Received: June 20, 2022. Revised: July 21, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Classification of Acute Lymphoblastic Leukemia on White Blood Cell Microscopy Images Based on Instance Segmentation Using Mask R-CNN

Aldinata Rizky Revanda[1] Chastine Fatichah[1]* Nanik Suciati[1]

[1]*Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
* Corresponding author's Email: chastine@if.its.ac.id

Abstract: Manual classification of acute lymphoblastic leukemia carried out by doctors will certainly take a lot of time and effort. The challenges in automated computer-based systems for classification of acute lymphoblastic leukemia are when providing proper lightning in stained white blood cell microscopy images and when segmenting the touching or overlapping cells in the image. Previous studies related to the classification of acute lymphoblastic leukemia still require many steps when using conventional methods, whereas when using deep learning methods are still limited to classification without providing analysis for the instance segmentation of lymphoblast in the image. Therefore, we propose instance segmentation using Mask R-CNN on white blood cell microscopy images to classify acute lymphoblastic leukemia that can support the diagnosis process efficiently and effectively. In this study, we implemented Mask R-CNN by transfer learning method to fit the instance segmentation task on white blood cell microscopy images. We added a contrast enhancement process to the image dataset to overcome the bad lightning problem in stained white blood cell microscopy images. We used the real dataset obtained from hospital to evaluate our method. The method we used was able to get 83.72 % accuracy, 85.17 % precision, and 81.61 % sensitivity.

Keywords: Acute lymphoblastic leukemia, Classification, Instance segmentation, Mask R-CNN, White blood cell.

1 Introduction

White blood cells are one of the other particles in the human body along with red blood cells and platelets. White blood cells are a crucial element because they are closely related to the human immune system to help fight bacteria and viruses [1]. However, an excess or deficiency of white blood cells in the body can be an indication of a dangerous disease [2]. Consequently, the analysis of white blood cells is necessary for early prevention and to avoid delays in the treatment of patients. The most common conditions associated with white blood cell counts are cancer and leukemia.

In this study, leukemia was focused on acute lymphoblastic leukemia (ALL). While most cases of ALL occur in children, representing 80 % of cases, ALL can be devastating in adults [3]. There are five types of white blood cells, namely neutrophils, basophils, eosinophils, monocytes, lymphocytes [4]. ALL is an acute malignancy of white blood cells that causes excess production of immature lymphocytes or called lymphoblasts in the bone marrow [5]. Lymphoblasts are characterized by their irregular shape, the presence of small cavities in the cytoplasm, called vacuoles, and the presence of spherical particles in the nuclei, called nucleoli [6]. According to the French American British (FAB) classification, ALL is further classified into three subtypes, which are L1, L2, and L3, with the classification as follows [7, 8]. Lymphoblast cells of the L1 type tend to be small and homogeneous. The nuclei are round and regular, with little cleft and inconspicuous nucleoli. The cytoplasm is scanty and usually without vacuoles. Lymphoblast cells of the L2 type tend to be large and heterogeneous. The nuclei are irregular and often cleft. One or more large nucleoli are present. The volume of the cytoplasm is variable, but often abundant and may contain vacuoles. Lymphoblast cells of the L3 type tend to be moderate-large in size

and homogeneous. The nuclei are regular and round-oval in shape. One or more prominent nucleoli are present. The volume of cytoplasm is moderate and contains prominent vacuoles.

Pathologists typically count and identify white blood cells by inspecting the blood smears visually through a microscope. However, this manual technique requires a lot of time and effort [9]. Manual technique could potentially cause problems as it produces inaccurate results that are highly dependent on the pathologist's skills and experiences [10]. Different pathologists may produce different results, which can lead to confusion. Nowadays, the industry has turned to automated machines called hematology counters to perform rapid analysis of white blood cells, but these machines are still so expensive [11]. Hence, a computer-based system is needed to help analyze white blood cells, especially lymphoblasts. Some researchers have applied conventional methods of machine learning. A conventional method such as traditional image processing can help but the process will be more complex because it involves interconnected steps such as segmentation, feature extraction, and classification [12]. Before the segmentation step, it is often necessary to preprocess the input image, this preprocessing step includes image enhancement such as contrast adjustment, noise removal, and image sharpening. It makes the conventional method requires more steps depending on the condition of the input image [13]. Moreover, accurate nuclei detection, separation of borders to recover overlapped cells, segmenting RoI, robust features extraction, and best features selection have become challenging and time-consuming using these approaches [14].

