International Journal of Advanced Research in Engineering Technology & Science

ISSN: 2349-2819

Email: editor@ijarets.org Volume-8, Issue-12 December- 2021 www.ijarets.org

# A STUDY ON CLASSIFICATION OF LEUKEMIA DISEASE IN PERIPHERAL BLOOD CELL IMAGES USING CONVOLUTIONAL NEURAL NETWORK

**Prayaga Siddappa**

Assistant Professor,

Department of Computer Science

Government College(A), Kalaburagi

## ABSTRACT

Classification is extensively used in clinical snap shots to categorize patients and non-sufferers. However, conventional type requires a complex manner, consisting of a few rigid steps consisting of preprocessing, segmentation, feature extraction, detection, and class. In this paper, we advocate a novel convolutional neural network (CNN), called Leukemia Net, to specially classify two one of a kind styles of leukemia, consisting of acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), and non-cancerous sufferers. To amplify the restricted dataset, a PCA coloration augmentation method is applied before photos are input into the LeukemiaNet. This augmentation approach enhances the accuracy of our proposed CNN architecture from ninety six.Nine% to ninety seven.2% for distinguishing ALL, AML, and ordinary mobile photos.

***Keywords-*** *Leukemia, classification*

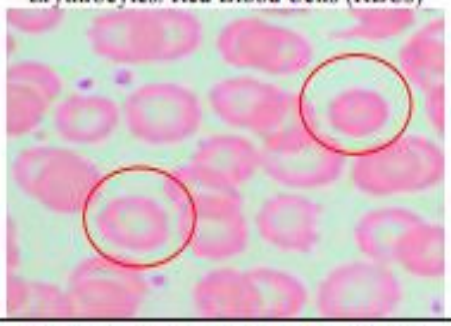
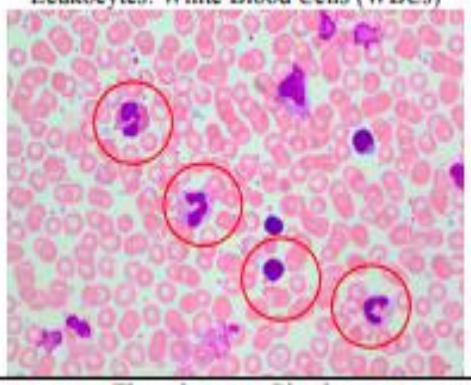
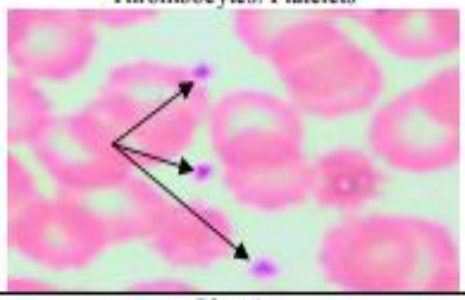
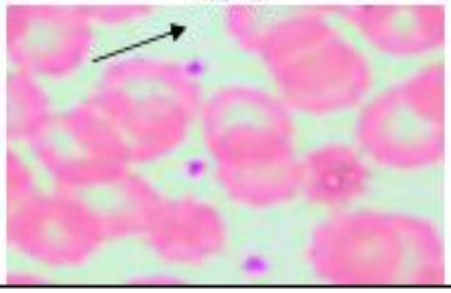
## INTRODUCTION

Leukemia is a group of heterogeneous blood-related cancers, differing in its aetiology, pathogenesis, prognosis and response to remedy (Bain, 2017). Leukemia is taken into consideration as a serious difficulty in contemporary society, because it influences both kids and adults or even from time to time babies under the age of 365 days. In youngsters, leukemia is taken into consideration as the most not unusual type of most cancers, whilst, in adults, the World Health Organization file indicates that leukemia is one of the top 15 most commonplace varieties of most cancers (Kampen, 2016). To higher recognize leukemia, the subsequent sections are dedicated to the discussion of the blood cells lineage, styles of leukemia, diagnostic strategies presently in use, treatments options in addition to prognostic elements.

