

A mobile application based on efficient lightweight CNN model for classification of B-ALL cancer from non-cancerous cells: A design and implementation study



Azamossadat Hosseini ^a, Mohammad Amir Eshraghi ^b, Tania Taami ^c, Hamidreza Sadeghsalehi ^d, Zahra Hoseinzadeh ^e, Mustafa Ghaderzadeh ^{f,*}, Mohammad Rafiee ^g

^a Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

^b School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

^c Department of Computer Science, Tallahassee, FL, USA

^d Department of Neuroscience, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran

^e Physical Education Faculty, University of Tabriz, Tabriz, Iran

^f Department of Artificial Intelligence, Smart University of Medical Science, Tehran, Iran

^g Department of Medical Laboratory Sciences, School of Paramedical Sciences, Zanjan University of Medical Science, Zanjan, Iran

ARTICLE INFO

Keywords:

Lightweight
Mobile application
Acute lymphoblastic leukemia
efficientnetb0
mobilenetv2
NASNet mobile

ABSTRACT

Background: B-cell acute lymphoblastic leukemia (B-ALL) is one of the most widespread cancers, and its definitive diagnosis demands invasive and costly diagnostic tests with side effects for patients. Access to definitive diagnostic equipment for B-ALL is limited in many geographical areas. Blood microscopic examination has always been a major B-ALL screening and diagnosis technique. Still, the examination of blood microscopically by laboratory personnel and hematologists is riddled with disadvantages. Meanwhile, AI techniques can achieve remarkable results in blood microscopy image analysis. The present study aimed to design and implement a well-tuned based on deep CNN to detect B-ALL cases from hematogones and then determine the B-ALL subtype.

Methods: Based on the well-designed and tuned model, a mobile application was also designed for screening B-ALL from non-B-ALL cases. In the modeling stage, a unique segmentation technique was used for color thresholding in the color LAB space. By applying the K-means clustering algorithm, and adding a mask to the clustered images, a segmented image was obtained to eliminate unnecessary components. After comparing the efficiency of three notable architectures of lightweight CNN (EfficientNetB0, MobileNetV2, and NASNet Mobile), the most efficient model was selected, and the proposed model was accordingly configured and tuned.

Results: The proposed model achieved an accuracy of 100%. Finally, a mobile application was designed based on this state-of-the-art model. In the real laboratory setting, the mobile application based on the proposed model classified B-ALL cases from other classes and achieved a sensitivity and specificity of 100% as a robust screening tool.

Conclusions: The application that relies on preprocessing and DL algorithms can be used as a powerful screening tool by hematologists and clinical specialists to ignore or minimize unnecessary bone marrow biopsy cases and decrease the B-ALL diagnosis time.

1. Background

B-Acute lymphoblastic leukemia (B-ALL) is one of the most prevalent types of leukemia in different age groups. An accurate diagnosis of leukemia demands a range of tools and methods for diagnostics. The definitive method for diagnosis involves the analysis of bone mass

aspiration (BMA) or bone mass biopsy (BMB) samples as an invasive laboratory method that requires sophisticated equipment. As most patients with B-ALL are children, these invasive diagnostic methods cause complications for them, and the diagnosis may take several days. Moreover, these surgical procedures impose heavy costs on healthcare systems [1,2] As a result, the laboratory staff's early analysis of blood

* Corresponding author.

E-mail address: Mustaf.ghaderzadeh@sbmu.ac.ir (M. Ghaderzadeh).

Table 1

The local dataset of B-ALL and its subclasses.

Type	Classes	Samples	Samples	Image Dimensions
Benign	Hematogones	512	25	1024 × 768
Malignant	Early Pre-B ALL	979	20	1024 × 768
	Pre-B ALL	955	21	1024 × 768
	Pro-B ALL	796	23	1024 × 768
		3242	89	

microscopic samples is critical in diagnosing leukemia. On the other hand, the analysis of the sheer amount of blood microscopic samples is a problem for laboratories, resulting in diagnostic errors. B-ALL detection is prone to error due to the time-consuming nature of the blood microscopic survey under the microscope, the expertise of the laboratory staff, and the fatigue they experience daily. Furthermore, the signs of some autoimmune and infectious diseases (caused by an increase in hematogones or benign lymphocytes) resemble B-ALL, leading to a misdiagnosis of cancer and confusing the hematologist [1,3,4]. Hematogones, or benign B-ALL, have a cellular heterogeneity similar to B-ALL cells and, in most cases, a morphology similar to that of cancerous B cells, which is why they are sometimes mistaken for cancer cells. Meanwhile, hematogones do not require complex treatments. It is complicated to differentiate hematogones from lymphoblasts on the basis of morphology and difficult process, by which the hematologist distinguishes them from malignant lymphoblasts under a microscope with great difficulty. The definitive method involves the use of bone marrow samples via flow cytometry [5–7].

In this study, a model relying on deep learning (DL) concepts is proposed for B-ALL diagnosis and classification. Previous studies had used classical Artificial Intelligence methods to propose models for B-ALL diagnosis, but these methods were purely theoretical and not implemented in clinical settings. Herein, after introducing a tuned CNN model, its application was designed for mobile phones and PDAs to be implemented in the test phase in a laboratory setting.

