LeukoSegmenter: A Double Encoder-decoder Based Network for Leukocyte Segmentation From Blood Smear Images

Sabrina Dhalla (✉ dhallasabrina@gmail.com)

Panjab University Faculty of Engineering and Technology <https://orcid.org/0000-0002-5820-3666>

Ajay Mittal

Panjab University Faculty of Engineering and Technology

Savita Gupta

Panjab University Faculty of Engineering and Technology

Research Article

Keywords: LeukoSegmenter, double encoder-decoder, leukocyte segmentation, blood smear images, blood cells

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-997876/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

LeukoSegmenter: A double encoder-decoder based network for leukocyte segmentation from blood smear images

Sabrina Dhalla · Ajay Mittal · Savita Gupta

the date of receipt and acceptance should be inserted later

Abstract Segmentation of blood cells is a prerequisite step in automated morphological analysis of blood smear images, cell count determination, and diagnosis of various diseases such as leukemia. It is extremely challenging due to the different sizes, shapes, morphological characteristics, and overlapping of blood cells. Due to its complicated nature, it is generally performed as a sequence of steps. However, sequential segmentation results in restricted accuracy due to cascading of errors that creep during each stage. On the contrary, pixel-wise segmentation of blood cells is a single-step task and gives promising results. In this paper, we propose LeukoSegmenter, a double encoder-decoder for precise pixel-wise segmentation of leukocytes from blood smear images. It uses pre-trained ResNet18 based encoders and U-Net-based decoders. Feature maps obtained from the first network are utilized as attention maps. These are used as input in conjunction with the original 3-channel image to obtain the final mask from the second network. This mechanism allows the latter encoder-decoder pair to focus explicitly on leukocytes and ignore other blood cells and debris, thus improving the segmentation accuracy. Experiments on ALL-IDB1 dataset show that the proposed LeukoSegmenter achieves an intersection-over-union score of 94.6827% and a Dice score of 97.1987% which is superior to that of state-of-the-art methods.

1 Introduction

Leukemia, a type of blood cancer, is a deadly disease that causes thousands of fatalities every year globally. More than sixty thousand cases were registered in United States (US) alone in the year 2019 [1]. In the same year, Global Cancer Observatory positioned it among the top 10 deadly cancers in India [2]. In general, human peripheral blood consists of three types of cells namely red blood cells (RBCs), white blood cells (WBCs), and platelets (Fig. 1). Out of these,

UIET, Panjab University, Chandigarh

Fig. 1 Types of blood cells

WBCs are more prone to become cancerous. In leukemia, blood-forming tissues start producing a large number of abnormal WBCs, slowly killing the normal cells in the bloodstream. Depending on the type of WBCs affected, leukemia is classified into two types as lymphoblastic (affects lymphocytes) and myeloid (affects monocytes and granulocytes). These can be further categorized on the basis of the rate of progress of leukemia as acute (i.e., progression rate is fast) and chronic (i.e., progression rate is slow).

Technological advancements during the past five decades have introduced various semi-automated methods in the field of hematology such as flow cytometry, blood cell count using a hemocytometer and blood smear image processing. Flow cytometry and hemocytometers are used for differential cell count and cannot be used for detailed morphological analysis of the blood cells. On the contrary, the image processing-based systems use pattern recognition algorithms to analyze the geometry, size, color, and texture of blood cells, much like a human expert. The early image processing-based systems capture various hematological features and apply different rules to quantify cancerous characteristics. These rule-based systems are ad hoc, fragile and do not fit into one generic model. Furthermore, various slide preparation issues like variation in optical density, overlapping cells, disrupted cells, stain debris, stain variations and image acquisition issues like lighting, scale, noise, compression cause significant variance in their results.

Advancements in machine learning, specifically deep learning, have led to the development of methods that have given benchmark performances in various computer vision tasks [3–6]. Though machine learning algorithms are based on an explicit selection of feature sets, deep learning algorithms automate this process [7]. Automated feature extraction allows an extremely large number of features to be extracted, beyond human comprehension. In a deep neural

network (DNN), input passes through a sequence of stacked layers. Each layer utilizes the low-level features selected from the preceding layer(s) and passes high-level features to the succeeding layer(s). This process continues till the final layer of the network is reached which gives the comprehensive representation of the data. The depth of a network or the number of stacked layers in a network significantly impacts the number of parameters and the performance of the network. Fine-tuning a large number of trainable parameters through training requires enormous computational power, generally met using graphics processing units (GPUs). To reduce DNN's computational requirement, a suggested way is to preprocess the input image, extract region-of-interest (RoI) and then pass it to the network. For leukemia, WBCs from blood smear images are segmented, stacked and then passed to a DNN-based computer-aided diagnosis (CADx) system as a block for analysis. This would help the network to focus on smaller ROIs, and perform faster and accurate analysis.