### Blood and its Components

Blood is a purple colored, life-maintaining fluid which circulates through the coronary heart and blood vessels (Veins and Arteries). Blood is vital for existence. Blood flows all through the human frame carrying oxygen and vitamins to the tissues and promises leftover products of metabolism to the lungs, liver and kidneys, where they are then eliminated from the body (Bain, 2018). Blood accommodates of 4 essential factors particularly plasma, pink blood cells (RBC), white blood cells (WBC) and platelets (Starr et al., 2017; Ciesla, 2017). Table 1 demonstrates the 4 principal components of blood.

**Table 1: The Four Major Components of Blood**

| Blood Element | Description |
|---|---|
| Erythrocytes: Red Blood Cells (RBCs)  | RBCs are responsible for carrying oxygen from lungs to the body tissues and organs and bringing back carbon dioxide to the lung (Paul, 2006) |
| Leukocytes: White Blood Cells (WBCs)  | WBCs are part of the immune system where they defend the body against both infections and foreign bodies. (Brooks, 2008) |
| Thrombocytes: Platelets  | Platelets are responsible for aiding in the blood clotting and subsequent wound healing, which occur at a site of injury (Manfred et al., 1999) |
| Plasma  | Blood plasma carries many important substances such as nutrients, waste, gases, and antibodies. (Aehlert & Vroman, 2011) |

All blood cells originate from the BM, developing from the hematopoietic stem cells (lymphoid and myeloid) (Ciesla, 2016). Figure 2. Shows the maturation path of various blood cells originating from the haematopoietic stem cells such as the lymphoid and myeloid stem cells.

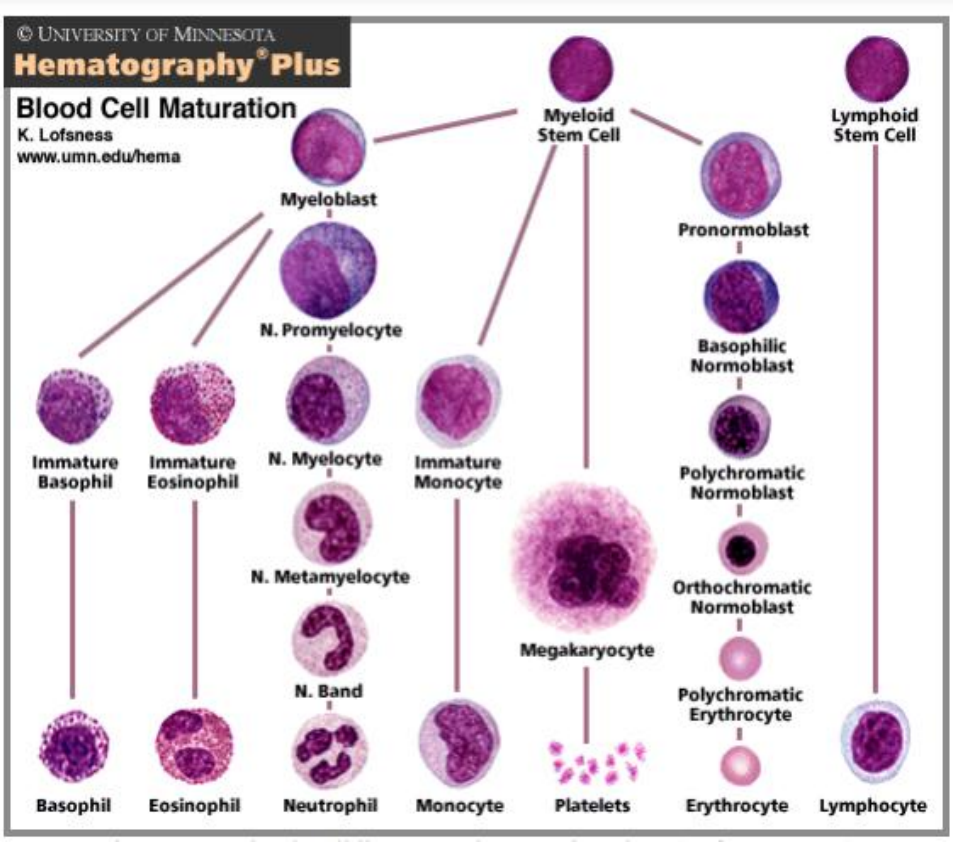



Figure 2: Blood Cell lineage and maturation chart (Lofsness. 2008)

**White Blood Cells (Leukocytes)**

Normally, WBCs are large in length than RBCs and platelets (Zamani & Safabakhsh, 2006). However, WBCs are the least numerous issue of blood cells where each micro liter of blood includes about 5000-ten thousand WBCs (Esteridge et al., 2019) instead of a hundred and fifty,000 platelets in the same volume. WBCs are a component of the immune machine and offer the first best defense against both infections and foreign bodies. (Brooks, 2018).

