







# Computer-aided deep learning model for identification of lymphoblast cell using microscopic leukocyte images

Abhishek Kumar<sup>1</sup>  | Jyoti Rawat<sup>2</sup>  | Indrajeet Kumar<sup>3</sup>  | Mamoon Rashid<sup>4</sup>  |  
Kamred Uddham Singh<sup>5</sup> | Yasser D. Al-Otaibi<sup>6</sup>  | Usman Tariq<sup>7</sup> 

<sup>1</sup>School of Computer Science and IT, JAIN (Deemed to be University), Bangalore, India

<sup>2</sup>School of Computing, DIT University, Dehradun, India

<sup>3</sup>Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, India

<sup>4</sup>Department of Computer Engineering, Faculty of Science and Technology, Vishwakarma University, Pune, India

<sup>5</sup>Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan

<sup>6</sup>Department of Information Systems, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia

<sup>7</sup>College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

## Correspondence

Mamoon Rashid, Department of Computer Engineering, Faculty of Science and Technology, Vishwakarma University, Pune, India.  
Email: mamoon873@gmail.com

## Abstract

The conventional technique of leukocyte cell classification involves segmenting the required portion of cells from input image, extracting features of the segmented nuclei, reducing and optimizing these features and then implements the classifier. Thus, designing a good classifier by using such techniques increases the time complexity of the system. In order to resolve such issues, the proposed work implements the deep convolutional neural network (DCNN)-based models for classifying malignant versus normal WBCs. The proposed system is validated on 108 images of ALL-IDB 1. Due to limited number of training samples, data augmentation is used to create a similar type of virtual image. In this work, experimentation is carried out for discrimination between normal and infected WBC using DCNN with four different activation functions. By using this method, a set of 6000 samples are generated and used for proper training of the DL model for all activation functions. The performance of each trained model is evaluated in terms of accuracy, recall, precision and F-measure with the maximum values of 98.1%, 98.3%, 98.3% and 98.3% are achieved, respectively. Finally, it has been concluded that the defined DCNN model and ReLu activation function yield outstanding performance for lymphoblast characterization using microscopic blood images.

## KEYWORDS

acute lymphoblastic leukaemia, convolutional neural network, data augmentation, leukocyte, lymphoblast identification

## 1 | INTRODUCTION

Leukaemia is a fatal ailment associated with the human WBCs. Among all types of most common cancers, leukaemia stands at the tenth position. According to facts and figures provided by American Cancer Society (ACS) in 2019, a total of 176,200 new cases are recorded for myeloma, lymphoma and leukaemia. Among these, leukaemia recorded about 35%, that is, 61,780 cases. Acute lymphoblastic leukaemia (ALL) is a disorder due proliferation of blasts cell, that is, immature blood cell inside the bone marrow, which can include solid or peripheral blood organs. In normal bone marrow, traditionally, 30% or more lymphoblast cells are required for a diagnosis of ALL based on certain morphologic and cytogenetic features present in the white blood cell. Conventional clinical investigative aided systems for the medical images are prone to inconsistency and misinterpretations. Computer vision technologies like artificial intelligence and biomedical image processing work together with full potential to enhance the effectiveness of diagnostic methods by swiftly traversing the medical images and proficiently classifying the disorders. However, there are always challenges involved in implementing the artificial intelligence (AI) in the medical images (Lewis et al., 2019).

Leukaemia is a commonly occurring abnormal tumour with high mortality and frequency, and thus, developing a technique for diagnosing such abnormalities in the lymphoblast cell plays a vital role in medical field (Vogado et al., 2018). Conventional biochemical and immune-

histochemical diagnostic techniques for detecting the lymphoblast leukaemic cells, which typically need fluorescence-labelling, are labour intensive and consume a lot of time. Computer-based diagnosis of leukaemic cell from a given microscopic blood image is important but is equally cumbersome task, especially in case of inconsistency among the training and testing data for example: blood cell images that contain non-negligible noises due to illumination or from staining. In order to deal with such problems, we proposed an automated diagnostic and classification system to evaluate and rank two classes of leukocyte cell images for early detection of acute lymphoblastic leukaemia (ALL) to avoid unnecessary deaths by providing a helping hand to the hematopathologists (Hegde et al., 2019). As per the requirement of pathologists in real-life examination scenario, the proposed system is able to discriminate between the lymphoblast cell and lymphocyte cell using deep convolutional neural networks (DCNN). The outcome of the work verifies the capabilities of the computer-aided system to substitute the conventional diagnostic systems. The proposed system is fully automated, thereby minimizing the chance of incorrect labelling of leukaemia classes due to overlapping of lymphoblast cells. As the proposed approach is automatic, thus, it does not require the implementation of segmentation and feature extraction that helps in fine-tuning the existing models for yielding desired results. Figure 1 shows the organization of the WBC and type of leukaemia disease.

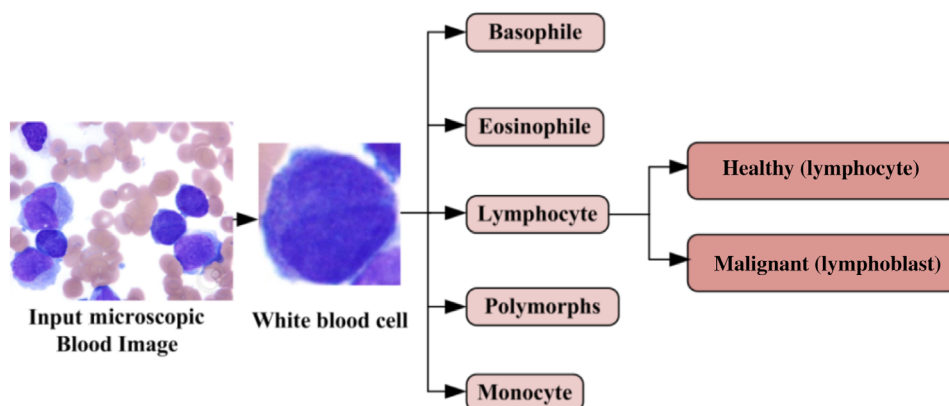
Based on the challenges given in the above paragraphs, we highlight the significant contributions of this paper as follows:

- i. The proposed work implements the deep convolutional neural network (DCNN)-based models for classifying malignant versus normal WBCs.
- ii. This work is validated on 108 images of ALL-IDB 1. Data augmentation is used to create a similar type of virtual image due to limited number of training samples.
- iii. In this work, experimentation is carried out for discrimination between normal and infected WBC using DCNN with four different activation functions.

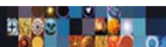
The rest of the paper is ordered as: Related studies section as Section 2, in which state-of-the-art work related to identification and classification of leukocyte with their limitations have been demonstrated. Section 3 provides material and methods where data set description, data augmentation, dataset bifurcation, and proposed model are explained. In Section 4, all the experiments and results of this study are discussed, and finally, the entire work is summarized in Section 5 as a conclusion.