In this study, we chose to use the convolutional neural network (CNN) to perform segmentation and classification of lymphoblast to determine the subtypes of ALL. With deep learning we can use images as direct input, so that analysis can be carried out faster and because each pixel in a medical image contains meaningful data, data loss caused by feature extraction can be prevented [15]. Computer vision based on CNN has enabled people to accomplish what had been considered impossible in the past few centuries, such as object detection, face recognition, autonomous vehicles, self-service supermarkets, and intelligent medical treatment [16, 17].

The type of CNN architecture we selected for this study is mask region-based convolutional neural network (Mask R-CNN). Mask R-CNN is a neural network in deep learning that aims to solve the task of instance segmentation in the field of digital image processing or computer vision. Mask R-CNN is a conceptually simple, flexible, and general framework for object instance segmentation [18]. Mask R-CNN can efficiently realize the accurate detection and classification of objects in an image and generate a high-quality segmentation mask for each object simultaneously [18, 19]. This way, Mask R-CNN can separate various objects within an image or video, including objects that touch or overlap. So far, Mask R-CNN has been applied to various computer vision tasks such as fruit detection [20], microscopic algae detection [21], inspection of surface defects [22], parking space detection [23], skin lesion detection and segmentation [19, 24], multiorgan segmentation [25], gastrointestinal diseases segmentation and classification [26], pulmonary embolism detection [27], and nuclei segmentation [28].

Therefore, we propose the enhanced Mask R-CNN on white blood cell microscopy images to classify acute lymphoblastic leukemia. So that it can support the diagnosis process by pathologists efficiently and effectively. In this study, we implemented Mask R-CNN by transfer learning method to fit the instance segmentation task on white blood cell microscopy images, then simultaneously to classify the subtypes of ALL, namely L1, L2 and L3. To improve the performance of Mask R-CNN on this task, we add contrast enhancement and augmentation processes to the image, search the best scenario, and refined the parameters of the Mask R-CNN.

The rest of the paper is organized as follows. Section 2 discusses related works on acute lymphoblastic leukemia detection and classification in general. Section 3 presents the proposed methodology for classification of acute lymphoblastic leukemia on white blood cell microscopy images. Section 4 explains the experiments that have been carried out and shows the results. Finally, conclusion and future work from this study presented in section 5.

2 Literature review

Research on the detection and classification of ALL has been carried out by several researchers before. There are several types of approaches used for this task, namely conventional methods and deep learning methods. In addition, there are several studies that use public dataset, and there are also studies that use their own dataset collected from the hospitals that collaborate with them.

2.1 Conventional method

Detection of ALL based on the color orthonormal basis entropy (COBE) and the distribution of the pixel intensity (DoPI) has been proposed [29]. The method has four main steps, these include the image

enhancement of the green channel by a gamma correction of the contrast stretching then convert to hue channel, the segmentation using Otsu algorithm, the feature extraction using the COBE and the DoPI, and then the similarity measurement using Manhattan method. The acute lymphoblastic leukemia image database-2 (ALL-IDB2) is used to evaluate the method. That method with the best scenario can produce an accuracy of 91.67 %. The deficiency of that method is that there is no feature selection process, object features are only based on object texture, while the shape of the object is not considered as feature, similarity measurements only consider the distance between pixels in the image.

Classification of ALL using the color-based hybrid modeling has been proposed [30]. The method has four main steps, not much different from the previous method, these include the preprocessing using the Otsu thresholding-based contrast stretching, the image segmentation using binary image transformation with Otsu algorithm and the biggest object area selection, the features extraction by combining the color-based densitometry and shape features, and then the similarity measurement by the Euclidian distance and Manhattan method. The dataset used to evaluate that method is the acute lymphoblastic leukemia image database-1 (ALL-IDB1). That method can produce an accuracy of 95.38 %. However, the drawback of that method is that it can only classify positive and negative ALL, not include to classify the subtypes of ALL.