The human blood accommodates of five styles of WBCs particularly basophil, eosinophil, neutrophil, monocyte, and lymphocytes. In wholesome human blood, each kind of WBC has a particular percentage of WBCs as follows: neutrophils 50- 70%, eosinophils 1-four%, basophils 0-1%, monocytes 2-8%, lymphocytes 20.40%. Calculating the proportion of different kind of WBC is referred to as differential blood be counted (Rl Bijlani & Manjunatha, M. 2010, GK & Pravati, 2006).Types of Leukemia

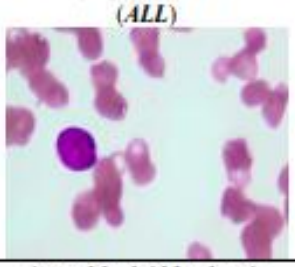
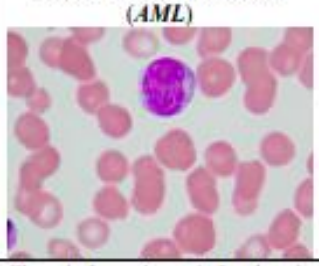
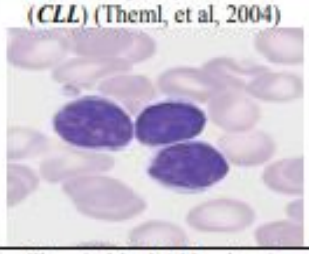
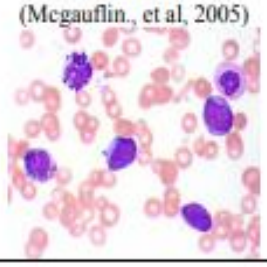
Figure 2 suggests the maturation process of the numerous blood cells and levels of maturation before becoming adult cells and are launched into the bloodstream. In the case of leukemia, an interruption in the WBC maturation method happens where the immature cells (blasts) stay immature, can not perform normal feature, multiply continuously and ultimately invade the BM, replacing all the everyday cells.

Leukemia is usually divided into two types referred to as acute leukemia or continual leukemia depending on how quick the immature cells (blasts) proliferate (Bain, 2018). Leukemia can be

similarly grouped based at the kind of cell that predominates in the PB and the BM described in keeping with cell lineage as either myeloid or lymphoid (Please Refer to Figure 2) (Ciesla, 2007).

Hence, there are 4 kinds of leukemia, particularly, acute lymphoblastic leukemia (A1L), acute myeloid leukemia (AML), persistent lymphocytic leukemia (CLL) and continual myeloid leukemia (CML). (Edward, 2017; Bain, 2013; Bain, 2016; Norman, 2019). Table 1 reveals the four kinds of leukemia.

**Table 1 : The Four Main Types of Leukemia (Hoffbrand et al., 2019)**

| Progression | Stem Cell | Type | Description |
|---|---|---|---|
| Acute | Lymphoid | Acute Lymphocytic Leukemia (*ALL*)  | The most common type of leukemia in young children. This disease also affects adults, especially over the age of 65. |
| | Myeloid | Acute Myeloid Leukemia (*AML*)  | It develops in both adults and children. |
| Chronic | Lymphoid | Chronic Lymphocytic Leukemia (*CLL*) (Theml, et al, 2004)  | It occurs mainly in adults and almost never seen in children |
| | Myeloid | Chronic Myeloid Leukemia (*CML*) (Mcgauflin., et al, 2005)  | Most commonly it affects adults over the age of 55. It sometimes occurs in younger adults, but it is rare in children. |

