Multiple Type of Acute lymphocytic leukemia blood Cancer Detection System Using CNN

Sushmita Jagtap
Student, School of Computer Engineering and Technology
MITWPU , Pune
1032211691@mitwpu.edu.in

Dr.Shamla Mantri
Professor, School of Computer Engineering and Technology
MITWPU , Pune
Shamla.mantri@mitwpu.edu.in

***Abstract*—Leukemia is a blood cancer that has two main types: acute and chronic. Diagnosing acute lymphoblastic leukemia (ALL) can be challenging. However, the subgroups for each type are lymphoid and myeloid, making four different kinds of leukemia in total. To enhance the data, seven distinct image transformation methods were used for data augmentation. A CNN framework was developed to identify all subtypes of leukemia. A multiple type detection system for acute lymphocytic leukemia blood cancer was proposed, which uses a Convolution Neural Network (CNN) model to detect a subtype of ALL through the input provided by the user in the form of a blood cell image. CNNs are now widely used for analyzing medical images, but the classical models require large image databases to achieve high accuracy. To address this problem, a powerful deep CNNs approach was proposed, which produces more precise ALL detection and improves the accuracy of the model using a large dataset. The system generates a report of the overall result, providing the user with information about the blood cancer and precautions to be taken for a particular type of blood cancer.**

Keywords—Leukemia, Lymphoma, Multiple Myeloma, Classification algorithms, Deep Learning, Convolution neural networks, Image processing, Acute lymphocytic leukemia.

I. Introduction

Image processing has become a popular method for the early detection of various medical conditions, including blood cancers such as leukemia. Leukemia is a type of blood cancer that can affect people of all ages, and it is categorized into two types based on the affected cells: myelogenous (AML) and lymphoblastic (ALL). Computer-based algorithms for image processing have made it easier to classify leukemia accurately and quickly, eliminating the potential for errors that can occur with manual detection. According to [1], the use of computer-based algorithms has greatly improved detection accuracy in the field of leukemia diagnosis. Healthcare professionals who diagnose leukemia using microscopic pictures may make mistakes due to inaccurate information or lack of knowledge in this area. Therefore, computer-based algorithms are a reliable solution for accurate diagnosis. [2]

Blood cancer can be classified into four main types, namely acute lymphoblastic leukemia (ALL), acute myelogenous leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myelogenous leukemia (CML). The classification is based on the type of white blood cells (WBCs) that are affected in the human blood .

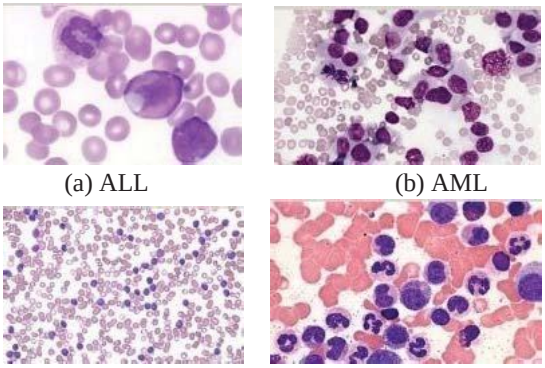
There are different types of leukemia blood cancer :

a) Acute lymphocytic leukemia (ALL)

b) Acute myelogenous leukemia (AML)

c) Chronic lymphocytic leukemia (CLL)

d) Chronic myelogenous leukemia (CML)



(a) ALL (b) AML (c) CLL (d) CML

Fig.1. Types of leukemia disease

Blood is made up of plasma and three types of blood cells, including red blood cells, white blood cells, and platelets. Each type of blood cell has a specific function. Red blood cells transport oxygen from the lungs to the body's organs and tissues, while white blood cells help fight infections and illnesses. Platelets play a vital role in controlling blood clotting and hemorrhage. Leukemia is a type of blood cancer that affects the white blood cells and causes them to multiply in an uncontrolled manner, often leading to immature cells that can interfere with the development of other blood cells, including red blood cells and platelets.

The normal ratio of white blood cells to red blood cells in the body is approximately 1:1000 [1].