Detection of ALL using image segmentation and data mining algorithms has been proposed [31]. The method has four main steps, these include the segmentation of the nuclei and cytoplasm using K-Medoids algorithm and granulometric analysis, the feature extraction based on shape, visual, pixel value, texture, and fractal dimension, then the feature selection using the InfoGainAttributeEval and the Ranker-Search method, and then the training of the classifier model. The dataset used to evaluate that method comes from several sources, including the Atlas of Hematology, ALL-IDB1, and the Internet. That method with the best scenario can produce an accuracy of 98.60 %. The deficiency of that method is that each step is too complex, especially in the segmentation of the touching cells and the extraction of so many features.

Identification of ALL subtypes in touching cells based on enhanced edge detection has been proposed [32]. The method has three main steps, these include the segmentation based on edge detection and edge linking, the feature extraction of cells and nuclei using the geometry and texture features, and then the classification of ALL subtypes by support vector machines (SVM) classifier and majority vote. That method can produce an accuracy of 75.42 %. The dataset used to evaluate that method is white blood cell microscopy images obtained from the dr. Soetomo hospital of Surabaya. However, the drawback of that method is that the accuracy result is not yet optimized, the feature extraction step requires quite complex calculations, and the majority vote is too simple and less reliable for ALL classification.

2.2 Deep learning method

Classification of ALL with data augmentation and CNN has been proposed [33]. The method used data augmentation to balance the training and validation dataset. The architectures of CNN that used in that method were visual geometry group-16 (VGG-16), visual geometry group-19 (VGG-19), and extreme inception (Xception). The dataset used to evaluate that method is the classification of normal and malignant cells (C-NMC) dataset. That method with the best scenario can produce an accuracy of 92.48 %. The deficiency of that method is that it is not used for classification of ALL subtypes.

Classification of ALL using an attention-based CNN has been proposed [34]. The method used CNN-based model that applies attention module called efficient channel attention (ECA) with VGG-16 architectures. The method also used augmentation techniques to increase the quality and quantity of the dataset. The C-NMC dataset is used to evaluate that method. That method can produce an accuracy of 91.10 %. The drawback of that method is that it is only used to classify healthy cells or ALL cells, and it is only used for single cell images.

Detection and classification of ALL using an efficient deep CNN has been proposed [35]. The method used a probability-based weight factor to make the model of hybridizing the mobile neural network version-2 (MobileNetV2) and the residual neural network-18 (ResNet18) architecture more efficient. The performance of that method is validated using public benchmark dataset, ALL-IDB1 and ALL-IDB2. That method with the best scenario can produce accuracy of 99.39 % and 97.18 % in ALL-IDB1 and ALL-IDB2 dataset, respectively. The deficiency of that method is that it still only to classify of the healthy or ALL cells, although it can be used for multi-cell images.

Detection and classification of ALL subtypes using pretrained deep CNN has been proposed [36]. The method used pretrained Alex-Krizhevsky neural network (AlexNet) which was fine-tuned to their task and dataset. The method also used augmentation technique to reduce overtraining. To evaluate that

method, it used the ALL-IDB2 dataset. For ALL detection task, that method can produce an accuracy of 99.50 %. While for ALL classification task, that method can produce an accuracy of 96.06 %. The drawback of that method is that it still uses single cell images for the dataset and does not provide segmentation result of detected cells.

Classification of ALL subtypes using CNN also has been proposed [37]. the method has an initial segmentation step to separate lymphoblast cells in the image. The segmentation results are then used to train the CNN model to be able to classify the ALL subtypes. That method is evaluated using the dataset that obtained from Amreek Clinical Laboratory, Saidu Sharif, Swat, Khyber Pakhtunkhwa, Pakistan. That can produce an accuracy of 97.78 %. The drawback of that method is that it uses semantic segmentation for multi-cell images, even still using a conventional segmentation method which is separate from CNN architecture.