## 2 | LITERATURE REVIEW

There are various studies made in state-of-art classifiers for differentiating among the leukaemic and normal cells by utilizing the microscopic images of leukocyte. The study (Vogado et al., 2018) utilized CNNs to define, select and reduce the features from input image, whereas the SVM is used for classifying the pathological image from given data set. The work proposed in article (Hegde et al., 2019), WBCs are categorized into six types of cells, namely: monocytes, abnormal cells, neutrophils, lymphocytes, eosinophils and basophils. It also utilizes neural networks with various types of hand-crafted features yielding an accuracy of 99.8%. Full training and transfer approaches can be used with CNNs to yield such accuracy. CNNs are also used for distinguishing ALL from normal blood images in study (Safuan et al., 2020). Various pre-trained models based on CNNs, that is, GoogleNet, VGG and AlexNet, are also examined to compare their results yielded in classifying lymphoblast from normal cells images taken from the dataset present in IDB database. Technically, the ratio of images used for training to that of testing is 7:3. From the experiments, the VGG and GoogleNet are found to be better than AlexNet with 100% accuracy with respect to given dataset during training, whereas VGG is reported best testing accuracy if 99.13% (Safuan et al., 2020).



**FIGURE 1** Classification of leukocyte



In the study (Di Ruberto et al., 2020), a method for identification of WBCs followed by its classification to normal or leukaemic cells from microscopic blood images is implemented on datasets available on public domain and yields 94.1% accuracy. The work done in the literature (Acevedo et al., 2019) uses a classifier based on CNNs that can categorize given peripheral blood cells into eight classes. The performance of conventional methods of identifying ALL entirely depends upon the expertise and skills of pathologists who manually analyse the blood images (Arslan et al., 2014; Duan & Yu, 2011). Manual analysis can lead to incorrect diagnosis and can, in fact, make it inefficient and harder with increase in the size of dataset as it is difficult to analyse each image manually. Moreover, due to non-uniform illumination, other types of noises manual analysis become unreliable (Nemane et al., 2013). Thus, the automated computer system is the key in resolving the above mentioned problems associated with blood diagnosis (Nemane et al., 2013), and these systems can be developed through implementing machine learning (ML) to diagnose the lymphoblast and classify it as ALL by using microscopic blood images. However, ML requires some other techniques like segmentation, feature extraction, reduction and classification for its proper functioning (Jiang et al., 2018; Shah et al., 2020), which can affect the performance of the system, because the results yielded by the system depend upon the performance of these techniques; that is, if an efficient segmentation method is not utilized, then classification would not be able to produce desirable results. In this work, a deep convolutional neural network (DCNN) is proposed to resolve such issues.

Deep learning-based techniques like DCNN are proficient in evaluating the features of an object without being dependent upon the yields of other techniques like segmentation and feature extraction and, thus, are used in every field of computer-based medical application to perform remarkably over other techniques (Levy & Jain, 2016). Deep learning is basically a type of machine learning, which combines the use of multiple layers depending upon the need and approach of the developers and their applications. It could be implemented to work in non-supervised and supervised mode for extracting, evaluating and transforming the features and then utilize them to classify the required input data. The multilayer architecture of CNN comprises of four main features like shared weights, multiple layers, local connections and pooling (LeCun et al., 2015).

### 3 | MATERIALS AND METHODS

The proposed work focuses on to develop a computer-aided system for leukaemia identification and classification. The system will take images of peripheral leukocyte cells as input and will give a label to the image representing the class to which the cell belongs to. The proposed system is fully automated, thereby minimizing the chances of incorrect labelling of leukaemia classes due to overlapping of lymphoblast cells. The system takes a less time to diagnose and classify with respect to the currently available system, so that adequate scheduling of treatment can be performed. Ground truth labelling is performed by more than two haematology experts having more than 15 years of experience.

#### 3.1 | Dataset description and preparation

The dataset used in the present work is taken from online available benchmark dataset repository ALL-IDB 1 version 1.0 that contains the images, which are in 24 bit colour depth JPG format with  $2592 \times 1944$  resolution. A total of 108 microscopic blood images have been taken from the dataset (Labati et al., 2011). The view of infected and healthy blood cell images is shown in Figure 2.

#### 3.2 | Proposed work

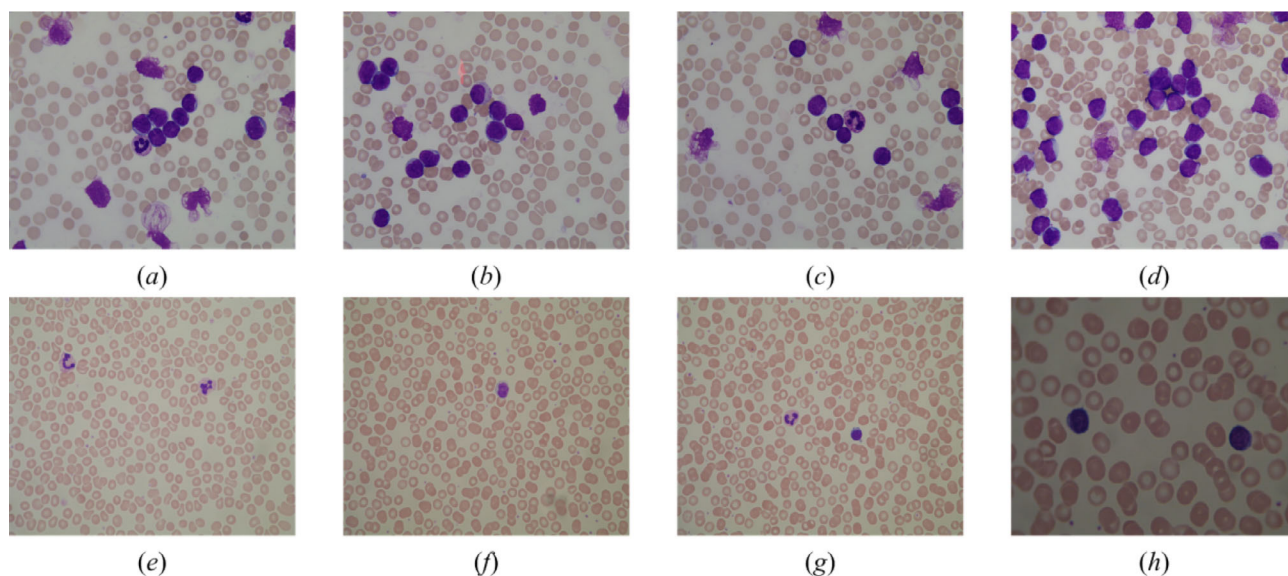
The workflow of the proposed work is given in Figure 3. The proposed work contains various modules, that is, preprocessing module, model training module and decision module. The description of each module is briefly explained in subsequent order.