Biomedical image segmentation using deep learning is generally performed using convolutional neural networks (CNNs), the networks best at processing spatial data. During the recent years, CNNs such as fully convolutional network (FCN) [5], and encoder-decoder networks such as SegNet [8], UNet [9] and others [10–12] have been used to perform biomedical segmentation with good results. The encoder-decoder networks generally use models such as AlexNet [3], VGG16 [4] and ResNet [13], pre-trained on natural images dataset such as ImageNet [14], as encoder. The encoder is then followed by a decoder which projects the low-level features extracted by the encoder to a high-resolution pixel space such that each pixel is classified into respective classes. In this paper, we propose LeukoSegmenter, a double encoder-decoder network to segment leukocytes from blood smear images. The contributions of this paper are as follows:

1. The proposed model cascades encoder-decoder twice with the concatenation of feature maps in between them. The concatenated input allows the latter encoder-decoder to explicitly focus on leukocytes while ignoring other blood cells and cell debris, thereby giving better-segmented results.
2. The model has a symmetrical structure and is based on ResNet18 and UNet architectures. It leverages the inherent advantages of ResNet18 architecture for feature extraction and UNet architecture for hierarchical upscaling to recreate the segmented image. The usage of pretrained networks also resulted in faster convergence of the proposed model.
3. We present a comprehensive empirical study including an ablation study validating the need of cascading two encoder-decoder networks. The qualitative and quantitative results from the ablation study indicate that the proposed dual ResNet18-UNet encoder-decoder network gives better segmentation results in terms of IoU and Dice score on ALL-IDB1 dataset as compared to single ResNet18-UNet network, single ResNet34-UNet++ network [15], and double ResNet34-UNet++ network.
4. The proposed LeukoSegmenter gives consistent and reproducible results when evaluated on the standard dataset using the benchmark metrics.

Fig. 2 Sample images and their masks (a-d) and (A-D) for ALL Patients, (e-h) and (E-H) for Healthy Individuals

The rest of the paper is organized as follows. Section 2 presents the work related to leukocyte segmentation from blood smear images. The architecture of LeukoSegmenter, dataset details, and the evaluation metrics used to evaluate the performance of the proposed model are presented in Section 3. Section 4 presents an ablation study, the results of the proposed model, and their comparison with that of state-of-the-art methods. Finally, conclusions are drawn in Section 5.

2 Related Work

The existing methods for leukocyte segmentation from a blood smear image can be broadly classified as traditional methods and deep learning-based methods.

2.1 Traditional methods

The traditional methods for leukocyte segmentation are further categorized as:

1. *Rule-based methods:* These methods use heuristic rules formulated on the basis of prior knowledge of cells such as their shape, texture and size to perform segmentation. The methods have the advantage of being simple and the rules can be applied in a different order, giving notable freedom. Rawat et al. [16] used global thresholding to separate the nucleus and cytoplasm of each cell to segment leukocytes. Other thresholding algorithms such as Otsu [17–20] and color-based algorithms [21] have also been used for leukocyte segmentation from blood smear images. Harun et al. [22] exploited the neighboring pixels in a connected region using the seeded region growing area extraction (SRGAE) method on customized dataset. The segmented cells with more than 1000 pixels are treated as blast cells and are retained

in the segmented result. Edges also play a vital role in segmenting WBCs [23, 24] but the segmentation results from the edge detection-based methods generally need post-processing using morphological operations such as hole filling, erosion and dilation. Another popular sub-category of rule-based methods is based on clustering. This method finds similarities in pixels based on some parameters and groups them into various clusters. In [25, 26], authors used centroid-based clustering for leukocyte segmentation. The rule-based methods have a limitation in that they are fragile, and their performance varies with the complexity of the input image.

2. *Deformable model-based methods*: These methods use flexible 2D or 3D curves that iteratively evolve under the influence of internal and external forces, and user-defined constraints to segmented leukocytes. Sadeghian et al. [27] used active contour models (ACM), also known as snakes, to separate the nucleus and cytoplasm of each cell. While the segmentation accuracy for the nucleus was 92%, it dropped to 70% for cytoplasm segmentation. Authors in [28–30] used level-set method for nuclei and cytoplasm detection. The method needs to be used in conjunction or be followed by other algorithms such as active contours and edge detection. The deformable model-based methods have limitations that their convergence is expensive and is highly dependent upon the initialization of the curves.