2. Methods

To the best of our knowledge, there is no mobile application for B-ALL detection and classification using blood microscopic images and a lightweight CNN model. Thus, the application developed herein can bridge the gap in the literature and serve as a useful tool for clinicians and researchers working in the laboratory. The main objective of this research was to design a CNN algorithm-based application for portable devices such as mobile phones and PDAs that could be applied to blood microscopic images to diagnose B-ALL and classify its classes (Early-Pre-B, Pro-B-ALL, and Pre-B-ALL).

2.1. Dataset

A local dataset was used to train and assess the suggested model's efficiency. These images were taken at several Tehran hospitals (Iran). A total of 3242 blood microscopic images were obtained from 89 patients suspected of B-ALL, whose blood samples were prepared and stained by skilled laboratory staff. There are benign and malignant classes in the dataset. The first class (benign) contains hematogones that greatly resemble B-ALL cases, but this hematopoietic precursor is benign, does not require chemotherapy, and usually disappears into the blood on its own. There are three subtypes of malignant lymphoblasts in the second class (ALL). The images were captured with a Zeiss camera from the optical microscope and saved as JPG files. This cell type and its subclasses were identified using flow cytometry. After shuffling the data, Table 1 displays detailed information about this dataset.

In the present study, all the collected slides were first stained by a number of laboratory specialists in a controlled environment in order to maintain the reliability of the images and prevent differences in the

structure of the images due to different types of staining and photography. Additionally, a specified hematology specialist, the same Zeiss camera, and a microscope were all used to capture every image of the slides. These procedures were followed in an effort to uphold the concepts of consistency and dependability.

In computer-based vision recognition and its applications, pre-processing directly affects the performance of ML-based models. Pre-processing techniques can prepare input data, remove noise, improve diagnosis, highlight the useful features of the image, or even speed up DL model training [3, 18]. In the proposed method, three stages of data pre-processing were included: (1) preparation (decoding and resizing), (2) segmentation, and (3) augmentation, rescaling, and categorization.

2.2. Data preparation (decoding, resizing)

Images of research data obtained in PNG or JPG formats were converted into numeric arrays corresponding to their pixels (decoding). In order to preserve flexibility to the network design and decrease computational effort, input pictures are often downsized in CNNs. In the pre-trained network selected for this study, the inputs must have a certain size. Moreover, in many conventional architectures, the dimensions of the input images should be 224 × 224, and recent studies on medical diagnosis also use these dimensions. Thus, herein, the image dimensions were converted to 224 × 224.

2.3. Segmentation

In this stage, the ROI areas are segmented by depicting the important parts of the model. In this study, segmentation aimed to extract and display areas that are clinically significant to the clinical specialist and, therefore, the ML model. One of the goals of segmentation in this study was to eliminate and ignore unnecessary parts that are insignificant for learning models and disrupt the learning process as noise. As a way to sensitize the network to lymphoblast cells and enhance the model's precision and reliability, the input images were segmented using a unique method, that separated the blast cells from other blood parts (erythrocytes and plasma). For blast cells' segmentation, we used an approach based on color interval separation and color thresholding that was both successful and cost-efficient. By default, blood microscopic images in the RGB space contain three channels of red, green, and blue. Image segmentation under the clustering paradigm was employed herein. This segmentation method utilizes a robust clustering algorithm in the L*a*b color space of pixels. In this method, image segmentation is performed directly by adjusting each pixel to the corresponding cluster.

This color space was originally defined by CAE and specified by the International Commission on Illumination. A lab color space is a color-opponent space with dimension L for lightness and a and b for the color opponent dimensions, based on nonlinearly compressed CIE XYZ color space coordinates. The coordinates of the Hunter 1948 L, a, b color space are L, a, and b. The nonlinear relations for L*, a*, and b* are intended to mimic the nonlinear response of the eye. Furthermore, uniform changes of components in the L*a*b* color space aim to correspond to uniform variations in perceived color, so the relative perceptual differences between any two colors in L*a*b* can be approximated by treating each color as a point in a three-dimensional (3D) space (with three components L*, a*, b*) and taking the Euclidean distance between them [8,9].

Transformations 1 to 3 are required to convert the images from the RGB space given as (XYZ) into the LAB space:

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \quad (1)$$

$$a^* = 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \quad (2)$$

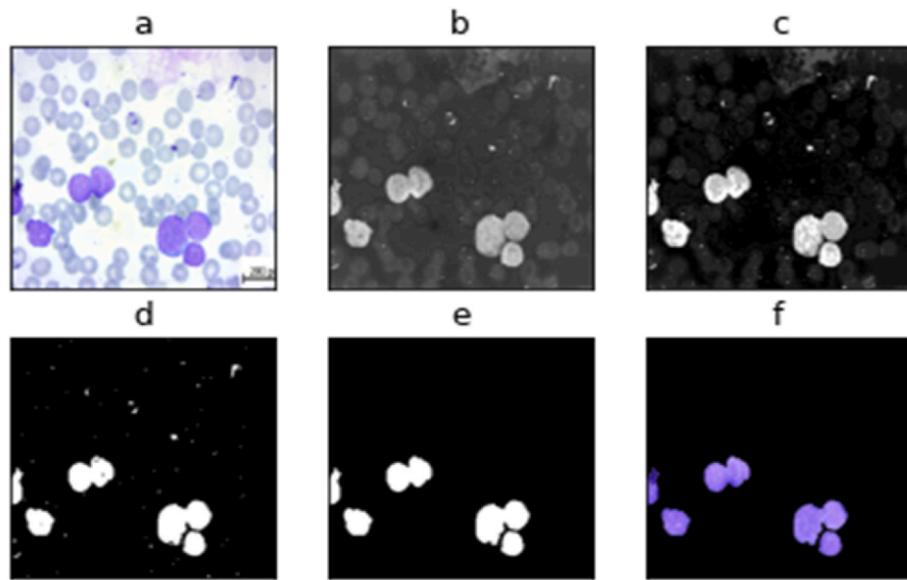


Fig. 1. Segmentation stages: a) RGB color space, b) LAB color space, c) K-means clustering d) Binary thresholding, e) Cleaning methods and mask generation, f) Applied Mask on the RGB original image. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$$b^* = 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \quad (3a)$$

where function f is obtained from Equation (4):