Leukemia is a type of blood cancer that affects white blood cells. There are two types of white blood cells that can develop leukemia: lymphoid cells and myeloid cells. Lymphoblastic or lymphocytic leukemia is caused by lymphoid cells, while myelogenous or myeloid leukemia is caused by myeloid cells. Leukemia is classified as acute or chronic depending on how quickly the abnormal blood cells are developing. In acute leukemia, young, immature cells called blasts are the abnormal cells that multiply rapidly and do not function properly. Without prompt treatment, acute leukemia can quickly worsen. In contrast, chronic leukemia produces both young and functional adult cells, and the growth of blasts is gradual, causing the disease to progress more slowly.

we have developed a system that uses Convolutional Neural Networks (CNNs) to detect different subtypes of Acute Lymphocytic Leukemia (ALL) and provide users with information and precautions. This system can help detect ALL at an early stage and improve the chances of successful treatment. By using our proposed application, users can receive detailed information about the subtype of ALL and necessary precautions to be taken.

II. RELATED WORK

The detection of cancerous blood cells through image analysis typically involves three stages: preprocessing, feature extraction and selection, and classification. Many studies have been conducted on various types of cancer, including leukemia, lymphoma, and myeloma. Zhang et al. proposed a convolutional neural network model for direct classification of infected and uninfected cells without segmenting the cervical cells first. Zhao et al. suggested using machine learning algorithms such as CNN, SVM, and Random Forests for classifying different types of white blood cells. Foran et al. developed an image-based clinical decision support prototype that can distinguish between various hematologic malignancies and recommend therapies using a plurality justification of the cases gathered. Mahajan et al. proposed an SVM-based system for identifying Acute Lymphocytic Leukemia (ALL) using changes in texture, geometry, and histogram as classifier inputs. Markiewicz et al. proposed a classification scheme for the 17 different types of blood cells found in myelogenous leukemia using Gaussian Kernel Support Vector Machine (SVM). Halim et al. suggested an automatic technique for counting the number of ALL and acute myeloid leukemia (AML)-infected cells in a leukemia image slide, while Horie et al. demonstrated the diagnostic capacity of CNN for esophageal cancer. Other studies have proposed automated approaches using CNN for detecting skin lesions, separating infected images from healthy ones, and classifying thermal images to distinguish between breast cancer images.

III. PROPOSED METHODOLOGY

A. Block Digram

The proposed detection system for multiple types of Acute Lymphocytic Leukemia (ALL) blood cancer using a CNN architecture involves the following steps:

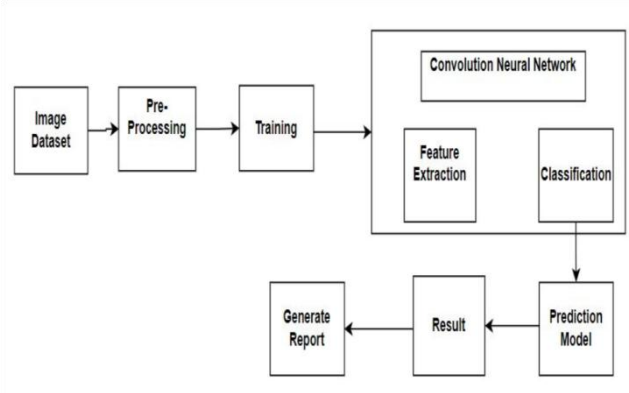


Fig2. Block Diagram of the proposed system

The system utilized an image from "Acute Lymphoblastic Leukemia (ALL) The Blood Cancer Image Database for Image Processing" as input [7]. Subsequently, the input image underwent pre-processing for enhancing image quality by performing color conversion. Next, the segmented image underwent feature extraction, where cells were examined for form, color, and texture attributes.

- Image Dataset Collection : The image dataset used in this study was obtained from the publicly available Acute Lymphoblastic Leukemia (ALL) dataset on Kaggle. The images were produced by the bone marrow laboratory at Taleqani Hospital in Tehran, Iran and consisted of 3256 PBS images from 89 patients suspected to have ALL, with blood samples properly prepared and stained by expert laboratory personnel. The ALL group was further divided into four subgroups: Early, Benign, Pre, and Pro.
- Data Prepocessing **:** During the data preprocessing stage, the raw data is processed and analyzed using techniques such as data reduction, data integration, data conversion, and data cleaning to transform it into a structured format that can be easily interpreted.
- Segmentation : Segmentation refers to the procedure of dividing an image into smaller

portions, which can be represented by a label or mask. Segmentation allows for processing of only the relevant segments, rather than the entire image, which can save time and computational resources.

B. Methodology

Research methodology follows:

1. Microscopic Image Acquisition:

The first section of the dataset contains images of individuals with different subgroups of Acute Lymphoblastic Leukemia (ALL), including pre-ALL, pro-ALL, early-ALL, and Benign-ALL. The second section of the dataset comprises 100 images of patients diagnosed with multiple myeloma.