##### 3.2.1 | Preprocessing

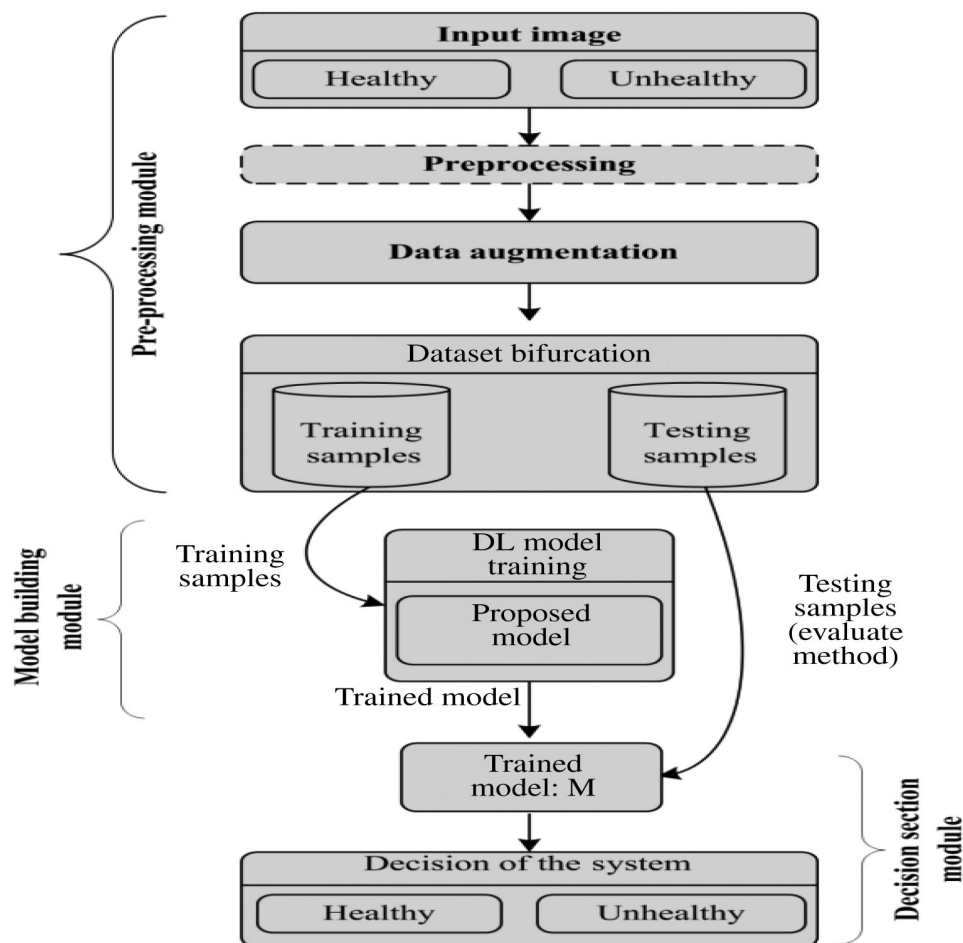
The pre-processing section includes different image enhancement methods to increase the visual perception of humans as well as increase the accuracy of classifications (Rashid et al., 2020; Rawat et al., 2015). It also incorporates the various enhancement methods of spatial domain like histogram equalization for improving the image quality and various frequency domain filtering techniques, which are used to reduce the noise and to give a smoothing effect to the image. The experimental workflow containing pre-processing is shown in Figure 3.

##### 3.2.2 | Data augmentation

In order to resolve the problem related to the imbalanced dataset, which would be the result of the heterogeneous distribution of WBCs, classes with a small amount of data items are subjected to oversampling. Data preparation is done for classes with limited data by cropping various



**FIGURE 2** Sample images taken from dataset18. (A-D) Infected blood cell and (E-H) healthy blood cell image



**FIGURE 3** Experimental method containing pre-processing stage

images related to the same cell at marginally different centers. This helps in providing more diverse data and, thus, increasing the amount of data, which reduces the inconsistency or overtraining of classifiers (Hassan et al., 2020; Liang et al., 2018; Shafique & Tehsin, 2018).

In this work, for ensuring the suitable quantity of data required in training, module images are first augmented and then passed for convolution. Every operation uses a predefined probability; that is, 0.25 probability would mean that out of 100 images, and 25 would be affected

through augmentation. The operations for augmentation involve maximum rotation of  $50^\circ$  to the left and right with 0.8 probability, flipping image left to right or upside down with 0.5 probability and zooming the random area of an image with 0.5 probability and 0.1–0.8 zoom ratio. The final model, thus, uses 6000 images that are produced by augmenting 108 images. Final augmented images are shown in Figure 4.

### 3.2.3 | Training and testing dataset preparation

The prepared dataset of images is divided into testing and training data sets with a ratio of 80% and 20%, respectively. The  $n$  cross-validation is used for training and optimization of hyper-parameters like weight decay, learning rate decay, momentum and learning ratio.

### 3.2.4 | Classification

In this work, to detect and classify the leukocyte cell with high accuracy, deep convolutional neural network is used. The description of DCNN and proposed DCNN is given in the next section (LeCun et al., 2015).

#### *Deep convolutional neural networks*

DCNN is among the most popular techniques used in image classification (Abdolali et al., 2020; Kaur & Gandhi, 2020; Kumar et al., 2021; Mondal et al., 2019). There are three ways for implementing DCNN-based classification of medical images, that is, DCNN as feature generator, as transfer learning and as full-training-based DCNN. The given network needs to be designed from scratch for the full training that may need a large quantity of labelled data, which is not easy to obtain in the medical domain. While in transfer learning approach, fine-tuning of non-medical data is done with the help of the application-specific pre-trained network. In the third approach, extraction of features is done with certain layers of pre-trained network, which are then used for training the external classifiers. Figure 5 illustrates the structure of DCNN.

Presently, researchers are creating robust artificial intelligence-based computational systems through deep learning methods (Duan & Yu, 2011; Nemane et al., 2013). These techniques have been outperformed by CNNs for extractions of features by yielding far more accuracy in diagnosis (Mansour, 2020; Wu et al., 2016; Yu et al., 2017). Therefore, CNNs have been proposed for carrying out our approach for classification of peripheral blood images of WBCs. Recently, studies have been done for utilizing the power of CNNs in two different ways, that is, first, by training the network with a large set of data and, second, by utilizing the pre-trained network related to transfer of learning (Choi et al., 2017; Hassan et al., 2019; Kutia et al., 2019; Rehman et al., 2018; Tajbakhsh et al., 2016; Tiwari et al., 2018; Zhao et al., 2017).