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \quad (4)$$

The $f(t)$ function was defined in two domains to prevent an infinite slope at $t = 0$. $f(t)$ is assumed to be linear below $t = t_0$, and is assumed to match the $t^{1/3}$ part of the function at t_0 in terms of both value and slope [9,10].

The blast cells were segregated by using this mask on the input pictures. This procedure is depicted in Fig. 1 for a sample.

After changing the color space channels from RGB to LAB, the A component was extracted from the LAB space due to its focus on blast color and greater differentiation between the blast and background, and the output of original images was obtained as B. Then, by applying the K-means clustering, the image outputs were obtained as C. Subsequently, by applying binary thresholding to the image output (Image d), many unnecessary background components were blurred or ignored. The image output at this stage had diagnostically non-significant components; thus, to extract the features of the blasts and ignore other features, a mask had to be applied to the images to clean and remove their backgrounds. Thus, after applying the mask, the image outputs (Image e) were free of any unnecessary blood component or cell. Finally, by mapping this output to the original images, segmented images of the original images were extracted in order to extract ROI areas.

2.4. Data rescaling, augmentation, and categorization

To present the images in the same format, the corresponding values of the images had to be displayed in a single interval. Pixel intensity normalization in the $[0, 1]$ interval was used to homogenize the input data at this phase. The relevance of data augmentation to the ML-based model's performance cannot be underestimated. It entails modifying pictures to increase the variance of training samples while preserving semantic and intrinsic information. Changes are made to the images that represent the noise of real-world data by employing two transformations

Table 2

Division of dataset samples into training, test, and validation sets.

	Training dataset	Testing dataset	Validation dataset	Data
Percentage (%)	85	5	10	100
No. of paired samples	5250	325	584	6159

in training samples. These transformations include vertical flip and horizontal flip. To train the pre-trained networks and the proposed model, the data had to be categorized. After data segmentation, data rescaling, and categorization, the research data were chosen based on Flowchart 1 and underwent splitting (Table 2).

2.5. Convolutional neural network structure

As the goal was to use a well-turning model in low-specificity equipment, having a lighter weight was one of the main features of efficient models. We employed three lightweight models so that, by assessing their efficiency and based on their lightness, one of these pre-trained networks could be selected as the feature extraction block. The networks selected in this study included *EfficientNetB0*, *MobileNetV2* and *Nanette Mobile*.

2.5.1. EfficientNet

As one of the most up-to-date, cost-efficient, and robust models developed by scaling three parameters of depth, width, and resolution, the *EfficientNetB0* is used in the present study. Its baseline model is our entry point, which takes in an input image using $224 \times 224 \times 3$ dimensions. The model then extracts features throughout the layers by using multiple convolutional layers using a 3×3 receptive field and the mobile inverted bottleneck Conv (MBConv). We chose to employ *EfficientNetB0* due to its balanced depth, width, and resolution that can produce a scalable yet accurate and easily deployable model. Compared to other DCNNs, *EfficientNetB0* scales each dimension using a fixed set of scaling coefficients. This approach surpassed other state-of-the-art models trained on the ImageNet dataset. Even with transfer learning, *EfficientNet* still achieved exceptional results, indicating its effectiveness beyond the usual ImageNet dataset. In its released version, the model had scales of 0–7 (*EfficientNet B0* to *B7*), showing an increase in

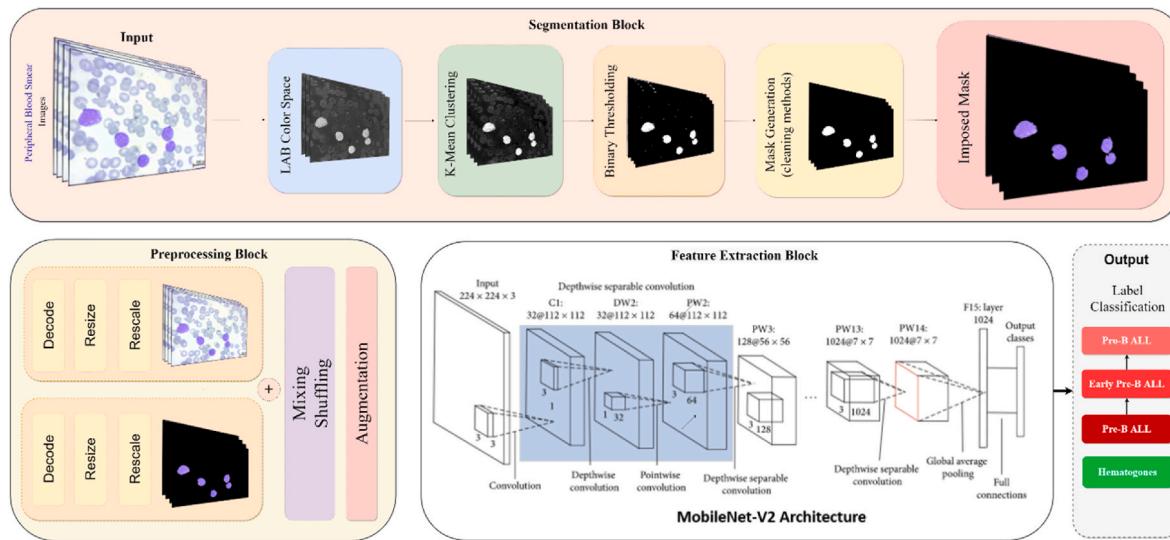


Fig. 2. The suggested model's schematic.

parameter size and accuracy. With the recent EfficientNet, users and developers can access and provide improved ubiquitous computing imbued with DL capabilities on several platforms for various needs [11].

