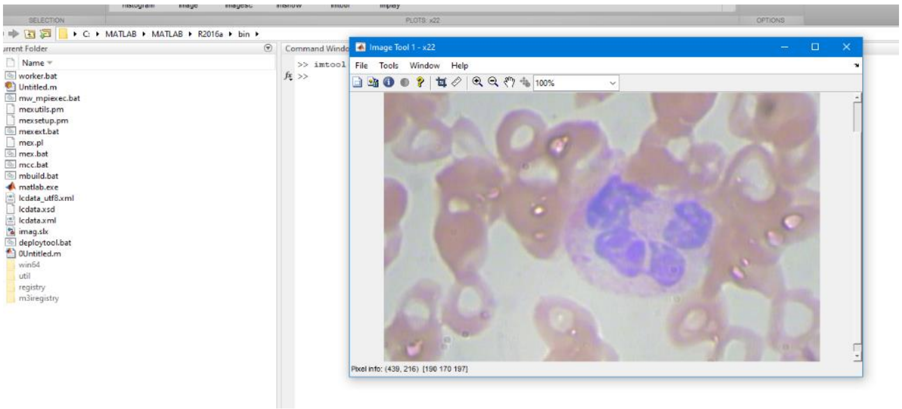


Figure 2 below is the blood cell image recognition algorithm recognition process suggested in this article

It is considered an algorithm that saves damage to the recognition of the images of the called cells described in the selected article. On the basis of this algorithm, they restored the effective sequence of the process. First, the image dataset was divided into pre-training and test data; next, reading data is said more to get updated reading data; then the resulting call information is analyzed in a cellular image recognition network. Last, the test data is outputted to the test network sample and classifies captured images.



```
>> imtool
>>
```

```
Pixel info: (439, 216)  [190 170 197]
```

Fig 3: Cancer image output

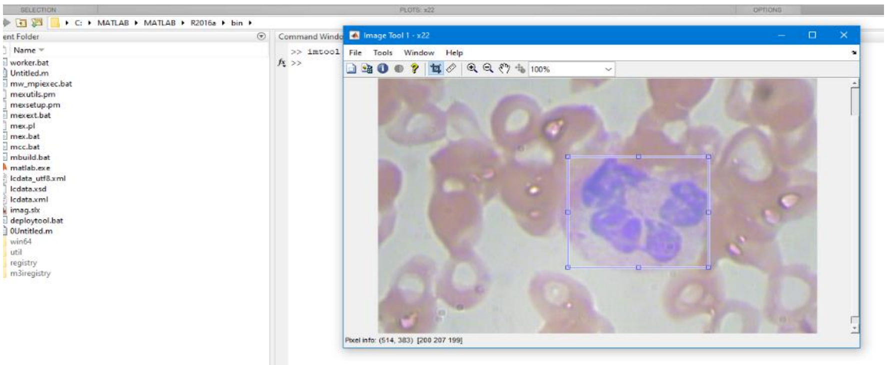


Fig 4: Determining the disease in the image

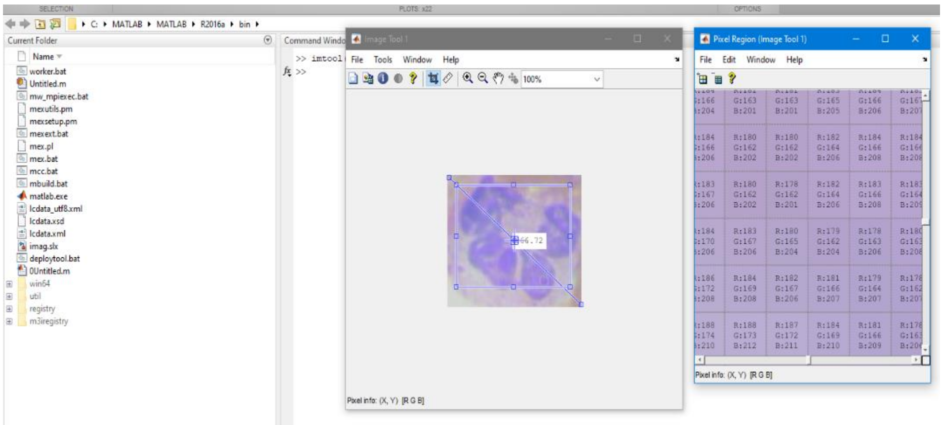


Fig 5. diagnosis of blood cell imaging

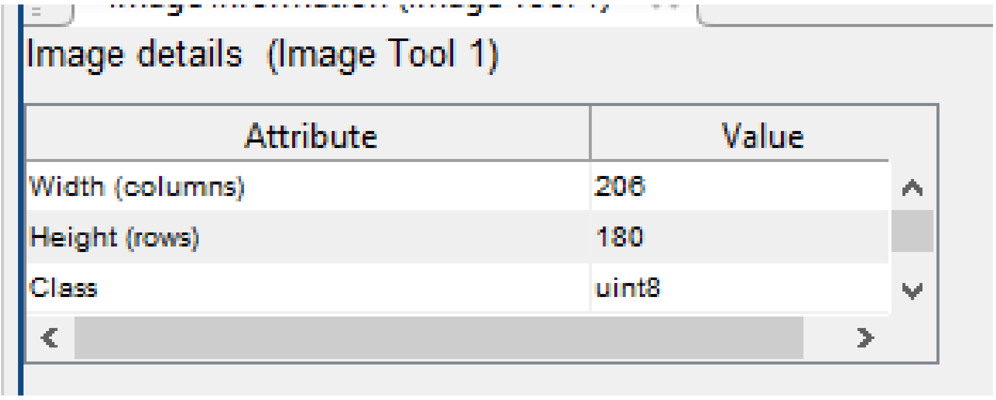


Image details (Image Tool 1)