#### *Proposed model description*

In this model, we first had to resize the image, because the model designed would accept images in the resolution of  $128 \times 128$ , so the images had to be scaled down from the original resolution to the required size; then, the model first consisted of four convolutional layers (C1, C2, C3 and C4), each having 128 of all the convolutional layers that were followed by max pooling of  $2 \times 2$  (P1, P2, P3 and P4) neurons. The activation function used in each of the layer was 'sigmoid', 'Relu', 'tanh' and 'LeakyReLU'; each layer also had a max pooling of  $(3 \times 3)$  after that I used a dense layer (D1) to condense the result of the images from a wide network and spread it again after that before the final result extraction, I flattened the input and used another dense layer (D2) with the 'softmax' activation function to achieve the final result of the model, as shown in Figure 6.

#### *Model details*

The details description of each layer of proposed model is given in Figure 7.

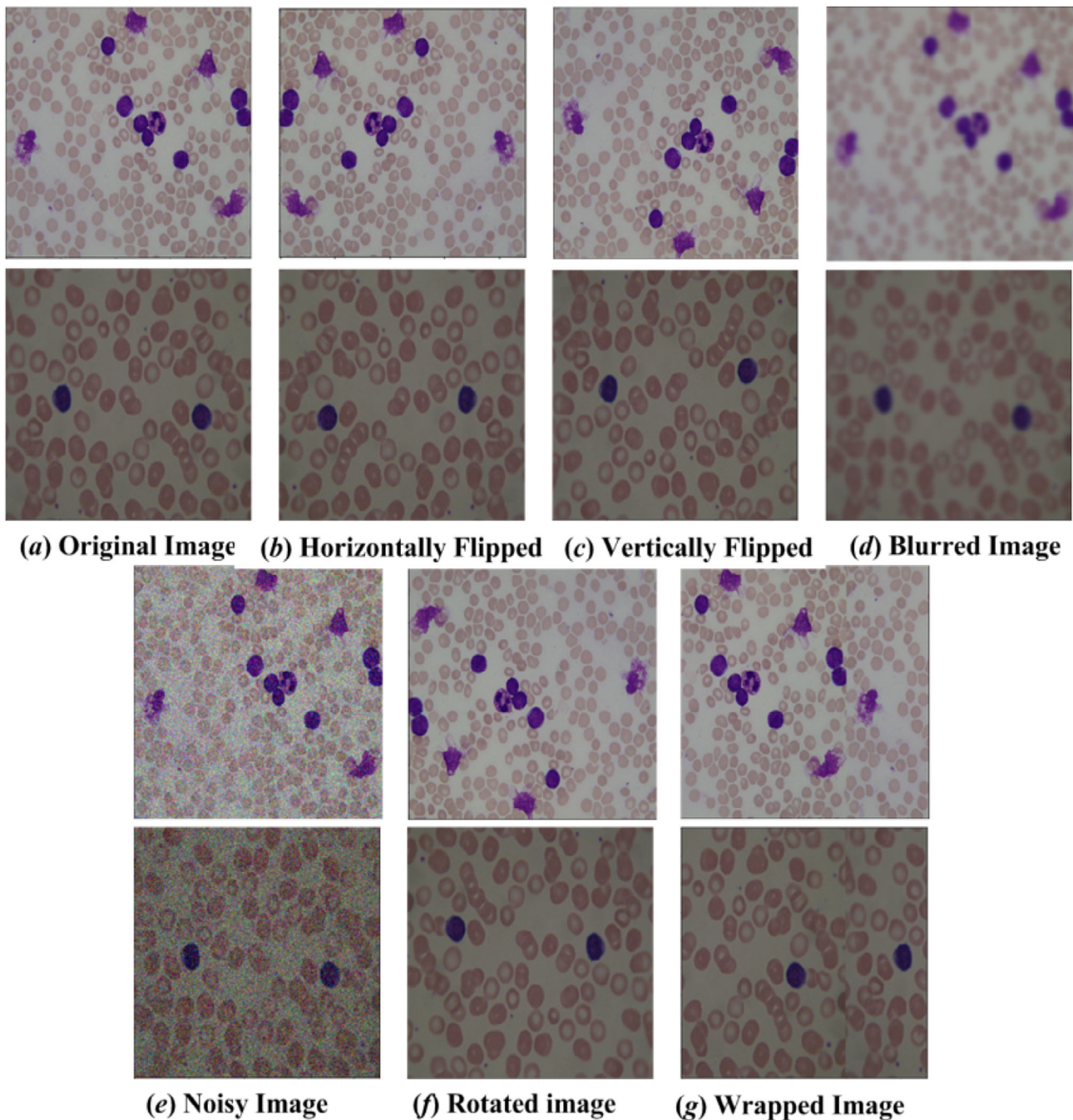
## 4 | EXPERIMENTAL TEST AND RESULTS

The experiment conducted on ALL-IDB1 dataset on workstation with specification of the system is given as: 3.2 GHz Intel Xeon W-2014 CPU, RAM 64 GB, 2 TB SATA HDD and 256 GB SSD and NVIDIA Quadro P10004GB. All the images are stored at this system, and Python environment is used for experimentations.

### 4.1 | Experiment description

In this work, initially, experiment is carried out without any data augmentation with four different type activation functions. In next experiment, data augmentation is performed, and 6000 virtual images are generated and train the model using augmented dataset. All experiments



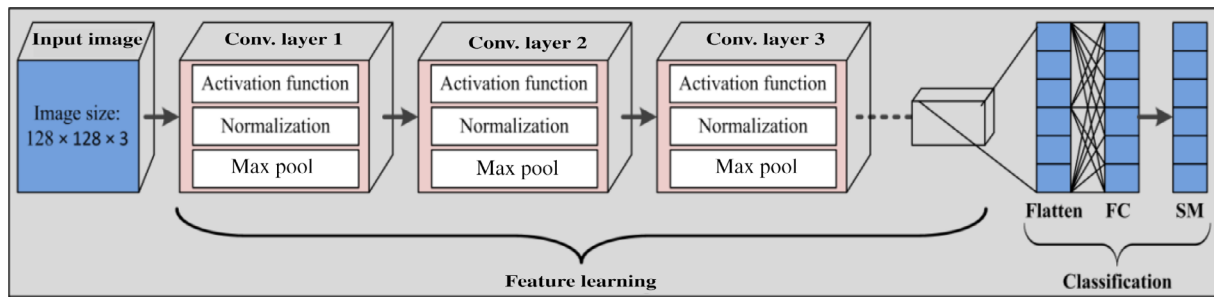


**FIGURE 4** Augmented images from the original image. (a) Original image. (b) Horizontally flipped. (c) Vertically flipped. (d) Blurred image. (e) Noisy image. (f) Rotated image. (g) Wrapped image

have been performed for four different activation functions '*sigmoid*', '*tanh*', '*ReLU*' and '*LeakyReLU*'. The description of experiment is shown in Table 1.