2. Image augmentation :

To improve the performance of the Acute Lymphoblastic Leukemia (ALL) detection system, the input images are subjected to several processing steps. The images are first rotated, and edges are extracted to enhance their quality. Additionally, the dataset is augmented using various image manipulation techniques to generate new training and evaluation sets. This approach ensures that the model can generalize well to different sizes, poses, and lighting conditions. By creating additional data from existing data, the system can leverage a larger quantity of images and improve its accuracy.

3. Image Feature Extraction

During feature extraction, pre-processed images with various abnormalities are analyzed to extract multiple features from them. In Deep Learning, feature selection is crucial as it heavily impacts the model's performance. A model with too many features can suffer from over-fitting and reduced accuracy due to redundant and noisy data.

Some benefits of feature extraction include reducing over-fitting by eliminating noise-based predictions, shortening the training period by using fewer training samples, and increasing accuracy by eliminating false data and anomalies.

- Geometrical Features : Geometrical features include area, radius, perimeter, symmetry, border, concavity, compactness, solidity, eccentricity, elongation, and form factor.
- Texture Features : Texture features can include homogeneity, energy, correlation, entropy, contrast, and angular second momentum.
- Color Features : To extract color features, the RGB color spaces are converted into HSV color spaces, and then the mean color values are computed.
- Statistical Features : Statistical features encompass various attributes of an image such as the gradient matrix for RGB or HSV colour space, along with the mean value, variance, skewness, and kurtosis of the image matrix's histograms.

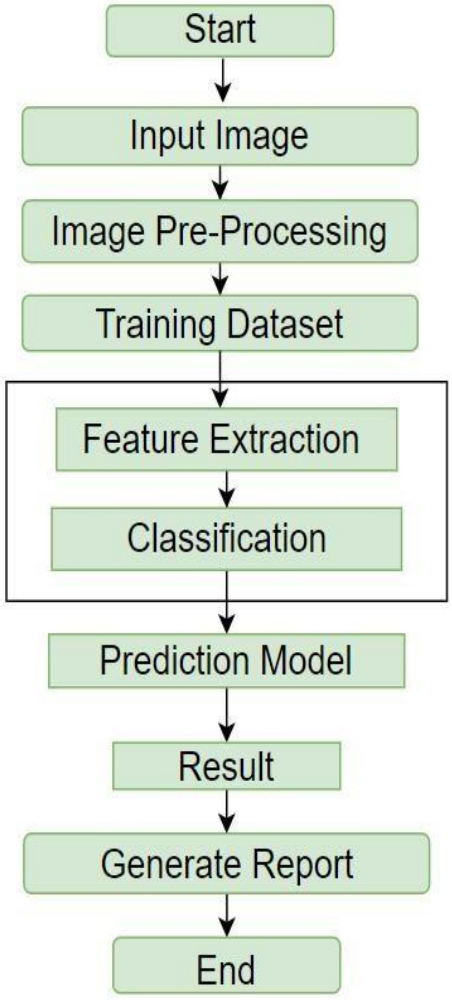


Fig3. flowchart of the proposed system

4. Image Classification:

The next step is classification, where the extracted features are analyzed by the classifier to determine if a cell is normal or cancerous. This is accomplished by comparing geometric, statistical, texture, and size ratio features from segmented regions with standard

features. The analysis results are then used to identify the subtypes and types of acute leukemia.

Neural networks are utilized to automatically detect cancer in blood samples due to their capability as a reliable classifier for various practical applications. One of the key factors for constructing an accurate CNN-based model is the training and validation processes. During training, the CNN model is trained using the training feature set, while the accuracy of the trained model is evaluated using the testing feature set with a feed-forward backpropagation network. The dataset for the training and validation processes is split into two sections. During the training phase, the connection weights are continuously adjusted until the specified iteration number or the desired error rate is reached. The use of neural networks enables automatic detection of cancer.

C. Proposed Convolution neural Network And Architecture

The study proposes an improved CNN model for accurately classifying different types of ALL blood cancer. The architecture of the CNN used in the study is shown in Figure 5. Convolutional neural networks (CNNs) are widely used for image analysis and are known to be the backbone of image classification systems. They are capable of performing image classification quickly and accurately with minimal pre-processing. The CNN model comprises of several hidden layers, including input, output, and fully connected layers.