However, to support the diagnosis process by pathologists, microscopic image segmentation is still required. Especially, the instance segmentation which can detect and separate every cell in the microscopic image, even including the touching cells. Therefore, we carried out a study to classify acute lymphoblastic leukemia on white blood cell microscopy images based on instance segmentation.

3 Materials and method

In this study, several processes were carried out for the classification of acute lymphoblastic leukemia. The processes are contrast enhancement, augmentation, lymphoblast annotation and labeling, training of Mask R-CNN, testing of trained Mask R-CNN, and ALL classification. The contrast enhancement method we use is the exposure fusion framework (EFF). The augmentation that we do is flip the image horizontally and vertically. The lymphoblast annotation and labeling are done using the VGG image annotator (VIA). The training of Mask R-CNN is done by transfer learning. The testing of trained Mask R-CNN aims to obtain instance segmentation of lymphoblast. ALL classification is done by majority vote. The flowchart in this study can be seen in Fig. 1.

3.1 Dataset

The dataset used in this study is the dataset of previous study by Nenden Siti Fatonah which was obtained from dr. Soetomo hospital of Surabaya [32]. This dataset is a collection of white blood cell microscopy images entirely from patients with ALL, so there are no images of healthy patients. This dataset consists of two types of images, there are single-cell and multi-cell images. Since in this study we propose an instance segmentation-based approach that can detect and separating every lymphoblast cell, including the touching cells, we chose to use only multi-cell images from the dataset.

The number of data we use from this dataset is 301 multi-cell ALL images which consists of 128 L1 type images, 63 L2 type images, and 110 L3 type images. White blood cell microscopy images in this dataset have the same shooting scale that is 1000x and the same staining process. White blood cell microscopy images of the three types of ALL in the dataset can be seen in Fig. 2.

This dataset was divided into training data and testing data. For the training data, we generate 60 new images by performing the augmentation process, with a balanced composition of 20 L1 type images, 20 L2 type images, and 20 L3 type images. For the testing

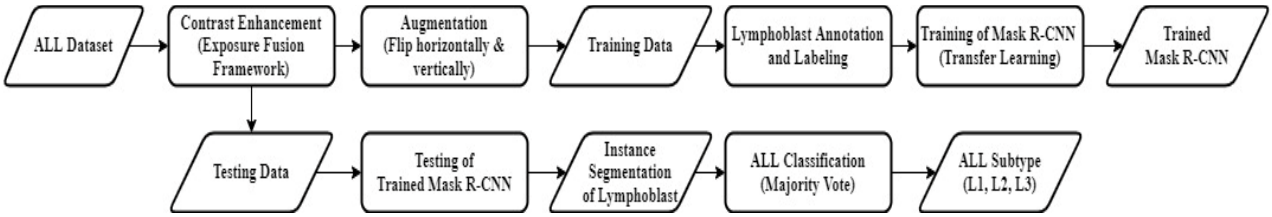
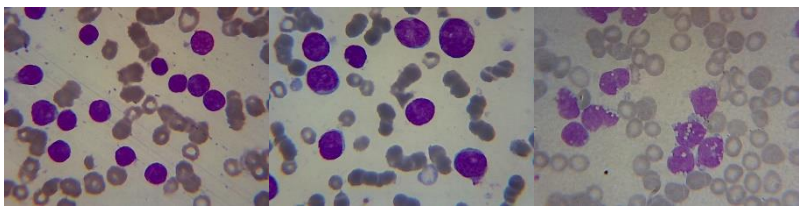


Figure. 1 ALL classification processes based on instance segmentation



(a) (b) (c)

Figure. 2 White blood cell microscopy images of acute lymphoblastic leukemia: (a) L1, (b) L2, and (c) L3

data, we use the entire initial dataset without the segmentation results, 301 multi-cell ALL images.