## 4.2 | Performance evaluation

**Classification Rate/Accuracy:** This is generally defined as the accuracy of the model. Mathematically, it is described as the ratio of sum of true positive (TP), true negative (TN) to the sum of TP, TN, false positive (FP) and false negative (FN) using confusion matrix (Rawat et al., 2015; Wu et al., 2016) and is represented in Equation (1) as:



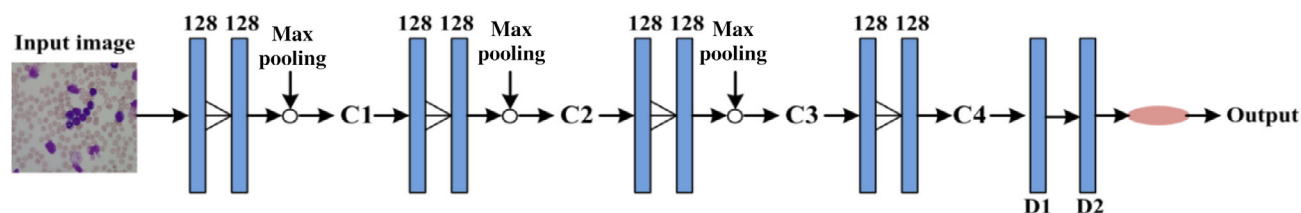
**FIGURE 5** Illustration of DCNN architecture. Conv., convolutional layer; FC, fully connected layer; SM, softmax layer

| Model: "sequential"            |                       |         |
|--------------------------------|-----------------------|---------|
| Layer (type)                   | Output Shape          | Param # |
| conv2d (Conv2D)                | (None, 118, 118, 128) | 1280    |
| activation (Activation)        | (None, 118, 118, 128) | 0       |
| max_pooling2d (MaxPooling2D)   | (None, 59, 59, 128)   | 0       |
| conv2d_1 (Conv2D)              | (None, 57, 57, 128)   | 147584  |
| activation_1 (Activation)      | (None, 57, 57, 128)   | 0       |
| max_pooling2d_1 (MaxPooling2D) | (None, 28, 28, 128)   | 0       |
| conv2d_2 (Conv2D)              | (None, 26, 26, 128)   | 147584  |
| activation_2 (Activation)      | (None, 26, 26, 128)   | 0       |
| max_pooling2d_2 (MaxPooling2D) | (None, 13, 13, 128)   | 0       |
| conv2d_3 (Conv2D)              | (None, 11, 11, 128)   | 147584  |
| activation_3 (Activation)      | (None, 11, 11, 128)   | 0       |
| max_pooling2d_3 (MaxPooling2D) | (None, 5, 5, 128)     | 0       |
| dense (Dense)                  | (None, 5, 5, 64)      | 8256    |
| activation_4 (Activation)      | (None, 5, 5, 64)      | 0       |
| flatten (Flatten)              | (None, 1600)          | 0       |
| dense_1 (Dense)                | (None, 1)             | 1601    |
| activation_5 (Activation)      | (None, 1)             | 0       |
| Total params: 453,889          |                       |         |
| Trainable params: 453,889      |                       |         |
| Non-trainable params: 0        |                       |         |

**FIGURE 6** Proposed model description

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

*Recall:* Mathematically, this is the ratio of the TP to the sum of TP and FN and is represented in Equation (2) as:



#### Model details:

C1: Convolutional layer 1 [128 × 128, ReLu, Max Pool: 2 × 2, stride: 2];

C2: Convolutional layer 2 [128 × 128, ReLu, Max Pool: 2 × 2, stride: 2];

C3: Convolutional layer 3 [128 × 128, ReLu, Max Pool: 2 × 2, stride: 2];

C4: Convolutional layer 4 [128 × 128, ReLu];

D1: Dense layer 1 [Neurons: 32, ReLu];

D2: Dense layer 2 [Neurons: 2, Softmax].

**FIGURE 7** Proposed model description

**TABLE 1** Description of experiment

| Experiment no. | Description   | Total no. of samples |         |
|----------------|---|----------------------|---------|
|                |   | Training             | Testing |
| Experiment 1   | Development of deep learning-based model for identification of lymphoblast cell from microscopic blood images without data augmentation | 54                   | 54      |
| Experiment 2   | Development of deep learning-based model for identification of lymphoblast cell from microscopic blood images with data augmentation    | 6000                 | 108     |

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

*Precision*: Mathematically, this is the ratio of TP and TP, FP and is represented in Equation (3) as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

*F-measure*: This is the ratio of twice of the product of recall, precision and the sum of recall, precision and is represented in Equation (4) as:

$$\text{F-measure} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (4)$$

## 4.3 | Result analysis

The result analysis has been carried out in terms of various experiments as given in the following.

### 4.3.1 | Experiment 1

In this experiment, the complete set of input image is equally divided into two set; one is called training set, and the other is called testing set. The training and testing sets are evaluated for four different functions. The obtained result for each function is given in Table 2.

### 4.3.2 | Experiment 2

In this experiment, a set of 6000 samples are used to train the model using four different activation functions. The trained model is tested on 108 samples of testing set. After the extensive computation of the proposed model, the results yielded through the experimentation are given in Table 3.



**TABLE 2** Obtained results for Experiment 1

| Sr. no. | Activation function |         | Confusion matrix |          | Accuracy (%) | Recall (%) | Precision (%) | F-measure (%) |
|---------|---------------------|---------|------------------|----------|--------------|------------|---------------|---------------|
|         |                     |         | ALL              | Healthy  |              |            |               |               |
| 1       | sigmoid             | ALL     | (TN): 24         | (FN): 1  | 92.6         | 96.3       | 89.6          | 92.8          |
|         |                     | Healthy | (FP): 3          | (TP): 26 |              |            |               |               |
| 2       | tanh                | ALL     | (TN): 23         | (FN): 2  | 88.8         | 92.6       | 86.2          | 89.3          |
|         |                     | Healthy | (FP): 4          | (TP): 25 |              |            |               |               |
| 3       | ReLu                | ALL     | (TN): 24         | (FN): 1  | 96.2         | 96.5       | 96.5          | 96.5          |
|         |                     | Healthy | (FP): 1          | (TP): 28 |              |            |               |               |
| 4       | LeakyReLu           | ALL     | (TN): 23         | (FN): 2  | 92.6         | 93.1       | 93.1          | 93.1          |
|         |                     | Healthy | (FP): 2          | (TP): 27 |              |            |               |               |