The model proposed in this study takes an image as input and outputs the type of ALL blood cancer. The CNN model is optimized for two-dimensional pattern recognition and includes three types of layers: convolutional layers, pooling layers, and fully connected layers.

1) Convolutional Layer

The convolutional neural network (CNN) used in the study consists of six levels of convolution layers. The first two convolution layers apply 16 filters of size 3x3 to the input image, while the second layer applies 32 filter combinations of the same size. The final two convolution layers apply 64 3x3 filters to the input image. The ReLU activation function is used in the nonlinear transformation sublayer.

In the convolution layer, neurons serve as feature extraction units. The input image is convolved with a k by k matrix, also known as a filter, to create an activation map. The "stride" refers to the predetermined distance at which the filter moves along the image. The result of convolutioning an input image of size a b with a kernel of size k, padding p, and stride s is of size ((a - k + 2p) /s + 1) x ((b - k + 2p) /s + 1).

2) MaxPooling Layer

CNNs typically consist of several layers, with each layer performing specific functions. The first six layers in this study were convolution layers, where the input image was convolved with a set of filters. The first two convolution layers used 16 3x3 filters, while the third and fourth layers used 32 3x3 filters, and the final two layers applied 64 3x3 filters. Nonlinear transformations were applied using the ReLU activation function.

Another important layer in CNNs is the pooling layer, which downsampled the output of the convolutional layers to reduce the number of parameters and minimize overfitting. In this study, the max-pooling function was used, where the highest value within each non-overlapping region of the image was retained. The pooling layer used a kernel of size k and stride of size s, resulting in an output size of (a/k) x (b/k).

Fig4: Maxpool Function

After the pooling layer, the eighth layer was a flattening layer that converted the multidimensional array into a one-dimensional array of 4800 values. The ninth layer was a fully connected ANN layer that used the ReLU activation function to map the 4800 input values to 64 output values. To prevent overfitting, a dropout layer was added as the eleventh layer, where 50% of the input values were randomly set to zero.

3) Fully Connected Layer

The eleventh and last layer of the model is a fully connected artificial neural network (ANN) that uses the sigmoid activation function to map 64 input values to 2 class labels. This layer is also known as a Multi-Layer Perceptron and is added after the convolutional layers. A dense layer is added to the end of the convolutional layers,

which is then followed by the fully connected layer. These layers operate as regular neural networks and classify images based on the features that were extracted from the convolutions. The error is calculated at this stage and then backpropagated to improve the accuracy of the model.

The proposed model includes five fully connected layers, with all four preceding the output layer utilizing the softmax activation function. The sigmoid activation function, which is used in the output layer, produces a probability value between 0 and 1 for each classification label that the model is attempting to predict. The sigmoid function is defined in Equation (1), where the incoming vector is denoted by z.

$$Sig(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{e^z+1} \quad (1)$$

Here, the incoming vector is z.

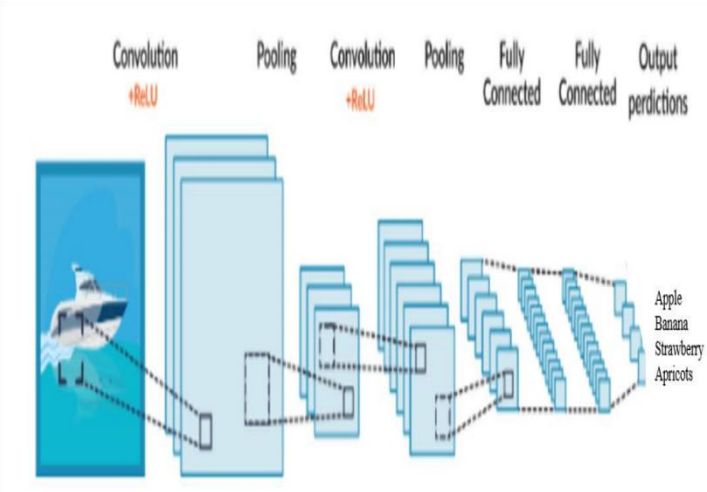


Fig5: Proposed CNN Model Architecture.