2.5.2. MobileNetV2 and NASNet mobile

The MobileNet model is developed for efficiency and designed for possibly running in embedded devices or mobile devices. The depthwise separable convolution architecture is the primary layer of this model to reduce the number of features. Its DW separable convolution has two layers: DW convolution and point convolution. With separable convolution and DW convolution, MobileNet substantially reduced the parameters of convolutional layers while maintaining good classification performance. In MobileNetV2, residual bottleneck blocks were also proposed to reuse the feature maps. Therefore, MobileNetV2 can achieve proper classification performance with much less training time. We used both pre-trained networks in the current study. MobileNetV2 is quite similar to the original MobileNet, except that it uses inverted residual blocks with bottle necking features. It has a drastically lower parameter count than the original MobileNet. MobileNets support any input size greater than 32×32 , with larger image sizes offering superior performance. One of the main differences between the two versions lies in the convolution layers. The image below depicts convolution blocks for both versions.

NASNet is one of the main scalable CNN architectures (constructed by neural architecture search) that consists of basic building blocks (cells) optimized using reinforcement learning. A cell consists of only a few operations (several separable convolutions and pooling) and is repeated multiple times according to the required network capacity. The mobile version (NASNet-Mobile) consists of 12 cells with 5.3 million parameters and 564 million multiple accumulates (MACs). MobileNetV2 and NASNet-Mobile are two lightweight CNN architectures.

2.6. Selective structure, train and evaluation

We opted to employ segmented images to sensitize the network to lymphoblast cells for boosting accuracy and reliability after constructing and analyzing alternative fitting methods and designs. To select the proposed model, selected pre-trained models were tested after dataset pre-processing and segmentation. After calculating the performance indices, the results of the pre-trained models were not acceptable. Therefore, after discussions and consultations with hematologists, we concluded that the weakness in the results could be due to the incomplete extraction of image features.

Finally, the model was designed such that a combination of segmented images and original images would be fed into the networks. Herein, to extract maximum feature space, an original image channel and a segmented image channel were fed to the selected models together. An advantage of this method (receiving images from two channels) is that almost all the features of research images are covered. The general framework of the suggested technique is depicted in Fig. 2.

2.7. Metric evaluation

In this study the suggested detection and classification model was assessed using clinically relevant statistical metrics. Specificity, recall, and accuracy are suggested metrics. The following parameters were required to calculate these evaluation metrics: True Positive (TP): the number of correct predictions of positive cases by the method; True Negative (TN): the number of correct predictions of negative cases; False Positive (FP): the number of incorrect predictions of positive cases; and False Negative (FN): the number of incorrect predictions of negative cases by the method. However, in the case of an asymmetrical dataset, accuracy is not necessarily.

The best statistic to use when evaluating model performance is accuracy, but in this research, the accuracy metrics were used to select the efficient model through the selected pre-trained networks. The following are some of these metrics in a nutshell:

Accuracy is a parameter that evaluates the capability of a method by measuring the ratio of correctly predicted cases out of the total number of cases. It may be stated mathematically as [14]:

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

Recall: The proportion of accurately anticipated positive observations to all of the actual class observations [14].

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

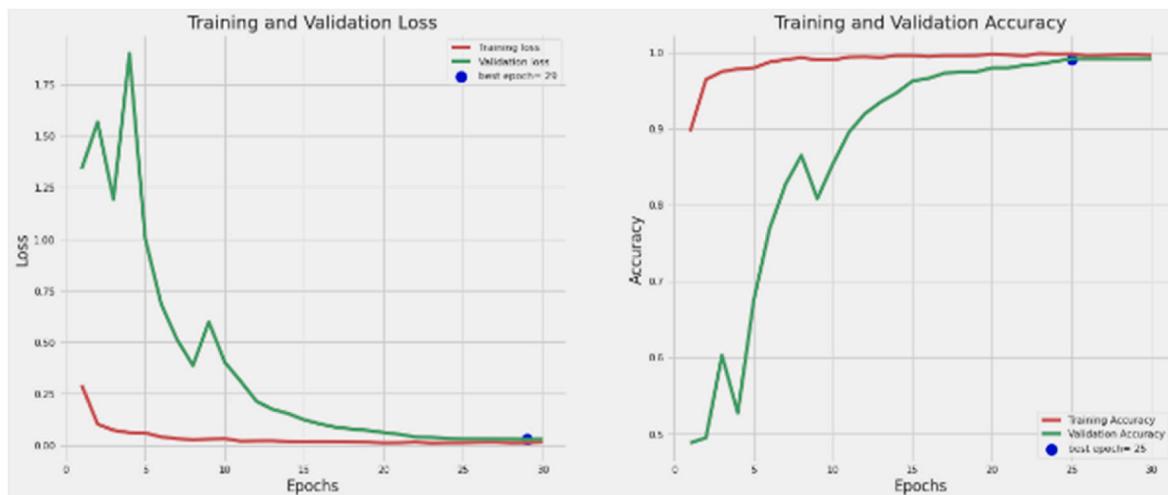
Specificity: the ratio of correctly detected negative observations to all the negative cases [14]. This metric represents the correctly classified rate of opposite disease classes.