## OBJECTIVE OF THE STUDY

To have a look at to look at Types of Leukemia

2. To study Blood and its Components

## RESEARCH METHODOLOGY

A sufficient dataset of leukemia photos is a mission for our proposed method due to the privacy of such pix. Therefore, we perform two kinds of information augmentation to educate accurate and robust CNN architecture. The first facts augmentation is traditional techniques, in which to be had samples are transformed into new samples the usage of a few conventional affine transformation methods which includes shear photograph, rotation, mirrored image, translation, histogram equalization and so forth. The second one is PCA coloration augmentation, as referred to in [10]

### Conventional image augmentation methods

In this paper, to boom the variety of pix within the original dataset, we increase the dataset by way of applying following transformation strategies.

### Conversion of RGB pix to grayscale images

All RGB photos inside the dataset are converted to grayscale intensity photos via discarding the saturation and hue in the unique photos at the same time as the luminance remains.

Blurring A Gaussian filter out is implemented not best within the original RGB pictures but additionally in grayscale pics with a scalar cost for sigma generated randomly.

### Histogram equalization

Histogram equalization is used to enhance the contrast of a grayscale photograph. The picture after being implemented histogram equalization has the higher comparison.Reflection

The image is symmetrically taken alongside the x-axis and the y-axis. Fig. Four (f) depicts a mirrored image photo of the photograph in Fig. 4 (a) along the x-axis

Rotation The image is rotated counterclockwise or clockwise some attitude that's generated randomly among. The photo rotates clockwise if the angle has a poor cost. The circled picture makes the output photograph big sufficient to incorporate the whole rotated photograph. Therefore, we should resize the output image to the equal size as the authentic image.

### Translation

We behavior horizontal and vertical translation for snap shots with the displacement fee randomly selected with a uniform distribution of among 25 and 50.

### Shearing photographs

A simple vertical and horizontal shear transformation is implemented to pics with the price of distortion is chosen randomly between [0.3, 1], as illustrated in Fig. 4 (h) and (i).

### PCA shade augmentation

This technique was first proposed by using Alex Krizhevsky to boom the dimensions of the dataset.The fundamental concept of this technique is altering the intensities of RGB channels in an

picture. PCA reduces the number of dimensions inside the facts by means of locating styles in the records. There are 6 steps for appearing PCA coloration augmentation which can be summarized as underneath**:**Each pixel of an image is composed of three vectors of red (R), green (G) and blue (B).

PCA is then computed on these statistics points. Each channel of an image is converted to a vector