IV.RESULTS AND ANALYSIS

The authors of the study utilized TensorFlow, an open-source framework, to develop their classification model. Over 25 iterations, the model was trained on a dataset of 2,356 images to perform binary classification of different types of cancer. The Adam Optimizer was used in each iteration to minimize the loss function and improve the model's accuracy. After training, the model was able to predict the type of cancer present in new images. Training was performed using a K80 GPU. In the following section, the authors report their results and compare the performance of their model with other state-of-the-art deep learning and machine learning models.

a. *DEEP LEARNING APPROACH*

The weights of our network are updated using equation (2) and trained using Adam Optimizer with a learning rate of 0.01. We use the sigmoid cross-entropy loss function, which is optimised by the Adam algorithm. We have depicted the reduction in loss as the number of rounds increases in Figure 5. Additionally, Figure 7 illustrates the training loss, testing loss, and training accuracy, while Figure 8 demonstrates the testing accuracy after 25 rounds. We have provided the confusion matrix for the binary classification of ALL blood cancers using the CNN model in training loss, testing loss, and training accuracy, while Figure 8 demonstrates the testing accuracy after 25 rounds. We have provided the confusion matrix for the binary classification of ALL blood cancers using the CNN model in Figure 6. Our CNN model achieved a precision of 97.25% and specificity of 95.19%.

$$w(n) = w(n-1) - \alpha * m(t) / \left(\sqrt{v(t)} + \epsilon\right) \quad (2)$$

where w is the weight matrix, is the learning rate, and m(t) and v(t) are the bias-corrected estimators for the first and second moments, respectively.

b. *ANALYSIS USING MACHINE LEARNING*

In the following subsection, we conducted a comparative study to demonstrate the effectiveness of deep learning as the volume of data increases. Our analysis is based on various picture classification algorithms provided by machine learning. We extracted two features from the images, namely the histogram and the Discrete Fourier Transform (DFT), and used them to train all machine learning classifiers.

The SVM algorithm, a supervised learning method, was used to construct the classification model with the "RBF" kernel. Naive Bayes, a probabilistic classifier, utilized the Gaussian version to differentiate between all types of blood cancer. The decision tree classifier model estimated the value of a variable using the input sequence of the feature vector. Finally, we utilized the random forest ensemble

learning algorithm to output the mean prediction of each individual decision tree and provide a decisive difference.

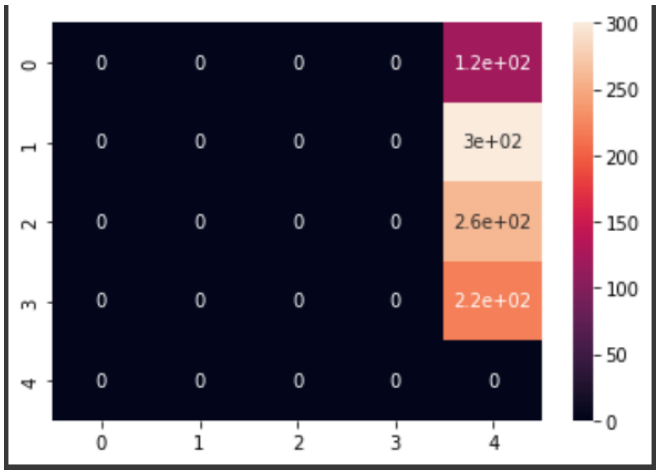


Fig6: Confusion Matrix

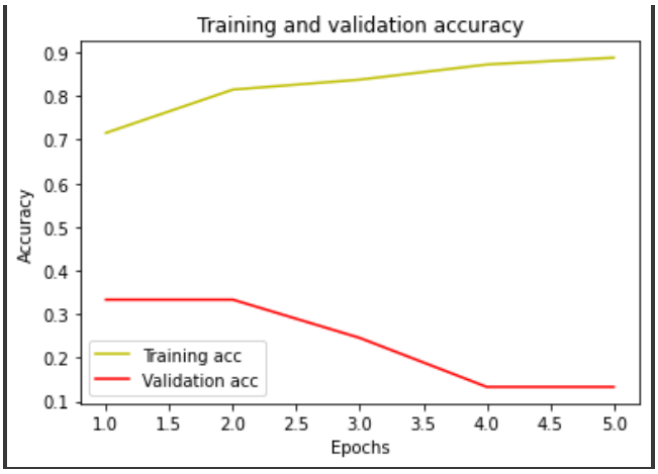


Fig7: Training and validation accuracy

Fig8: Training and validation loss

c. ANALYSIS USING TRANSFER LEARNING

Our proposed model's superior performance over transfer learning models such as VGG-16 has been demonstrated to reduce computation costs for similar problems. The VGG-16 convolutional neural network model has three layers stacked on top of each other, and it uses softmax for classification with two fully connected layers containing 4096 nodes each. We compared various metrics, including accuracy, precision, recall, specificity, and F1 score, using the following equations:

$$AC = (TP + TN)/(TP + TN + FP + FN) \quad (3)$$

$$P = TP/(TP + FP) \quad (4)$$

$$R = TP/(TP + FN) \quad (5)$$

$$S = TN/(TN + FP) \quad (6)$$

$$F = (2 * P * R)/(P + R) \quad (7)$$

TP refers to the number of samples correctly classified as positive, while TN is the number of samples correctly classified as negative. FP represents the number of samples that were wrongly classified as positive, while FN represents the number of samples that the model incorrectly predicted as negative.

V. CONCLUSION

The proposed method involves analyzing variations in texture, geometry, colors, and statistical inputs to extract features in microscopic images. This system is designed to be fully automatic, reliable, accurate, and efficient, with faster data processing, fewer errors, and lower costs. Our team has developed a web-based program that focuses on detecting Acute Lymphocytic Leukemia, displaying necessary precautions, and generating a comprehensive report on the disease detection. Initially, we trained and tested our model using microscopic images of blood and

evaluated its accuracy against three previously trained deep learning models. The system can identify multiple types of blood cancers.

REFERENCE

1. C. C. Hortinela, J. R. Balbin, J. C. Fausto, P. Daniel C. Divina and J. P. T. Felice "Identification of Abnormal Red Blood Cells and Diagnosing SpecificTypes of Anemia Using Image Processing and Support Vector Machine" IEEE 11th International Conference,2019
2. R. P and S. D. P "Detection of Blood Cancer- Leukemia using K-means Algorithm"5th International Conference, 2021
3. A. Das Mou and P. Kumar Saha "A Comprehensive Study of Machine Learning algorithms for Predicting Leukemia Based on Biomedical Data" 2nd International Conference,2019
4. J. A. Akrimi, A. Suliman, L. E. George and A. R. Ahmad "Classification red blood cells using support vector machine" 6th International Conference ,2017
5. A. Ratley, J. Minj and P. Patre "Leukemia Disease Detection and Classification Using Machine Learning Approaches: A Review" First International Conference ,2020
6. R. Sigit, M. M. Bachtiar and M. I. Fikri "Identification Of Leukemia Diseases Based On Microscopic Human Blood Cells Using Image Processing" International Conference ,2018
7. S. Mandal, V. Daivajna and R. V."Machine Learning based System for Automatic Detection of Leukemia Cancer Cell" IEEE 16th India Council International Conference ,2019
8. M. Akter Hossain, M. Islam Sabik, I. Muntasir, A. K. M. Muzahidul Islam, S. Islam and A. Ahmed"Leukemia Detection Mechanism through Microscopic Image and ML Techniques" IEEE REGION 10 CONFERENCE ,2020
9. R. Agrawal, S. Satapathy, G. Bagla and K. Rajakumar, "Detection of White Blood Cell Cancer using Image Processing, "International Conference ,2019
10. C. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearsonand E. Ifeachor"Identification of Optimum Panel ofBlood-based Biomarkers for Alzheimer's Disease Diagnosis Using Machine Learning" 40th Annual International Conference of the IEEE ,2018
11. T. Xia, Y. Q. Fu, N. Jin, P. Chazot, P. Angelov and R. Jiang, "AI-enabled Microscopic Blood Analysis for Microfluidic COVID-19 Hematology," 5th International Conference,2020
12. .C. B. Alkan and Z. Isik,"Characterization of Cancer Types by Applying Machine Learning Methods on Blood RNA-Sequencing Data,"3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT,2019
13. R. Mosayebi, A. Ahmadzadeh, W. Wicke, V. Jamali, R. Schober and M. Nasiri-Kenari,"Early Cancer Detection in Blood Vessels Using Mobile Nanosensors"IEEE Transactions,April 2019
14. M. Sajjad "Leukocytes Classification and Segmentation in Microscopic Blood Smear: A Resource-Aware Healthcare Service in Smart Cities,"IEEE Access, 2017
15. O. Ceylan, G. K. Mishra, M. Yazici, A. Qureshi, J. H. Niazi and Y. Gurbuz"A Hand-Held Point-of-Care Biosensor Device for Detection of Multiple Cancer and Cardiac Disease Biomarkers Using Interdigitated Capacitive Arrays," IEEE Transactions,Dec. 2018,