3.2 Exposure fusion framework

Exposure fusion framework or abbreviated as EFF is a contrast enhancement method by putting an RGB image. The purpose of contrast enhancement process is to solve bad lighting problems, especially in stained white blood cell microscopic images. Contrast enhancement method that has been done was successfully solved low contrast problem and light distortion [38]. Increasing exposure can show areas that have bad exposure, but it can also make areas that already have good exposure become over-exposed at the same time. The EFF formula for obtaining an image with all pixels that are well-positioned to increase exposure is defined in Eq. (1). Well-exposed pixels are assigned a big weight value, whereas poor-exposed pixels are assigned a small weight value that is applied to the non-uniform tricolor component for all pixels, so the normalized weight is defined in Eq. (2).

$$R^c = \sum_{i=1}^{N} W_i \circ P_i^c \tag{1}$$

$$\sum_{i=1}^{N} W_i = 1 \tag{2}$$

Where R is the enhanced result, c is the index of three-color channels, N is the number of images, W_i is the weight map of the i-th image, and P_i is the i-th image in the exposure set. The operator $\circ$ means element-wise multiplication. The problem that may be encountered with this method is that images with other exposure settings are not available. That is because the image is only taken once, and there are no additional images for different exposures. In contrast to images taken at different exposures, the problem associated with image enhancement can be solved easily because the images are strongly correlated. The brightness transform function (BTF) is a mapping function between two images that differ only in exposure to solve the problem of different exposure settings [39]. Eq. (3) states that the input image P can be mapped to the i-th image in the exposure set by using BTF g and exposure ratio k_i.

$$P_i = g(P, k_i) \tag{3}$$

In the contrast enhancement process, the input image is combined with the image itself with other exposures to reduce complexity. Eq. (4) explains the formula for image merging.

$$R^c = W \circ P_i^c + (1 - W) \circ g(P^c, k) \tag{4}$$

The image enhancement problem can be divided into three parts, there are determining the values of W, g, and k. The key to a successful algorithm that can enhance the under-exposed area and maintain the well-exposed area is the design of W. Under-exposed pixels need to assign a small weight value and well-exposed pixels need to assign a big weight value. Intuitively, the scene illumination and the weight matrix will be positively correlated. Well-exposed pixels should be assigned with big weight value to maintain the contrast level of highly illuminated area. Eq. (5) states the weight matrix calculation.

$$W = T^{\mu} \tag{5}$$

Where T is the scene illumination map and μ is a parameter that controlling the enhance degree. Optimization is needed to get the estimated scene illumination map T. Optimization is done by refining T through the optimization formula in Eq. (6) that aims to minimize the difference between the initial map L and the refined map T, while at the same time maintaining the smoothness of the T.

$$T^o = \min_T ||T - L||_2^2 + \lambda ||M \circ \nabla T||_1 \tag{6}$$

Where $\|*\|_2$ and $\|*\|_1$ is the l_2 and l_1 norm, respectively. The first order derivative filter ∇ contains $\nabla_h T$ (horizontal) and $\nabla_v T$ (vertical). M is the weight matrix and λ is the coefficient. For the initial estimation of the illumination for each pixel x, we use the lightness component as defined in Eq. (7). Another important thing for the illumination map refinement is the design of M. As a result, the design of the weight matrix is defined in Eq. (8).

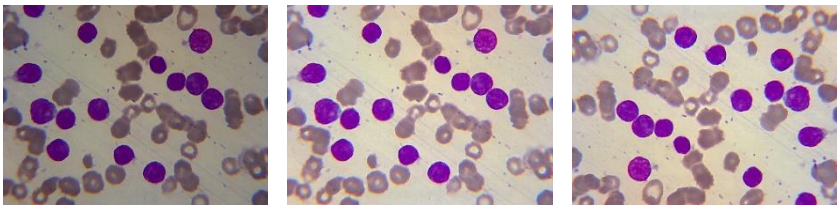
$$L(x) = \max_{c \in \{R,G,B\}} P_c(x) \tag{7}$$

$$M_d(x) = \frac{1}{|\sum_{y \in \omega(x)} \nabla_d L(y)| + \epsilon}, d \in \{h, v\} \tag{8}$$