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

Table 3

Performance of selected network in unique situation.

Feature extractor	Epoch	Training Accuracy	Batch Size	Size (MB)	Depth	Parameter	Test Accuracy
EfficientNetB0 [11]	30	99.1	32	29	132	5.3 million	100
MobileNetV2 [12]	30	99.62	32	14	88	3,538,984	100
NASNet Mobile [13]	30	86.3	32	23	771	5,326,716	94.1

**Fig. 3.** The suggested model's training and validation accuracy, as well as its loss during training. After 25 epochs, the model converged with training accuracy and loss of 100% and 0.0035, respectively.

3. Results

In this section, we first compare the efficiency of the pre-trained networks on the training and test datasets to select the most efficient network as the chosen network. Following hyperparameter tuning, the network's efficiency is assessed with the test dataset. Finally, based on the chosen model, a mobile application is designed and tested with real data in a laboratory environment.

3.1. Efficient model selection

Following training with the training dataset and testing on the test dataset after 30 epochs, the efficiency of pre-trained networks was obtained under the same conditions. The Table 3 displays the evaluation metrics on the training and test datasets for the three pre-trained models so that, by comparing these indices, the most efficient network for the data of this research can be selected.

The goal was to select a model for use in mobile devices; thus, we were looking for the most efficient network with the smallest size, depth, and learnable parameters. These indicators were chosen in order to select the lightest network in terms of processing with the highest efficiency to be installed on mobile phones. Given that the MobileNetV2 network achieved the highest efficiency and had the minimum size, layer depth, and parameters, it was selected as the chosen network of the proposed model.

3.2. Detection and classification performance of the selected model

In the training phase, the learning curve of the selected model with the tuning parameter for training and validation sets is illustrated in Fig. 3. The proposed model's behavior in dealing with new validation data demonstrated that with increased epochs, the model had reduced error and thus enhanced accuracy for the unknown data, which suggests its great potential in detecting new cases of blood microscopic images.

By receiving the original and segmented images from the input, the proposed model achieved maximum coverage of the features and their

extraction, such that all items of the test dataset were accurately classified. After achieving this model based on the pre-trained network MobileNetV2, an application was designed to utilize this network on laboratory devices.

3.3. Mobile application based on the proposed model in detection and diagnosis B-ALL

In accordance with the objectives of this research, the mobile application should have maximum usability for users. The application input should retrieve images from the camera and the device memory so that, by using the data preparation and data segmentation block, the input images can be segmented and then prepared to be fed to the proposed model. Fig. 4 shows a schematic of the mobile application interface.

This application then receives the original images from the mobile phone input, and after extracting their features, classifies them to determine the B-ALL class. The mobile performance is schematically depicted below. The image output is displayed after classifying the image features in the data classes. In this application, by embedding the SoftMax function and determining the probability of the model output in each class, the end-user can view the probability of the output belonging to each data class. We used the test dataset to test the model and assess the efficiency of the proposed application. By testing all the images in the test dataset, the model managed to classify all the items accurately. Table 4 shows the suggested mobile application's best performance in detecting and classifying each class.

Fig. 5 illustrates the confusion matrix of evaluation on the test set for four classes. The evaluation criteria based on the confusion matrix are given in Table 4.

After being evaluated on the test dataset, the suggested model produced outstanding results. The model was put to the test with a series of photographs, and it correctly identified the study data. The suggested model's efficiency assessment findings on 9 random samples of the test dataset are shown in Fig. 6. I stand for the image indices, P for the model's projected class, and L for the ground truth label. Because of the