**TABLE 3** Obtained results for Experiment 2

| Sr. no. | Activation function |         | Confusion matrix |          | Accuracy (%) | Recall (%) | Precision (%) | F-measure (%) |
|---------|---------------------|---------|------------------|----------|--------------|------------|---------------|---------------|
|         |                     |         | ALL              | Healthy  |              |            |               |               |
| 1       | sigmoid             | ALL     | (TN): 45         | (FN): 4  | 93.5         | 93.3       | 94.9          | 94.1          |
|         |                     | Healthy | (FP): 3          | (TP): 56 |              |            |               |               |
| 2       | tanh                | ALL     | (TN): 42         | (FN): 7  | 92.6         | 88.8       | 94.9          | 91.7          |
|         |                     | Healthy | (FP): 3          | (TP): 56 |              |            |               |               |
| 3       | ReLu                | ALL     | (TN): 48         | (FN): 1  | 98.1         | 98.3       | 98.3          | 98.3          |
|         |                     | Healthy | (FP): 1          | (TP): 58 |              |            |               |               |
| 4       | LeakyReLu           | ALL     | (TN): 47         | (FN): 2  | 95.3         | 96.5       | 94.9          | 95.7          |
|         |                     | Healthy | (FP): 3          | (TP): 56 |              |            |               |               |

After the execution of two extensive experiments, the major findings are observed as the following:

- For experiment, the complete set of ALL-IDB-1 is equally divided into two sets, that is, training and testing sets. In this experiment, 54 images are used for training, and the remaining 54 images are passed to trained model, and obtained results are reported in Table 1. After that, it has been found that the maximum accuracy of 96.2% is with ReLu activation function. The value of recall, precision and F-measure is obtained as 96.5%, 96.5% and 96.5%, respectively, for ReLu function.
- Experiment 1 also shows the result for four different activation functions. The accuracy of 92.6%, 88.8% and 92.6% is reported for sigmoid, tanh and LeakyReLu activation functions, respectively.
- The extensive experiment has been carried out for 50 epochs, and the obtained training, validation accuracy, training loss and validation loss for each epoch are shown in Figure 8.
- In Experiment 2, an augmented set of 6000 samples is used for the training purpose using four activation functions, and later on, trained model is tested on 108 samples. After testing, it has been seen that the model yields 98.1% of accuracy using ReLu function and 98.3% recall, 98.3% precision and 98.3% F-measure, respectively.
- Using Experiment 2, the accuracy of sigmoid, tanh and LeakyReLu functions is obtained as 93.5%, 92.6% and 95.3%, respectively. The obtained results show the importance of data augmentations and number of training samples for successful implementation of deep learning-based model.
- The complete experiment is executed for 50 epochs, and training loss, training accuracy, validation loss, and validation accuracy are measured for each epoch. The average training loss is 0.01, average training accuracy is reported as 99.8%, average validation loss is noted as 26.6 and average validation accuracy is reported as 98.1%. The performance curve for best performing model is shown in Figure 9.

#### 4.4 | Comparative analysis

The comparative analysis of results for Experiment 1 and Experiment 2 is given in Table 4. It is noted that the maximum accuracy of 98.1% is yielded by proposed model. The ReLu activation function gives maximum accuracy for both experiments. The model proposed in Experiment