2.2 Deep learning-based methods

Deep learning-based methods have outperformed state-of-the-art methods in several computer vision tasks including leukocyte segmentation. Three main DNN-architectures that have been used for segmentation of leukocytes from blood smear images are:

1. *SegNet*: Introduced by Badrinarayanan et al. [8], in the year 2017, SegNet architecture consists of an encoder network and a corresponding decoder network, followed by a pixel-wise classification layer. Encoder network, which is similar to VGG16 architecture, is used for feature extraction whereas the decoder network upsamples the resultant feature map using max-pool indices and the convolution operation. Trans et al. [31] used SegNet to segment WBCs from 42 images of ALL-IDB database. Using this architecture, an overall IoU score of 89.96% is obtained.
2. *Fully-convolutional networks (FCNs)*: These networks are modified CNNs [5], in which the last fully connected layers are removed and the output map is passed to the decoder network. Skip connections are embedded from the corresponding encoders to the decoders. In contrast to SegNet, FCNs use deconvolution operation to upsample the feature maps. Depending on the number of times the pooling operation is performed, they are further categorized as FCN8, FCN16 and FCN32. Shahzad et al. [32] used a variant of FCN where VGG16 was used as a feature extractor model to segment RBCs, WBCs and platelets from ALL-IDB blood slide images. Mean ac-

curacy of 91.96% is achieved for all the three cellular components of blood, which leaves scope for improvement in the future.

3. *U-Net*: This network is the most popular form of encoder-decoder networks that is used for semantic segmentation of medical images [9, 33]. Its network architecture is similar to SegNet and FCNs where the encoder downsamples the input image and the decoder performs up-sampling, thus taking shape of the letter ‘U’. The major difference between FCNs and UNet is that the former uses an addition operation between downsampled and upsampled feature maps on the same level whereas the latter uses a concatenation operation for the same. Thus, UNet prevents loss of any information and is specifically designed for medical images. Lu et al. [15] used a combination of UNet++, a variant of UNet with multi-skip connections, and ResNet50 to segment WBCs from the background using four types of datasets. The mean IoU value obtained for all the datasets is above 90%. However, the major drawback of the model is that it has not been tested on images having multiple WBCs.
4. *DeepLab*: This model has been designed in the year 2017 for semantic segmentation and consists of encoder-decoder phases [34]. It aims to overcome the major drawback of all the architectures defined above which is the incapability to handle varied sized inputs. This network can be trained on various sized images using Spatial Pyramid Pooling (SPP). To reduce the computational complexities and cost involved in this model, dilated convolutions are used. These types of convolutions introduce space in between kernel values so that field of view is increased and the number of parameters is not humongous. Roy et al. in 2021 trained blood cell images on DeepLabv3+ model (a variant of DeepLab which uses depthwise separable convolutions in both encoder and decoder phase) which used ResNet-50 to downsample the images [35]. The average IoU on three datasets obtained was 92.1%. Although the results are satisfactory it needs to be tested with images containing multiple overlapped cells too.

3 Materials and Methods

3.1 Dataset

In this work, Acute Lymphoblastic Leukemia Image Database (ALL-IDB1) [36] has been used for the research experiments. It consists of total of 108 blood slide images, out of which 49 images have been taken from patients suffering from ALL and 59 from healthy persons. Images have been captured by using Canon PowerShot G5 camera, have 24-bit color depth and are available in .jpg format. While the blast cells’ images have a resolution of 2592×1944 , healthy cells’ images have 1712×1368 image quality. Fig. 2 (a-h) shows sample blood slide images of healthy individuals and ALL patients. In our case, the entire dataset has been divided into 80:20 for training and testing purposes. Blast cell images are associated with the .xyc file which contains centroid coordinates

of each cell. Masks for the corresponding cells (WBCs, RBCs and platelets) have been generated and fine-tuned by Shahzad et al. [32]. These masks are available in .jpg format and are used as labels for training and testing semantic segmentation models. Fig. 2 (A-H) shows mask images corresponding to the images of ALL-IDB1 in the same figure. Apart from WBCs, the rest of the area is labeled as background.