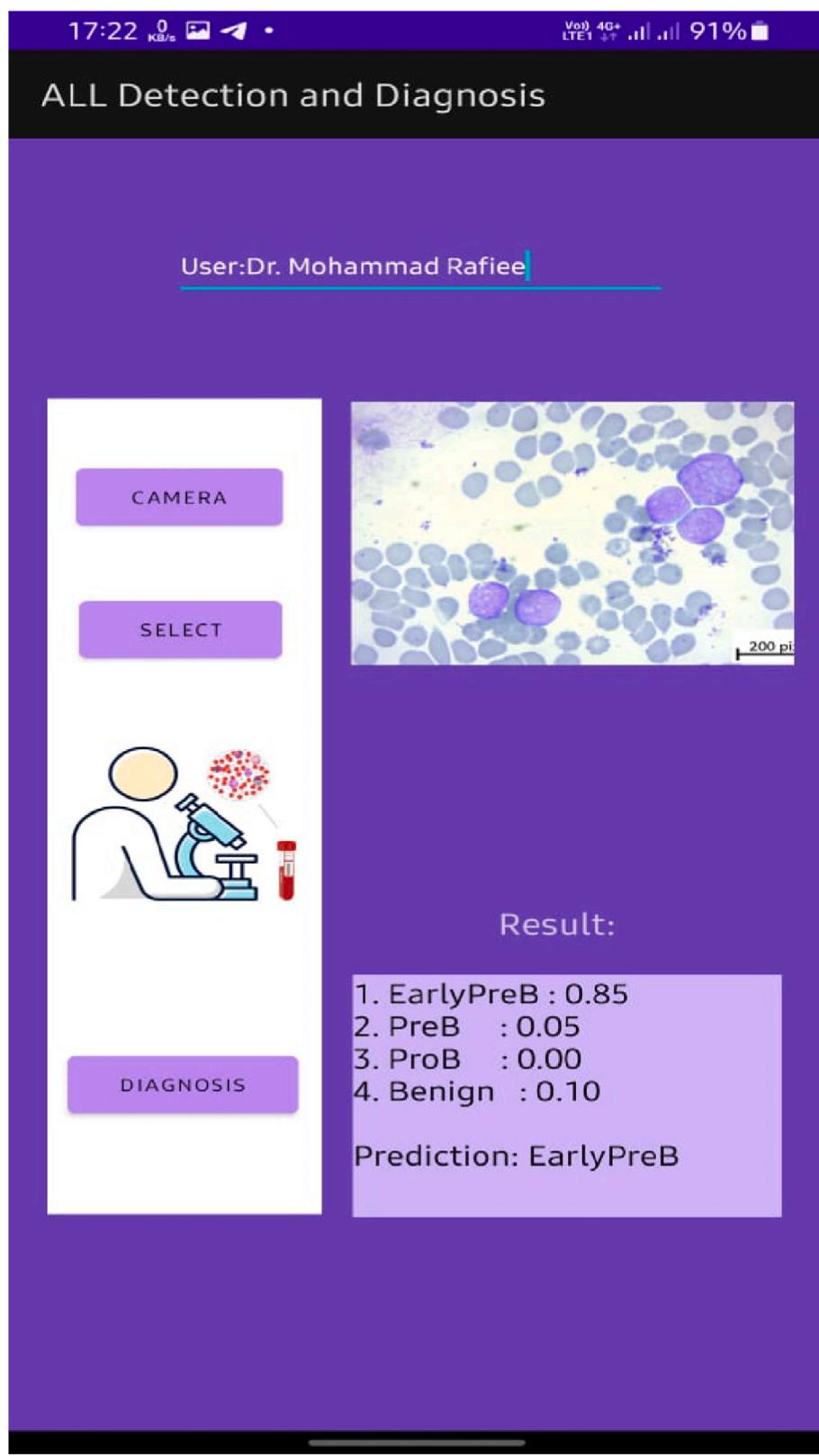


Fig. 4. The proposed Mobile application for the detection and classification of B-ALL.

model's high accuracy, there were no failed sample detection attempts to analyze the model's potential weaknesses.

4. Discussion

In this research, by designing and implementing a model based on the CNN algorithm and transfer learning concepts, innovations were made in the field of its better design. Among these innovations, it is

Table 4
Proposed mobile application evaluation criteria on the test dataset.

Metric	Value (%)
Sensitivity	100
Specificity	100
Precision	100

possible to present an image segmentation pipeline with the best performance by mixing segmented images and original images together and adding this combination to the model, comparing state of the art models, and presenting one of the most optimal structures.

By comparing the performance of the proposed model with previous models on other different datasets, it can be concluded that the performance of the model presented in this research has higher efficiency. Perhaps the proposed model's performance is due to several reasons. It is believed that the first reason was the extraction of the maximum features from two image channels, which have been able to cover all the features of the images. On the other hand, by comparing the models, the most efficient model has been selected, and the selection of this efficient pre-trained model has led to the maximum extraction of image features and optimal classification of these features. The main model based on the lightweight MobileNet V2 model pointed out that finally, to transfer the technology from the theoretical environment to the real environment of clinical laboratories, with the advice of oncology and hematology experts, the proposed model was transferred to the all-in-one system, and then a mobile application was designed, which is in the computing core. This software allows users to enter microscopic images of blood into the model with their mobile phone camera or from the device's memory and view the results. This software was tested by hematologists and achieved acceptable performance in case screening. In the design of the model, by changing each hyperparameter, it has been tried to get the model running optimally. Also, to avoid overfitting in the present model, he pointed out two techniques of data augmentation and added the dropout option to the model. Regardless of the high accuracy of the proposed model due to the sensitive segmentation techniques, the extraction of the maximum number of features, and the optimal topology of the classification block of these features, it is likely that the high percentage of the classification performance metrics of this data set is due to the small number of test datasets, and it is believed that if the number of test datasets increases, the performance of these metrics will decrease slightly.

Almost all previous studies discussed the models theoretically and used public data sets. But this study was an application for clinical

laboratory devices designed, used, and tested in the local laboratory of Tehran Hospital (Iran). This tool helped to reduce costs and complications caused by bone marrow aspiration. The proposed program performed real-time screening of blood microscopic images, thereby reducing diagnosis time and hospital resource consumption. As a result, only cases with a definitive diagnosis of B-ALL were referred to the surgery department for a bone marrow biopsy, and such a biopsy was avoided in unnecessary cases. These AI-based models can be used on portable devices such as smartphones and PDAs to assist clinicians by reducing healthcare resource utilization and diagnostic costs. However, models and programs based on the DL algorithm are merely tools to assist clinical professionals and do not replace them.

Ethics approval

This article, was approved by Iran National Committee for Ethics in Biomedical Research with Approval ID IR.SBMU.RETECH. REC.1400.591.