$$x=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

1.Compute the means for 3 dimensions of every photo mi with i : 1,2, 3

$$m_i = \frac{1}{n}\sum_{k=1}^{n} x_k \quad (1)$$

$n$: *size of image*

$m_i$: *mean of each channel of an image*

2. Compute the covariance matrix. The covariance matrix has a size of 3×3 because the data is 3 dimensional.

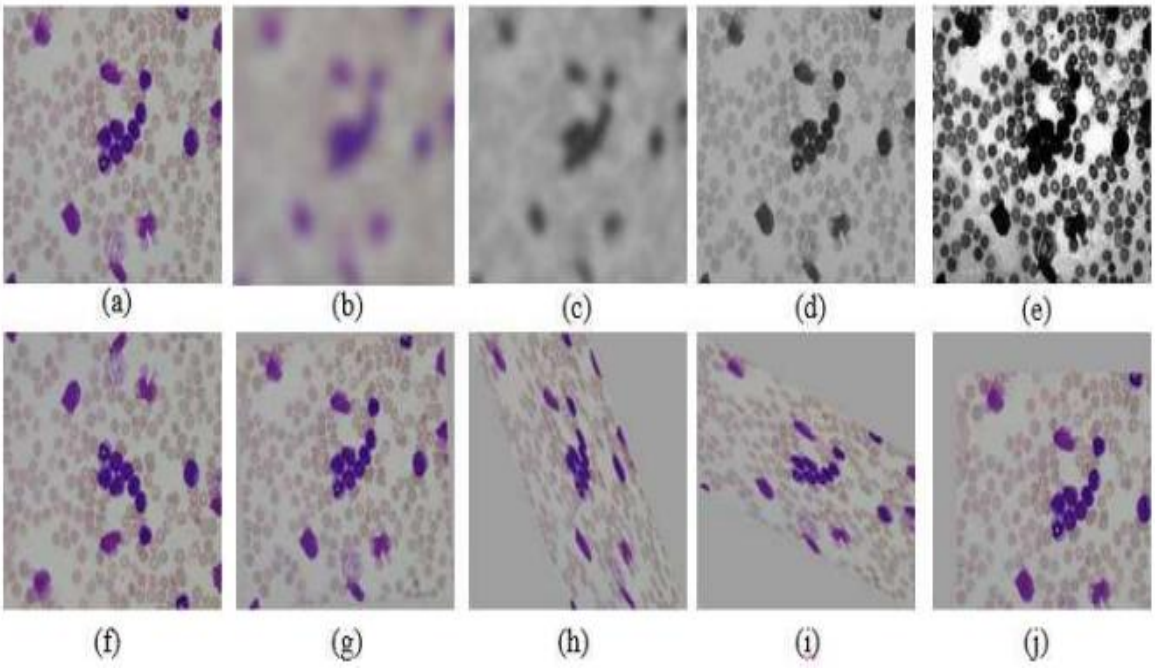
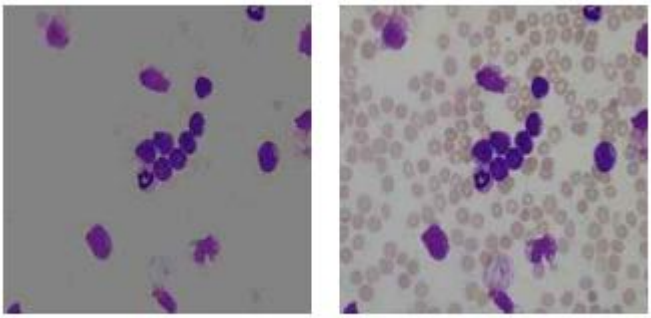



Fig. 4. Conventional facts augmentation, (a) An authentic ALL picture, (b) A blurring ALL picture, (c) A blurring grayscale image, (d) A grayscale image, (e) Histogram equalization, (f) Reflection image

thru x-axis, (g) Rotated photo, (h) Shearing photograph alongside x-axis, (i) A shearing photo alongside y-axis, and (j) Image translation.



(a) (b)

Fig. Five. PCA shade augmentation, (a) The authentic ALL photo, and (b) ALL photograph after being carried out PCA color augmentation.

$$S = \sum_{k=1}^{n}(x_k - m)(x_k - m)^T \quad (2)$$

1.Compute the eigenvector (e 1 , e 2 , e three) and corresponding eigenvalues (λ1 , λ2 , λ3)

2.Sort the eigenvectors from the highest to lowest and discard the lower eigenvalues. Choose okay eigenvectors with the largest eigenvalues from a dxk dimensional matrix W. In W, every column represents an eigenvector.

3.Take the transpose of the vector and multiply it to the transposed original dataset, after that, take the transpose of the result. We want to add multiples of located principal additives with magnitudes proportional to the corresponding eigenvalues instances a random variable drawn from a Gaussian distribution with suggest zero and fashionable deviation of zero.

## EXPERIMENTAL RESULT AND DISCUSSION

Our experiment become carried out on Matlab with NVIDIA GeForce GTX 1080 Ti. We carried out two experiments on special datasets to evaluate the robustness of augmentation techniques and the effectiveness of our proposed CNN structure. All of the experiments were run for 30 epochs on the gaining knowledge of charge of 0.001 with the enter size photograph of [228x227x3].

### Dataset

In this paper, the dataset includes 49 ALL snap shots and 59 everyday mobile pix from the ALLIDB1 database, amongst which 33 AML snap shots are accrued from the Internet. Because of insufficient information, the CNN structure can not research many wealthy capabilities. To enhance the accuracy of our proposed CNN architecture, we amplify the original dataset via applying some statistics augmentation techniques inclusive of blurring using Gaussian clear out, histogram equalization, converting photographs to grayscale picture, shearing photograph, photograph reflection, photograph translation, etc. The benefit of the use of statistics diversification demonstrates that our CNN architecture can handle a range of pix accumulated from specific studies organizations with special pleasant (noisy pictures, tilted photos, etc.).