Attribute	Value
Width (columns)	206
Height (rows)	180
Class	uint8

Fig 6. describe the image

Below is the sequence of mathematical calculation of the obtained results

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} * 100\% \qquad (18)$$

$$Sensitivity = \frac{TP}{TP+FN} * 100\% \qquad (19)$$

$$Specificity = \frac{TN}{TN+FP} * 100\% \qquad (20)$$

$$AUC = \frac{TruePositive\ Rate}{False\ Positive\ Rate} = \frac{Sensitivity}{Specificity} \qquad (21)$$

where the true positive (TP) is the unhealthy WBCs that have been correctly diagnosed, true negative (TN) is the healthy WBCs from correctly diagnosed normal patients, false negative

(FN) is the blasted WBCs diagnosed as normal, and false positive (FP) is a normal WBC count diagnosed as blasted WBCs.

4. Experimental results and analysis

Image size	Learning rate	Bacth size	epoch	Accuracy
224*224	0.001	32	200	92.3%
112*112	0.001	32	200	97.3%
224*224	0.0001	32	200	96.5%
224*224	0.001	64	200	96.9%
224*224	0.001	32	250	97.8%
224*224	0.0001	32	260	99.6%

Table 1 of accuracy gains versus the proposed algorithm used in cell image classification using a database.

Classification of cell images and obtaining a 90% accuracy data set. The accuracy obtained in the data set was 90.79%. There were also results on algorithm accuracy and computation time. Experimental results showed that the proposed algorithm was trained faster and achieved 94% accuracy on the dataset.

Conclusions

If hematologists find any anomalies in the immune system through blood tests, they would prescribe a peripheral smear test to understand vital information about the immune system and its disorders. This test relies on the manual examination of all WBC subtypes, which is error-prone and laborious. Artificial intelligence techniques can help overcome these challenges. In this work, high-speed diagnosis analytical systems were proposed for microscopic blood sample datasets. Averaging and Laplace filters are designed to enhance WBC contrast and reduce noise. Efficiency for the proposed algorithm is very notable, with a classification accuracy of 90.79% on the dataset. This algorithm trains faster and gives an accuracy of 94% in experimental results. Basically, it is composed of two components: feature map extractions by CNNs and classification of deep features using SVM. These hybrid techniques returned outstanding results in the detection of certain diseases in WBCs. Future research will extend this work by incorporating features from CNNs with additional datasets and hand-crafted features for better improvement in accuracy in WBC-type classification.

References

1. Bhuvan, C., Bansal, S., Gupta, R., & Bhan, A. (2020, February). Computer based diagnosis of malaria in thin blood smears using thresholding based approach. In *2020 7th International conference on signal processing and integrated networks (SPIN)* (pp. 1132-1135). IEEE.
2. Yue, G., Zhang, S., Li, Y., Zhou, X., Zhou, T., & Zhou, W. (2023, October). Subjective quality assessment of enhanced retinal images. In *2023 IEEE International Conference on Image Processing (ICIP)* (pp. 3005-3009). IEEE.
3. Kandhari, R., Bhan, A., Bhatnagar, P., & Goyal, A. (2021, February). Computer based diagnosis of Leukemia in blood smear images. In *2021 Third International Conference on*

Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 1462-1466). IEEE.

4. Guan, Y., & Wang, Z. (2022, August). Blood Cell Image Recognition Algorithm based on EfficientNet. In *2022 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 1640-1645). IEEE.

5. Hafeez, M. A., Imran, A., Khan, M. I., Khan, A. H., Nawaz, A., & Ahmed, S. (2022, May). Diagnosis of Liver Disease Induced by Hepatitis Virus Using Machine Learning Methods. In *2022 8th International Conference on Information Technology Trends (ITT)* (pp. 154-159). IEEE.

6. Elhassan, T. A. M., Rahim, M. S. M., Swee, T. T., Hashim, S. Z. M., & Aljurf, M (2022). Feature extraction of white blood cells using CMYK-moment localization and deep learning in acute myeloid leukemia blood smear microscopic images. *IEEE Access*, *10*, 16577-16591.

7. Eshmurodov D. Iskandarova S. Tulaganova F . (January 29-31, 2024) Algorithms for detection of cells in blood images

8.Ismailov Otabek, Iskandarova Sayora, Temirova Xosiyat Farxod qizi (international scientific journal science and innovation special issue: "modern problems and prospects of development of energy supply of digital technology facilities", march 1, 2024) “ detection and differential treatment of pathologies in x-ray dental images”

9.Kamol, L. va Raj, R. J. R. (2024). To'liq qon hujayralari sonini aniqlash orqali qon sifatini ta'minlash uchun chuqur o'rganishdan foydalanish. e-Prime- Elektrotexnika, elektronika va energiya sohasidagi yutuqlar, 100450.

10.Aggraval, H. O., Gosvami, D. va Agarval, V. (2023 yil, aprel). Nuqta izohlari bilan hujayralarni aniqlash uchun cheklov qutisi. 2023 yilda IEEE 20- Xalqaro biotibbiy tasvirlash simpoziumi (ISBI) (1-4-betlar). IEEE.