3.2 Methodology

During the end-to-end training process, the encoder tries to learn patterns in an image with the help of many convolutional layers, denoted by C^m . These layers capture local information in an image and pass it to the next layers in the form of feature maps, denoted by F^m . Thus, each input x is convolved with the weight matrix, W_i and bias value b_i is added to it. This information is passed through ReLU activation function, denoted by $f(x)$ to induce non-linearity in the deep-supervised network. This process is simplified using equation 1

$$c_i^{(1)} = f\{b_i^{(1)} + W_i^{(1)} \times x\} \quad (1)$$

where $i = 1, \dots, F^1$ and $m=1$. The encoder branch for the network is generally chosen as one of the Convolutional Neural Networks (CNNs) which is pre-trained on ImageNet dataset. This process of transfer learning helps to transfer the basic features learned on non-medical datasets to medical image datasets. Such networks are generally used for classification purposes but can also be used for the segmentation of objects by passing the downsampled image to the corresponding decoder. In this case, the original ResNet18 network [13] has been used as an encoder. However, the original architecture has been modified as the last fully connected and softmax layers have been removed and the feature maps are passed directly to the next module as shown in Table 1. The decoder branch, which forms a directed acyclic graph topography is used to restore and refine the set of semantic features learned from the encoder module. This restoration process involves transposed convolutions (also known as deconvolutions) that helps to upsample the feature map and thus expand feature dimensions. Spatial resolution is upsampled by 2×2 factor which reduces the number of feature maps by half. It is then followed by corresponding 3×3 convolution, batch normalization and ReLU activation function. The inclusion of skip connections from an encoder to decoder helps to retain fine information and enhance decoding performance by avoiding the loss of gradient information. A 1×1 convolution used at the end of the network helps to associate each of the n -component feature maps to its corresponding classes. Thus, relevant features are learned and the decoder outputs a mask corresponding to the input image. In this paper, a network of dual encoder-decoder has been proposed which consists of two sub-networks X_1 and X_2 as shown in Fig. 3. A 3-channel RGB input is passed to the encoder which downsamples the image feature map and increases the count of channels using convolution operation. Now, the spatial resolution of the image has been reduced and only significant

Layer	Operations	Input Channels	Output Channels	Stride	Padding	Kernel Size	Input Size	Output Size
	Conv + BN + ReLU	3	64	2	3	7×7	512×512	256×256
	Max Pool	64	64	2	1	3×3	256×256	128×128
EncoderLayer 1	Basic Block 1	64	64	1	1	3×3	128×128	128×128
	Basic Block 2	64	64	1	1	3×3	128×128	128×128
EncoderLayer 2	Basic Block 1	64	128	2	1	3×3	128×128	64×64
	Down Sample	64	128	2		1×1	128×128	64×64
	Basic Block 2	128	128	1	1	3×3	128×128	64×64
EncoderLayer 3	Basic Block 1	128	256	2	1	3×3	64×64	32×32
	Down Sample	128	256	2		1×1	64×64	32×32
	Basic Block 2	256	256	1	1	3×3	64×64	32×32
EncoderLayer 4	Basic Block 1	256	512	2	1	3×3	32×32	16×16
	Down Sample	256	512	2		1×1	32×32	16×16
	Basic Block 2	512	512	1	1	3×3	32×32	16×16
DecoderLayer 0	Decode Block 1	768	256	1	1	3×3	16×16	32×32
	Decode Block 2	256	256	1	1	3×3	32×32	32×32
DecoderLayer 1	Decode Block 1	384	128	1	1	3×3	32×32	64×64
	Decode Block 2	128	128	1	1	3×3	64×64	64×64
DecoderLayer 2	Decode Block 1	192	64	1	1	3×3	64×64	128×128
	Decode Block 2	64	64	1	1	3×3	128×128	128×128
DecoderLayer 3	Decode Block 1	128	32	1	1	3×3	128×128	256×256
	Decode Block 2	32	32	1	1	3×3	256×256	256×256
DecoderLayer 4	Decode Block 1	32	16	1	1	3×3	256×256	512×512
	Decode Block 2	16	16	1	1	3×3	512×512	512×512
Segmentation Head	Conv	16	1	1	1	3×3	512×512	512×512

Table 1 Architecture for single encoder-decoder. Basic Block 1 represents Conv2D, BN, ReLU, Conv2D, BN and Decode Block represents Conv2D, BN, ReLU in the same order.