**Metrics to evaluate the performance of LeukemiaNet**

In this paper, specific overall performance metrics are applied to assess the preferred outcomes of LeukemiaNet architecture.

**A. Confusion matrix**

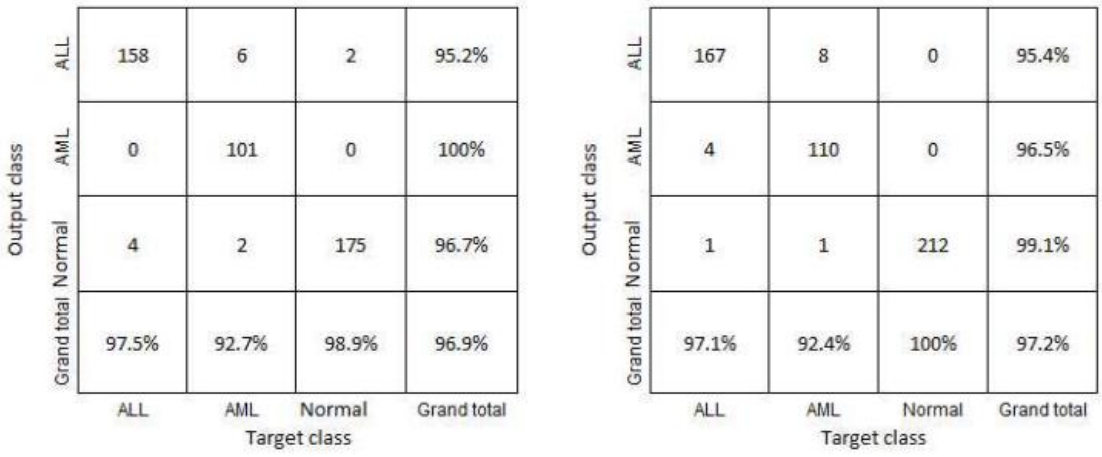
A confusion matrix figures out an in depth itemization of correct and inaccurate classifications for every magnificence, as illustrated in Fig. Three.

True positives (TP) is the case while the output class of an image and the goal magnificence are real. For instance, that takes place whilst someone is recognized with ALL and the LeukemiaNet classifies his case as ALL. In Fig. 6 (a), TP of ALL, AML, and the ordinary magnificence correspond with 158 photographs, a hundred and one snap shots, and 175 images.

On the opposite, authentic negatives (TN) is a case while the goal class of the image is fake and the output elegance is also false. The overall wide variety of TN for the ALL elegance, the AML elegance, and the regular magnificence in Fig. 6 (a) is 278 pix, 339 snap shots, and 265 pix, respectively.

False positives (FP), for example, is defined as the case while LeukemiaNet version predicts a noncancerous man or woman as a affected person who has ALL. In Fig. Five (a), the total range of FP for the ALL class is 8, for the AML elegance is 0, and for the ordinary magnificence is 6.

False negatives (FN) is a case whilst the target class of the photograph is real and the output elegance is fake. For example, a person having ALL and the LeukemiaNet classifies this case as a non-cancerous man or woman or AML patient. The general range of FN for the ALL class, the AML elegance, and the everyday magnificence are four pictures, 8 pictures, and 2 photographs,respectively.



**B. Accuracy**

Accuracy is a superb metric to assess Leukemia Net because the number of photographs in lessons is approximately balanced. Accuracy is calculated by using dividing the sum of correct classifications

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

via the total wide variety of pictures.

**C. Precision**

Precision is a dimension metric that figures out what percentage of pix that LeukemiaNet anticipated belong to a category, virtually belong to this magnificence.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Using Equation (4), the precision of 3 instructions in Fig. Five (a) is 95.2%, 100%, and ninety six.7%, similar to ALL, AML, and ordinary class. Similarly, the precision of ALL magnificence, AML elegance, and the regular magnificence in Fig. 6 (b) is 95.Four%, ninety six.Five%, and ninety nine.1%, respectively.

**D. Recall (sensitivity)**

Recall corresponds to the real-wonderful charge of the fascinated elegance**.**

$$Recall = Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

The sensitivity of 3 instructions in Fig. 3 (a) and (b) is proven in Table 1.