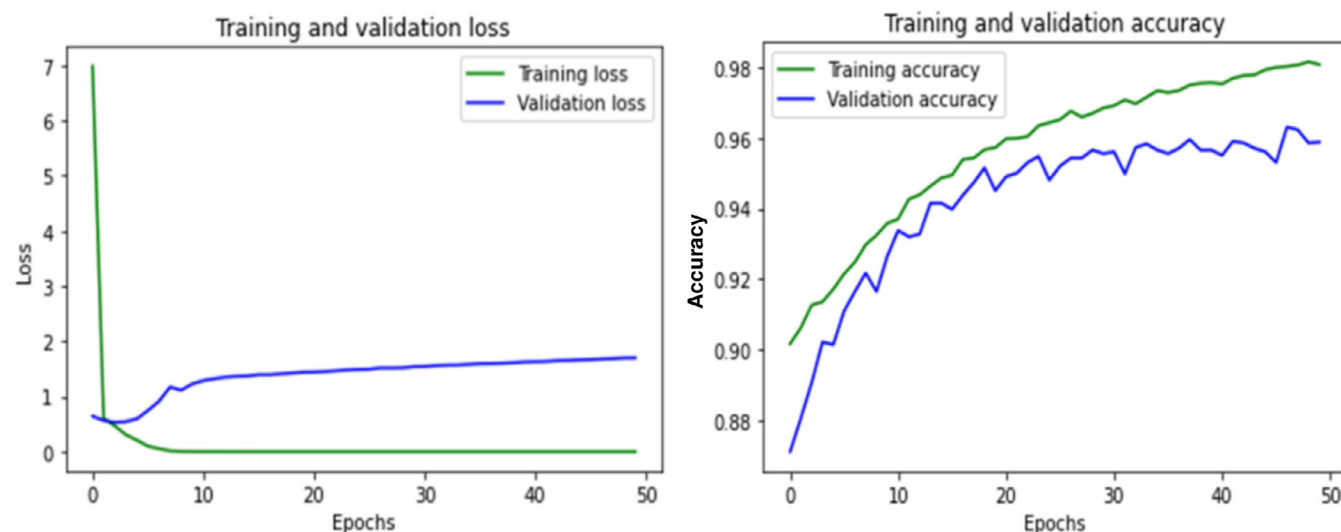


FIGURE 8 Performance curve for best performing model of Experiment 1

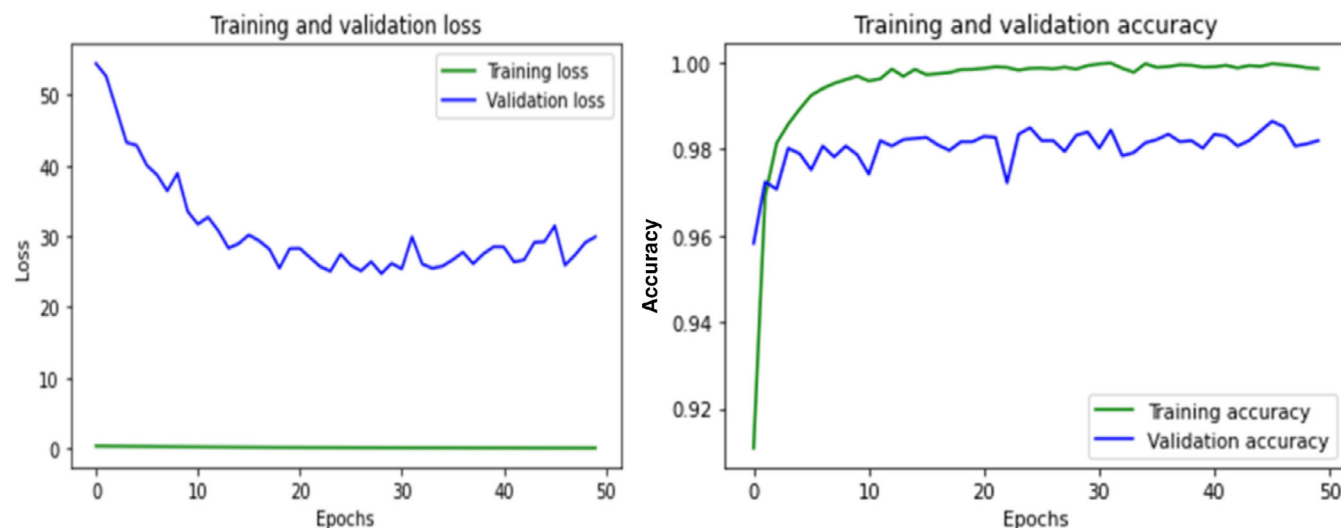


FIGURE 9 Performance curve for best performing model of Experiment 2

TABLE 4 Comparative analysis for Experiment 1 and Experiment 2

| Experiment no. | Number of samples | Accuracy (%) | Recall (%) | Precision (%) | F-measure (%) |
|----------------|-------------------|--------------|------------|---------------|---------------|
| Experiment 1   | 54                | 96.2         | 96.5       | 96.5          | 96.5          |
| Experiment 2   | 108               | 98.1         | 98.3       | 98.3          | 98.3          |

2 yields the highest recall as 98.3%, precision as 98.3% and F-measure as 98.3%. The obtained accuracy also reflects the significance of training samples and role of data augmentation. Since Experiment 1 model yields less accuracy, the model is trained by limited number of samples, while Experiment 2 model trained using too much number of samples and provides good accuracy respect to Experiment 1 model. Thus, it has been concluded that the deep learning model must be trained with appropriate number of training samples.

## 5 | CONCLUSION

Precisely identifying the different elements of blood is most important part of diagnosing the present of abnormalities in the field of computer-aided haematological applications. There have been various studies implementing the deep learning or machine learning-based networks for

diagnosing haematological disorders like acute leukaemia. In the present work, a deep learning technique is implemented for classification of lymphoblast cell from normal cell, which includes augmentation of images for enhancing the quantity of training set and then training the deep convolutional neural network classifier with the help of given data set. This network is then tested with the test cases to correctly distinguish between normal and abnormal cells, which resulted in drawing the conclusion that the proposed technique can classify the lymphoblast better than the present state-of-art techniques. It can be stated that the proposed DCNN method implemented with ReLU activation function yields an accuracy of 98.1% for classifying lymphoblast from given microscopic images of blood. For future scopes, a goal has been set to implement the proposed method to classify other cancers like CLL, CML and AML with sufficient amount of data belonging to both healthy as well as unhealthy individuals in correct proportion.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Abhishek Kumar  <https://orcid.org/0000-0002-9925-8137>

Jyoti Rawat  <https://orcid.org/0000-0003-3331-7568>

Indrajeet Kumar  <https://orcid.org/0000-0003-2814-2900>

Mamoon Rashid  <https://orcid.org/0000-0002-8302-4571>

Yasser D. Al-Otaibi  <https://orcid.org/0000-0002-1464-8401>

Usman Tariq  <https://orcid.org/0000-0001-7672-1187>

## REFERENCES

- Abdolali, F., Kapur, J., Jaremko, J. L., Noga, M., Hareendranathan, A. R., & Punithakumar, K. (2020). Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks. *Computers in Biology and Medicine*, 122, 103871.
- Acevedo, A., Alf  rez, S., Merino, A., Puigv  , L., & Rodellar, J. (2019). Recognition of peripheral blood cell images using convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 180, 105020.
- Arslan, S., Ozyurek, E., & Gunduz Demir, C. (2014). A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images. *Cytometry Part A*, 85(6), 480–490.
- Choi, J. W., Ku, Y., Yoo, B. W., Kim, J. A., Lee, D. S., Chai, Y. J., Kong, H. J., & 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), e0189259.
- Di Ruberto, C., Loddo, A., & Puglisi, G. (2020). Blob detection and deep learning for leukemic blood image analysis. *Applied Sciences*, 10(3), 1176.
- Duan, J. & Yu, L. (2011). A WBC segmentation method based on HSI color space. Paper presented at the 2011 4th IEEE International Conference on Broadband Network and Multimedia Technology (pp. 629–632). IEEE.
- Hassan, H., Bashir, A. K., Abbasi, R., Ahmad, W., & Luo, B. (2019). Single image defocus estimation by modified gaussian function. *Transactions on Emerging Telecommunications Technologies*, 30(6), e3611.
- Hassan, H., Bashir, A. K., Ahmad, M., Menon, V. G., Afridi, I. U., Nawaz, R., & Luo, B. (2020). Real-time image dehazing by superpixels segmentation and guidance filter. *Journal of Real-Time Image Processing*, 18(5), 1555–1575. <https://link.springer.com/article/10.1007/s11554-020-00953-4>
- Hegde, R. B., Prasad, K., Hebbar, H., & 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), 382–392.
- Jiang, M., Cheng, L., Qin, F., Du, L., & Zhang, M. (2018). White blood cells classification with deep convolutional neural networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(09), 1857006.
- Kaur, T., & Gandhi, T. K. (2020). Deep convolutional neural networks with transfer learning for automated brain image classification. *Machine Vision and Applications*, 31(3). <https://doi.org/10.1007/s00138-020-01069-2>
- Kumar V. D. A., Sharmila S., Kumar A., Bashir A. K., Rashid M., Gupta S. K., & Alnumay W. S. (2021). A novel solution for finding postpartum haemorrhage using fuzzy neural techniques. *Neural Computing and Applications*, <https://doi.org/10.1007/s00521-020-05683-z>
- Kutia S., Chauhdary S. H., Iwendi C., Liu L., Yong W., & Bashir A. K. (2019). Socio-Technological Factors Affecting User's Adoption of eHealth Functionalities: A Case Study of China and Ukraine eHealth Systems. *IEEE Access*, 7, 90777–90788. <https://doi.org/10.1109/access.2019.2924584>
- Labati, R. D., Piuri, V. & Scotti, F. (2011). All-IDB: The acute lymphoblastic leukemia image database for image processing. Paper presented at the 2011 18th IEEE International Conference on Image Processing (pp. 2045–2048). IEEE.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Levy, D. & Jain, A. (2016). Breast mass classification from mammograms using deep convolutional neural networks. arXiv preprint arXiv:1612.00542.
- Lewis, S. J., Gandomkar, Z., & Brennan, P. C. (2019). Artificial intelligence in medical imaging practice: Looking to the future. *Journal of Medical Radiation Sciences*, 66(4), 292–295.
- Liang, G., Hong, H., Xie, W., & Zheng, L. (2018). Combining convolutional neural network with recursive neural network for blood cell image classification. *IEEE Access*, 6, 36188–36197.
- Mansour R. F. (2020). A robust deep neural network based breast cancer detection and classification. *International Journal of Computational Intelligence and Applications*, 19(01), 2050007. <https://doi.org/10.1142/s1469026820500078>