features are retained by the network. This feature map is then passed to the decoder which up-scales it and helps to compensate for the loss of spatial resolution that occurred during the encoding phase. The motivation behind the double encoder-decoder model proposed in this paper is a network called W-net [37] which simply stacks two such architectures sequentially in the shape of “W”. However, modification in our model ensures that minute details in an image are preserved. Given x as a 3-channel input image to the first network X_1 , $X_1(x)$ is a single channel output of the first network. This output along with the original 3-channel input is passed to the second network (refer equation (2))

$$X_{\text{final}} = X_2(x, X_1(x)) \quad (2)$$

Here, x and $X_1(x)$ are concatenated and hence, four-channel input is passed to the second network X_2 . In general, attention modules are used by CNNs to target contextual information and discard non-useful features learned by the network. In our case, the output of the first U-net network acts as an attention map to suppress the irrelevant background pixels. Concatenation

Fig. 3 Double encoder-decoder network architecture

of output from the first decoder helps the model focus on fine and significant regions in an image. Although, architectures of both networks can vary we have restricted ourselves to using the same architectures to simplify understanding and reasoning of the novel concept.

3.3 Experimental Setup

In this section, various data augmentations and the experiment setup details have been discussed.

1. *Training Phase:* Since, medical image datasets are small in size, it is advised to perform augmentations to prevent over-fitting of the model [38]. Thus, various types of data augmentations have been applied to increase the number of images in our dataset namely, horizontal flip, vertical flip, and rotations (90° , 180° and 270°). In this way, each image in the dataset has been augmented four times randomly which increased the size of the training set from 88 images to 1408 images.
2. *Testing Phase:* A total of 20 images have been utilized to test the performance and efficiency of the proposed model. To improve the robustness of the model, Test Time Augmentation (TTA) has been adopted. This includes operations equivalent to augmentation during the training phase i.e, horizontal flip, vertical flip and rotations (90° and 180°). The average of the output predictions is concluded as a final result.
3. *Experiment Settings:* The proposed network is trained and tested on Nvidia GeForce RTX 2080Ti. The experiments have been implemented on Windows 10 operating system using the public PyTorch platform. Various

Fig. 4 Sample images from (a) Dataset (b) Ground Truth (c) Proposed double UNet model

hyper-parameters such as batch size, number of epochs, optimizer and scheduler have been fine-tuned to adapt to the network as a whole. Optimizer used during training is Adam (i.e. adaptive moment estimation) and its initial learning rate has been set to 0.001. In addition, CosineAnnealingLR is used as a scheduler. The whole network is trained for 200 epochs using a batch size of 8.

3.4 Mixed Loss Function

The microscopic images consists of various types of cells (WBCs, RBCs and platelets) and background (cytoplasm). The majority area in an image is covered by RBCs and cytoplasm, thus causing a typical problem of class imbalance. This problem can significantly dominate the loss function values while the model is training and can fall for local minima. Thus, a new loss is computed which refined the results and made the model more robust.

1. *Binary Cross-Entropy (BCE) loss*: It compares values of each pixel in the resultant mask images with the original pixel value in the ground truth and is shown in equation 3.

$$L_{BCE}(p, g) = - \sum_i^N [g_i \ln p_i + (1 - g_i) \ln(1 - p_i)] \quad (3)$$

where N corresponds to total number of pixels in an image, p corresponds to the probability that a pixel belongs to the indicated class and g is the

ground truth for the pixel. If value of g_i is 1, it means pixels belong to WBC class and for value of g_i as 0, it corresponds to background class

2. *Dice Coefficient (DC) loss*: Dice-loss is used to handle the problem of class-imbalance by adaptively weighing each class as per number of pixels. It can be found using equation 4.

$$L_{DC}(p, g) = 1 - \frac{2 \times \sum_{i=1}^N p_i g_i + \delta}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2 + \delta} \quad (4)$$

where p and g indicate the same as that in equation (3) and $\delta \in [0,1]$ is a small constant value that is used to prevent divide-by-zero error and helps negative values propagate in the network.