Availability of data and material

Implementation codes and the pretrained model are available on <https://github.com/AMirEshraghi/Lightweight-Deep-CNN-Based-Mobile-App-in-the-Screening-of-ALL>.

And the data is available at: <https://www.kaggle.com/datasets/mohammadmahreshraghi/blood-cell-cancer-all-4class>.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

Not applicable.

Ethical approval and consent

Not applicable.

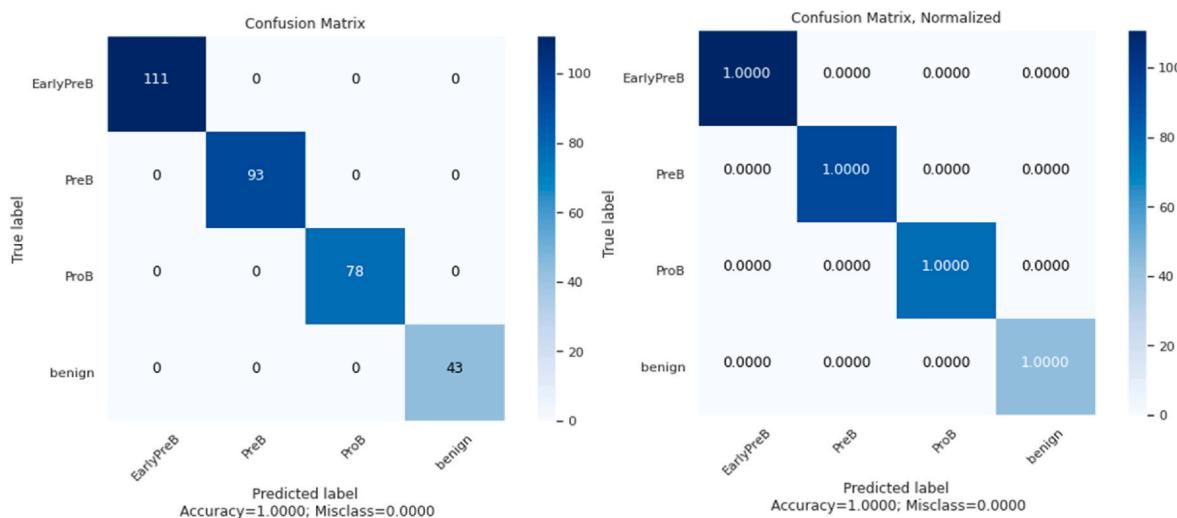


Fig. 5. a) Confusion matrix and b) Normalized confusion matrix in the evaluation of the proposed model's.

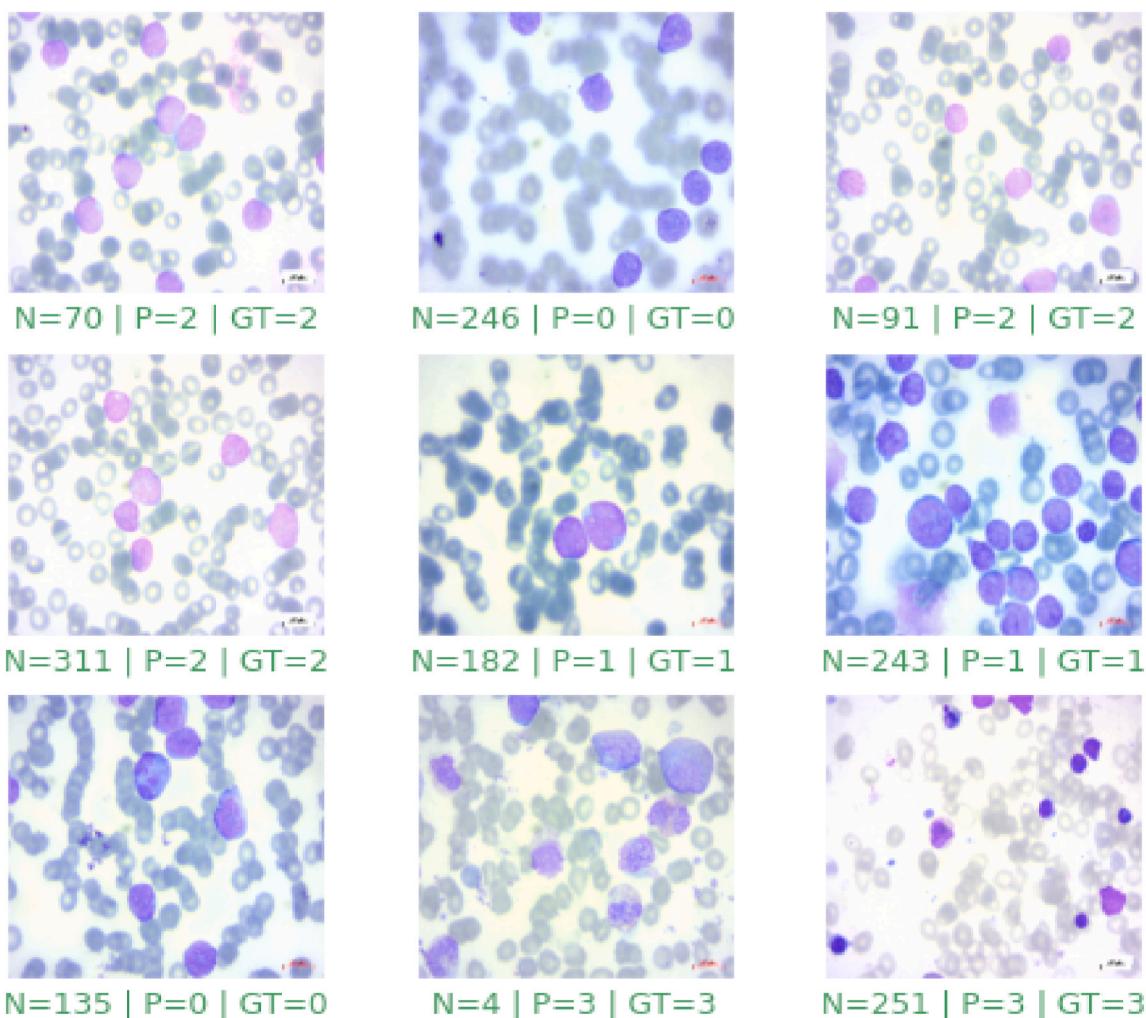
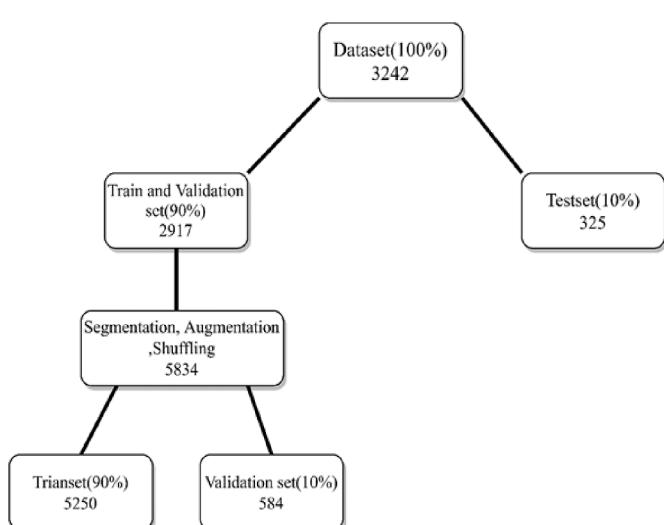