- Mondal, S., Agarwal, K. & Rashid, M. (2019). Deep Learning Approach for Automatic Classification of X-Ray Images using Convolutional Neural Network. Paper presented at the 2019 fifth international conference on image information processing (ICIIP) (pp. 326–331). IEEE.
- Nemane, J. B., Chakkarwar, V. A., & Lahoti, P. B. (2013). White blood cell segmentation and counting using global threshold. *International Journal of Emerging Technology and Advanced Engineering*, 3(6), 639–643.
- Rashid, M., Singh, H., & Goyal, V. (2020). The use of machine learning and deep learning algorithms in functional magnetic resonance imaging—A systematic review. *Expert Systems*, 37(6), e12644.
- Rawat, J., Singh, A., Bhadauria, H. S., & Virmani, J. (2015). Computer aided diagnostic system for detection of leukemia using microscopic images. *Procedia Computer Science*, 70, 748–756.
- Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2018). Classification of acute lymphoblastic leukemia using deep learning. *Microscopy Research and Technique*, 81(11), 1310–1317.
- Safuan, S. N. M., Tomari, M. R. M., Zakaria, W. N. W., Mohd, M. N. H., & Suriani, N. S. (2020). Investigation of white blood cell biomaker model for acute lymphoblastic leukemia detection based on convolutional neural network. *Bulletin of Electrical Engineering and Informatics*, 9(2), 611–618.
- Shafique, S., & Tehsin, S. (2018). Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technology in Cancer Research & Treatment*, 17, 1533033818802789.
- Shah, A. A., Parah, S. A., Rashid, M., & Elhoseny, M. (2020). Efficient image encryption scheme based on generalized logistic map for real time image processing. *Journal of Real-Time Image Processing*, 17(6), 2139–2151.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5), 1299–1312.
- Tiwari, P., Qian, J., Li, Q., Wang, B., Gupta, D., Khanna, A., Rodrigues, J. J., & de Albuquerque, V. H. C. (2018). Detection of subtype blood cells using deep learning. *Cognitive Systems Research*, 52, 1036–1044.
- Vogado, L. H., Veras, R. M., Araujo, F. H., Silva, R. R., & Aires, K. R. (2018). Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. *Engineering Applications of Artificial Intelligence*, 72, 415–422.
- Wu, Z., Ding, X., & Zhang, G. (2016). A novel method for classification of ECG arrhythmias using deep belief networks. *International Journal of Computational Intelligence and Applications*, 15(04), 1650021.
- Yu, L., Wang, Z., Tian, S., Ye, F., Ding, J., & Kong, J. (2017). Convolutional neural networks for water body extraction from Landsat imagery. *International Journal of Computational Intelligence and Applications*, 16(01), 1750001.
- Zhao, J., Zhang, M., Zhou, Z., Chu, J., & Cao, F. (2017). Automatic detection and classification of leukocytes using convolutional neural networks. *Medical & Biological Engineering & Computing*, 55(8), 1287–1301.

## AUTHOR BIOGRAPHIES

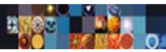
**Abhishek Kumar** is working as an Assistant Professor, School of Computer Science and I.T. at JAIN, Deemed to be University, Bangalore and have worked with Department of Computer Science, Institute of Science at Banaras Hindu University. Dr Abhishek Kumar is Apple Certified Associate (USA), Adobe Certified Educator (USA) Certified by Autodesk. Dr Abhishek holding 6 Patents in the field of V.R. and IOT Published 30+ Research papers Scopus/WOS indexed. Dr Abhishek has trained over 50,000+ students across the globe from 153 Countries; top 5 countries are India, Germany, United States, Spain Australia. He has completed their professional studies related to Animation, computer graphics, Virtual reality, Stereoscopy, Filmmaking, Visual Effects photography from the Norwich University of Arts, University of Edinburg, Wizcraft MIME FXPHD, Australia

**Jyoti Rawat** is working as an Assistant Professor in School of Computing, DIT University, Dehradun, Uttarakhand, India. She received her PhD in Computer Science and Engineering from G B Pant Engineering College, Pauri Garhwal, Uttarakhand, India. Her area of research interests include the application of digital image processing, machine learning, computer vision and soft computing techniques for the analysis of medical images.

**Indrajeet Kumar** is working as an Assistant Professor in Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, Uttarakhand, India. He received his PhD in Digital Image Processing from G B Pant Institute of Engineering and Technology, Pauri Garhwal, Uttarakhand, India. His area of research interests include Digital Image Processing, Biomedical Imaging, Machine learning, and Deep learning.

**Mamoon Rashid** is Assistant Professor in Department of Computer Engineering, Faculty of Science and Technology, Vishwakarma University, Pune, India. He has published 80+ papers indexed in SCI/SCIE journals and Conferences of International repute. He is a regular contributor to monthly Editorial of CSI Communications and serves as in Editorial Review Board for many journals. His research interests include Big Data Analytics, Machine Learning, Neuro Imaging, & Image Processing.

**Kamred Udhham Singh** is working as Postdoctoral Research Fellow in the Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan. He received his PhD from Baranas Hindu University, India. His area of research interests include Digital Image Processing, Steganography and Image Watermarking.



**Yasser D. Al-Otaibi** is currently an Assistant Professor in the Department of Information Systems at the Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia. He received a Ph.D. degree in information systems from Griffith University, Australia, in 2018. His current research interests include IT adoption and acceptance, wireless sensor networks, and IoT.

**Usman Tariq** is an associate professor with the college of computer engineering and sciences at PSAU. He holds a Ph.D. from the Ajou University, South Korea and led the design of a global data infrastructure simulator modeling, to evaluate the impact of competing architectures on the performance, availability, and reliability of the system for Industrial IoT infrastructure. His international collaborations/collaborators include but not limited to: NYIT, Ajou University, PSU, University of Sherbrooke, COMSATS, NUST, UET, National Security Research Institute (NSR), Embry-Riddle Aeronautical University, Korea University, Manchester Metropolitan University, University of Bremen and Virginia Commonwealth University. Currently he is interested in applied cyber security, advanced topics in Internet of Things, and health informatics. His research focus is on the theory of large complex networks, which includes network algorithms, stochastic networks, network information theory, and large-scale statistical inference.

**How to cite this article:** Kumar, A., Rawat, J., Kumar, I., Rashid, M., Singh, K. U., Al-Otaibi, Y. D., & Tariq, U. (2021). Computer-aided deep learning model for identification of lymphoblast cell using microscopic leukocyte images. *Expert Systems*, e12894. <https://doi.org/10.1111/exsy.12894>