3. *Resultant loss*: Based on the above two losses we propose a hybrid loss function which combines both the losses so that the convergence speed increases and the output results are improved. Thus, each of these is added to form the equation 5.

$$L_{Total} = L_{BCE} + L_{DC} \quad (5)$$

3.5 Evaluation Metrics

1. *Intersection-Over-Union (IoU or Jaccard Index)*: IoU is a very commonly used metric to evaluate the performance of a segmentation algorithm. It is defined as the percentage of the overlapped area between the target and predicted results. Hence, it is used to measure how close the resultant output is with the ground truth and is represented by equation 6.

$$IoU = \frac{1}{2} \left(\frac{|G_b \cap P_b|}{|G_b \cup P_b|} + \frac{|G_f \cap P_f|}{|G_f \cup P_f|} \right) \quad (6)$$

2. *Dice Coefficient (DC or F1 Score)*: It is defined as the total area of overlap of the respective classes divided by total area of both images and is represented by equation 7

$$DiceScore = \frac{2 |G_f \cap P_f|}{|G_f| + |P_f|} \quad (7)$$

For equations 6 and 7, G_f and G_b signify WBC and background region as per labels and P_f and P_b correspond to WBC and background region as per predicted results.

4 Results

This section compares the results of proposed methodology with various other methods.

Fig. 5 (a,c,e) Outputs from first decoder and (b,d,f) Corresponding outputs from second decoder

Fig. 6 Sample images of (a) Ground-truth (b) Single UNet model, (c) Proposed double UNet model (d) Single UNet++ model and (e) Double UNet++ model

4.1 Comparison with state-of-art

To check if our proposed model outperformed the state-of-art methods, a total of 20 images containing both blast and normal cells were used as test

References	Dataset Used	Over-lapping WBCs	Method Used	IoU	Dice Score	Accuracy
Reena et al. [39] 2020	LISC [40] and customised datasets	NO	DeepLabv3+ and AlexNet	84.22%	-	98.42%
Roy et al. [35] 2021	LISC [40] and customised datasets	NO	DeepLabv3+ and ResNet-50	92.1%	-	96.1%
Shahzad et al. [32] 2020	ALL-IDB 2	YES	SegNet and VGG16	40.62%	-	93.34%
Lu et al. [15] 2021	LISC [40] and customised datasets	NO	UNet++ and ResNet34	96.77%	96.54%	-
Tran et al. [31] 2018	ALL-IDB 2	YES	SegNet and VGG16	75.04%	-	94.93%
Proposed Method	ALL-IDB 1	YES	Double encoder-decoder (Resnet18 and U-Net)	94.68%	97.19%	98.24%

Table 2 Comparison of performance of state-of-art methods with the proposed method

images. As shown in Fig. 4, the qualitative similarity between the results of the proposed double encoder-decoder model and ground truth can be verified. This pixel-level segmentation model has an upper hand over classic object-based models too because it preserves the important information at pixel level, leading to much higher performance. Whole-slide images which contain multiple WBCs have been used to frame the solution for real-time problems for leukemia detection which is the major limitation for other state-of-art models. Moreover, it can be seen from Fig. 5 that resultant masks obtain from a single encoder-decoder model also segment unwanted areas (highlighted in red). This limitation is removed by the proposed model which pays attention to only leukocytes and ignores other unwanted areas in the images. In a word, the proposed model can effectively segment leukocytes from images of blood slides. In addition to the qualitative results, quantitative results have also been calculated using standard performance metrics such as IoU, DS and accuracy in Table. 2. The model has attained IoU of 94.68%, DS of 97.19% and accuracy of 98.24%.

4.2 Ablation Study

The performance of the proposed segmentation model was tested using three types of ablation studies. Fair comparison has been done qualitatively (Fig. 6) and quantitatively (Table 3) amongst various models. Fig. 7 shows the progression of L_{Total} over a total number of epochs for all three studies. The first ablation study was conducted to determine whether double or two cascaded UNets were superior to single UNet. In this experiment, a basic encoder (ResNet18) and all the hyper-parameters were kept the same. Although the size and the number of trainable parameters of the model have increased two

Method	Number of Trainable Parameters	Batch Size	Model Size (MB)	Time per Epoch (sec)	IoU	Dice Score	Accuracy
UNet and ResNet18	14,328,209	8	1196.66	207.47	94.06 %	97.12%	97.89%
Double UNet and ResNet18	28,659,554	8	2394.33	254.85	94.68%	97.19%	98.24%
UNet++ and ResNet34	26,078,609	8	3403.48	278.02	93.89 %	96.95 %	97.91%
Double UNet++ and ResNet34	52,160,354	8	6807.98	375.09	94.35%	97.16 %	98.16%

Table 3 Ablation Study

Fig. 7 Comparison of Double UNet with (a) Single UNet, (b) Single UNet++ (c) Double UNet++

folds but significant improvement in the segmentation could be observed. In this case, IoU increased by 0.62%, dice score by 0.07% and accuracy by 0.35%.