**E. Specificity**

Specificity corresponds to the genuine-bad charge of the considered elegance.

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

**Experiment consequences**

Two experiments have been performed, considered one of which turned into based totally on the dataset extended the usage of conventional information augmentation methods, at the same time as the

opposite applied the dataset extended the use of traditional facts augmentation techniques mixed with the PCA shade augmentation method to check the effectiveness of numerous augmentation methods

## .CONCLUSION

In this paintings, facts augmentation techniques to extend dataset to prevent overfitting and memorize the precise details of the education snap shots were implemented. While conventional augmentation methods are clean to use and not devour time, the opposite strategies to extend dataset show a promising solution to help to reduce the error charge of CNN structure. Besides, we proposed LeukemiaNet convolutional neural network structure to discover efficiently ALL, AML photographs from normal mobile pictures with high accuracy. The proposed CNN architecture changed into experimented on two separate datasets, one with data prolonged the usage of traditional methods simplest and the other with prolonged statistics the use of a mixture of conventional and PCA shade augmentation strategies, to affirm that our network can deal with the version of snap shots. The accuracy and performance of LeukemiaNet structure turned into additionally validated, in want reliable dataset of leukemia picture. In future work, we would like to apply this proposed technique in a massive dataset which include 4 forms of Leukemia (AML, ALL, Chronic lymphocytic leukemia - CLL and Chronic myelogenous leukemia – CML).

## REFERENCE

[1] Leukemia: What you need to know.

https:// www.medicalnewstoday.com/articles/142595. php (accessed Dec., 25, 2017).

[2] Leukemia stages. https://www.cancercenter. com/leukemia/stages/ (accessed Apr., 30, 2018).

[3] C. Raje and J. Rangole, "Detection of Leukemia in Microscopic Images using Image Processing," P roceedings of 2014 International Conference on Communication and Signal P rocessing, Melmaruvathur, pp. 255-259, 2014.

[4] I. Vincent, K.R. Kwon, S.H. Lee, and K.S. Moon, "Acute Lymphoid Leukemia Classification using Two-Step Neural Network Classifier," P roceedings of 2015 21st Korea-J apan J oint Workshop on Frontiers of Computer Vision (FCV), pp. 1-4, 2015.

[5] D. Goutam and S. Sailaja, "Classification of Acute Myelogenous Leukemia in Blood Microscopic Images using Supervised Classifier," P roceedings of 2015 IEEE International Conference on Engineering and Technology (ICETECH ), Coimbatore, pp. 1-5, 2015.

[6] M.C. Su, C.Y. Cheng, and P.C. Wang, "A Neural-Network-Based Approach to White Blood Cell Classification," The Scientific World J ournal, Vol. 2014, Article ID 796371, 9 pages, 2014.

[7] S.C. Neoh, W. Srisukkham, L. Zhang, S. Todryk, B. Greystoke, C.P. Lim, et al., "An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images," Scientific Reports, Vol. 5, pp. 1-14, 2015.

[8] J. Song, H.I. Kim, and Y.M. Ro, "Fast and Robust Face Detection based on CNN in Wild Environment," J ournal of Korea Multimedia Society, Vol. 19, No. 8, pp. 1310-1319, 2016.

[9] S.W. Park and D.Y. Kim, "Performance Comparison of Convolution Neural Network by Weight Initialization and Parameter Update Method1," J ournal of Korea Multimedia Society, Vol. 21, No. 4, pp. 441-449, 2018.

[10] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," N IPS' 12 P roceedings of the 25th International Conference on Neural Information P rocessing Systems, Vol. 1, pp. 1097-1105, 2012.

[11] S. Kansal, S. Purwar, and R.K. Tripathi, "Trade-off between Mean Brightness and Contrast in Histogram Equalization Technique for Image Enhancement," 2017 IEEE International Conference on Signal and Image P rocessing Applications (ICSIP A), pp. 195- 198, 2017.

**[12]** Padding and Shearing an Image Simultaneousl.

https://www.mathworks.com/help/images/padding-and-shearing-an-image-simultaneously.html (accessed Dec., 24, 2017).