Fig. 6. On 9 Blood Microscopic random samples, the model's performance was evaluated. Image index, model prediction, and groundtruth label are represented by N, P, and GT, respectively. The green caption denotes a correct forecast, whereas the red caption denotes an incorrect prediction. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



Flowchart 1. Overall view of splitting data.

Human and animal ethics

Not applicable.

Authors' detail

M.E and M.G implemented the code and developed the proposed model. They started the idea of the project. A.H was the leader of the project. M.R was the clinical assistance professor as a clinical adviser. H. S and Z.H gathered the data and cleaned it under clinical adviser.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was part of a research project conducted at Shahid Beheshti University of Medical Sciences, Tehran, Iran.

References

- [1] Ghaderzadeh M, Asadi F, Hosseini A, Bashash D, Abolghasemi H, Roshanpour A. Machine learning in detection and classification of leukemia using smear blood images: a systematic review. *Sci Program* 2021;2021:9933481. <https://doi.org/10.1155/2021/9933481>.
- [2] Trewitt KG. Bone marrow aspiration and biopsy: collection and interpretation. *Oncol Nurs Forum* 2001;28.
- [3] Dasariraju S, Huo M, McCalla S. Detection and classification of immature leukocytes for diagnosis of acute myeloid leukemia using random forest algorithm. 2020.
- [4] Agaian S, Madhukar M, Chronopoulos AT. Automated screening system for acute myelogenous leukemia detection in blood microscopic images. *IEEE Syst J* 2014;8: 995–1004.
- [5] Haworth C, Heppleston AD, Jones PHM, Campbell RH, Evans DI, Palmer MK. Routine bone marrow examination in the management of acute lymphoblastic leukaemia of childhood. *J Clin Pathol* 1981;34:483–5.
- [6] Loken MR, Chu S, Fritschle W, Kalnoski M, Wells DA. Normalization of bone marrow aspirates for hemodilution in flow cytometric analyses. *Cytometry Part B Clin Cytom J Int Soc Anal Cytol* 2009;76:27–36.
- [7] Ghaderzadeh M, Aria M, Hosseini A, Asadi F, Bashash D, Abolghasemi H. A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images. *Int. J. Intell. Syst.* 2022;37(8):5113–33.
- [8] Hunter RS. Photoelectric color difference meter. *Josa* 1958;48:985–95.
- [9] Sewall R. Accuracy precision and stability of new photoelectric color-difference meter. *Proc 33rd Annu Meet Opt Soc Am JOSA* 1948;38:651.
- [10] Hunter R. Accuracy, precision, and stability of new photoelectric color-difference meter. *J Opt Soc Am* 1948;38:1094. AMER INST PHYSICS CIRCULATION FULFILLMENT DIV, 500 SUNNYSIDE BLVD, WOODBURY.
- [11] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. *Int Conf Mach Learn* 2019:6105–14. PMLR.
- [12] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. Mobilenetv2: inverted residuals and linear bottlenecks. *Proc IEEE Conf Comput Vis pattern Recognit* 2018:4510–20.
- [13] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. *Proc IEEE Conf Comput Vis pattern Recognit* 2018: 8697–710.
- [14] Ghaderzadeh M, Asadi F, Jafari R, Bashash D, Abolghasemi H, Aria M. Deep convolutional neural network-based computer-aided detection system for COVID-19 using multiple lung scans: design and implementation study. *J Med Internet Res* 2021;23:e27468. <https://doi.org/10.2196/27468>.