As per the results in Table 2, Lu et al. [15] segmented WBCs using UNet++ and ResNet34, and it outperformed our suggested model in terms of IoU score. Hence, the same model was used to cross-verify the results on our dataset too. A score of the various performance metrics implies that our model is more suitable for segmentation task. Under the third ablation study, a double model of UNet++ and ResNet34 was designed and compared with our model. After various experiments, it can be concluded that performance patterns obtained by using this model are quite similar to those obtained from our proposed model. However, this model's size and the number of trainable parameters are high, which may have a detrimental impact on its practical application.

5 Conclusion

In this research, we have proposed an innovative design that involves the usage of a double encoder-decoder network to segment WBCs from microscopic blood smear images. The resultant single channel feature map obtained from

the encoder of the first network is used as an attention map that focuses on the ‘important’ areas in an image. The second decoder in the network generates the final resultant mask which is used for performance evaluation. A fair comparison between various strategies has been done to evaluate the advantage of this novel architecture. This work advocates the use of similar networks at this stage. However, combinations of various novel and pre-trained architectures should be investigated as future work.

6 Acknowledgment

The review paper is supported by University Grants Commission (UGC), New Delhi, India, which provides fellowship to research scholars under the scheme NET-JRF

7 Declarations

Author’s Contribution: Conceptualization: SD; Formal Analysis: AM, SG; Methodology: SD; Resources: SD, AM, SG; Software: SD; Supervision: AM, SG; Visualization: SD; Writing original draft: SD; Writing review and editing: SG, AM.

Conflict of interest: We declare that we do not have any commercial or associative interest, that represents a conflict of interest in connection with the work submitted.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. National Cancer Institute. <https://seer.cancer.gov/statfacts/html/leuks.html>, 2018.
2. The Global Cancer Observatory. <http://gco.iarc.fr/today/data/factsheets/populations/356-india-fact-sheets.pdf>, 2018.
3. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
4. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
5. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
6. Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and

- detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.
7. B Schmauch, P Herent, P Jehanno, O Dehaene, C Saillard, Christophe Aubé, Alain Luciani, N Lassau, and S Jégou. Diagnosis of focal liver lesions from ultrasound using deep learning. *Diagnostic and interventional imaging*, 100(4):227–233, 2019.
8. Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
9. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
10. Sihang Zhou, Dong Nie, Ehsan Adeli, Jianping Yin, Jun Lian, and Ding-gang Shen. High-resolution encoder-decoder networks for low-contrast medical image segmentation. *IEEE Transactions on Image Processing*, 29:461–475, 2019.
11. Jung Uk Kim, Hak Gu Kim, and Yong Man Ro. Iterative deep convolutional encoder-decoder network for medical image segmentation. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 685–688. IEEE, 2017.
12. Abbas Khan, Hyongsuk Kim, and Leon Chua. Pmed-net: Pyramid based multi-scale encoder-decoder network for medical image segmentation. *IEEE Access*, 9:55988–55998, 2021.
13. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
14. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
15. Yan Lu, Xuejun Qin, Haoyi Fan, Taotao Lai, and Zuoyong Li. Wbc-net: A white blood cell segmentation network based on unet++ and resnet. *Applied Soft Computing*, 101:107006, 2021.
16. Jyoti Rawat, Annapurna Singh, HS Bhadauria, Jitendra Virmani, and JS Devgun. Classification of acute lymphoblastic leukaemia using hybrid hierarchical classifiers. *Multimedia Tools and Applications*, 76(18):19057–19085, 2017.
17. Niranjana Chatap and Sini Shibu. Analysis of blood samples for counting leukemia cells using support vector machine and nearest neighbour. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 16(5):79–87, 2014.
18. Subhash Rajpurohit, Sanket Patil, Nitu Choudhary, Shreya Gavasane, and Pranali Kosamkar. Identification of acute lymphoblastic leukemia in microscopic blood image using image processing and machine learning algorithms. In *2018 International Conference on Advances in Comput-*

- ing, *Communications and Informatics (ICACCI)*, pages 2359–2363. IEEE, 2018.
19. Roopa B Hegde, Keerthana Prasad, Harishchandra Hebbar, and Brij Mohan Kumar Singh. Image processing approach for detection of leukocytes in peripheral blood smears. *Journal of medical systems*, 43(5):1–11, 2019.
20. Emad A Mohammed, Mostaja MA Mohamed, Christopher Naugler, and Behrouz H Far. Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding. In *2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–5. IEEE, 2013.
21. Leow Bin Toh, MY Mashor, P Ehkan, H Rosline, AK Junoh, and Nor Hazlyna Harun. Image segmentation for acute leukemia cells using color thresholding and median filter. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10(1-5):69–74, 2018.
22. Nor Hazlyna Harun, AS Abdul Nasir, Mohd Yusoff Mashor, and Rosline Hassan. Unsupervised segmentation technique for acute leukemia cells using clustering algorithms. *World Academy of Science, Engineering and Technology International Journal of Computer, Control, Quantum and Information Engineering*, 9:253–59, 2015.
23. Subrajeet Mohapatra, Dipti Patra, and Sanghamitra Satpathi. Image analysis of blood microscopic images for acute leukemia detection. In *2010 International Conference on Industrial Electronics, Control and Robotics*, pages 215–219. IEEE, 2010.
24. Subrajeet Mohapatra, Sushanta Shekhar Samanta, Dipti Patra, and Sanghamitra Satpathi. Fuzzy based blood image segmentation for automated leukemia detection. In *2011 International Conference on Devices and Communications (ICDeCom)*, pages 1–5. IEEE, 2011.
25. Sos Agaian, Monica Madhukar, and Anthony T Chronopoulos. Automated screening system for acute myelogenous leukemia detection in blood microscopic images. *IEEE Systems journal*, 8(3):995–1004, 2014.
26. Preetham Kumar and Shazad Maneck Udwadia. Automatic detection of acute myeloid leukemia from microscopic blood smear image. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1803–1807. IEEE, 2017.
27. Farnoosh Sadeghian, Zainina Seman, Abdul Rahman Ramli, Badrul Hisham Abdul Kahar, and M-Iqbal Saripan. A framework for white blood cell segmentation in microscopic blood images using digital image processing. *Biological procedures online*, 11(1):196–206, 2009.
28. Leyza Baldo Dorini, Rodrigo Minetto, and Neucimar Jeronimo Leite. Semiautomatic white blood cell segmentation based on multiscale analysis. *IEEE journal of biomedical and health informatics*, 17(1):250–256, 2012.
29. SJ Belekhar and SR Chougule. Wbc segmentation using morphological operation and smmt operator—a review. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(1):434–440, 2015.

30. Qiu Wenhua, Wang Liang, and Qiu Zhenzhen. White blood cell nucleus segmentation based on canny level set. *Sensors & Transducers*, 180(10):85, 2014.
31. Thanh Tran, Oh-Heum Kwon, Ki-Ryong Kwon, Suk-Hwan Lee, and Kyung-Won Kang. Blood cell images segmentation using deep learning semantic segmentation. In *2018 IEEE International Conference on Electronics and Communication Engineering (ICECE)*, pages 13–16. IEEE, 2018.
32. Muhammad Shahzad, Arif Iqbal Umar, Muazzam A Khan, Syed Hamad Shirazi, Zakir Khan, and Waqas Yousaf. Robust method for semantic segmentation of whole-slide blood cell microscopic images. *Computational and mathematical methods in medicine*, 2020, 2020.
33. Vincent Couteaux, Salim Si-Mohamed, Raphaele Renard-Penna, Olivier Nempont, Thierry Lefevre, Alexandre Popoff, Guillaume Pizaine, Nicolas Villain, Isabelle Bloch, Julien Behr, et al. Kidney cortex segmentation in 2d ct with u-nets ensemble aggregation. *Diagnostic and interventional imaging*, 100(4):211–217, 2019.
34. Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.
35. Reena M Roy and PM Ameer. Segmentation of leukocyte by semantic segmentation model: A deep learning approach. *Biomedical Signal Processing and Control*, 65:102385, 2021.
36. Ruggero Donida Labati, Vincenzo Piuri, and Fabio Scotti. All-idb: The acute lymphoblastic leukemia image database for image processing. In *2011 18th IEEE International Conference on Image Processing*, pages 2045–2048. IEEE, 2011.
37. Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation. *arXiv preprint arXiv:1711.08506*, 2017.
38. Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.
39. M Roy Reena and PM Ameer. Localization and recognition of leukocytes in peripheral blood: A deep learning approach. *Computers in Biology and Medicine*, 126:104034, 2020.
40. Seyed Hamid RezaTofighi and Hamid Soltanian-Zadeh. Automatic recognition of five types of white blood cells in peripheral blood. *Computerized Medical Imaging and Graphics*, 35(4):333–343, 2011.