Where $|*|$ is the absolute value operator, $\omega(x)$ is the local window centered at the pixel x and ϵ is a very small constant to avoid the zero denominator. To reduce the complexity of the proposed closed-form solution, we approximate Eq. (6) as in [40]. The approximate result is defined in Eq. (9), but the problem occurs because many quadratic operations are involved. Let m_d, l, t, and $\nabla_d l$ denotes vector versions of M_d, L, T, and $\nabla_d L$, respectively. Eq. (10) expresses a linear function to obtain the proposed solution, where $\oslash$ is the element-wise division, I is the unit matrix, the operator $Diag(v)$ is an operation to construct the diagonal matrix using the vector v,

$$T^o = \min_T \sum_x \left(\left(T(x) - L(x) \right)^2 + \lambda \sum_{d \in \{h,v\}} \frac{M_d(x)(\nabla_d T(x))^2}{|\nabla_d L(x)| + \epsilon} \right) \quad (9)$$

$$I + \lambda \sum_{d \in \{h,v\}} (D_d^T Diag(m_d \oslash (|\nabla_d l| + \epsilon)) D_d) t = l \quad (10)$$



(a) (b) (c)

Figure. 3 ALL images: (a) original image, (b) contrast enhancement result, and (c) augmentation result

and D_d is the Toeplitz matrices from the discrete gradient operators with forward difference. The design of weight matrix M is the main difference between the illumination map estimation method that used in [38] and [40]. An image with contrast enhancement is presented in Fig. 3.

3.3 Flip horizontally and vertically

In this study, we used the real dataset obtained from hospital and we only used multi-cell image types from the dataset, so the number of data we used was limited to only 301 images. Data augmentation techniques can be used to overcome the problem of the limited size of a dataset. Therefore, we decided to slightly augment our dataset by performing a simple image augmentation. Several previous studies have also proposed to perform image augmentation process to increase or balancing the variety and even quality of their dataset images [33, 34, 36]. In this study, we will use the results of image augmentation process as training data, while the initial 301 images will be used as testing data.

For this image augmentation we use a standard image transformation technique that is flipping horizontally and vertically. This technique will horizontally and vertically flip the entire rows and columns of an image to produce a mirror image, which means it will not affect the characteristics of each cell in the image. For the training data, we purposely created 60 new images with a balanced composition of 20 images for each type of ALL. An image with augmentation is presented in Fig. 3.

3.4 Lymphoblast annotation and labeling

Giving object annotation and labeling to the training data image is one of the mandatory steps, because it aims to create ground truth that will be used as input to train the Mask R-CNN model. Previous studies that have used Mask R-CNN also perform the process of annotating and labeling to create their ground truth [19, 20, 21, 26].

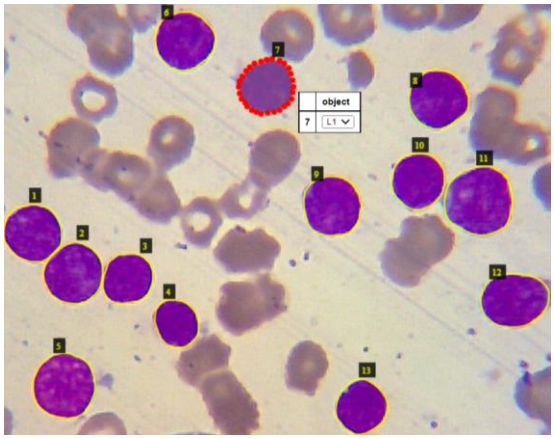


Figure. 4 Lymphoblast annotation and labeling result by using the VGG image annotator

The annotation means to provide mask or connected points around the lymphoblast and then give it the label or class, whether they belong to the L1, L2, or L3 type. This mask image is then used to calculate the reverse loss in model training and optimization of the model parameters. In addition, the performance of the trained model for instance segmentation was evaluated by comparing the annotated mask image as the ground truth against the predicted mask result. An image with annotation and labeling is presented in Fig. 4.

The annotation and labeling process is done using the VGG image annotator (VIA) and saved in the form of a JavaScript object notation (JSON) file. VIA is a simple and standalone manual annotation software for images, audio, and video [41]. VIA is just a single hyper-text markup language (HTML) file that can be opened in any standard web browser, without any installation or setup required.

3.5 Training of Mask R-CNN

Mask region-based convolutional neural network or Mask R-CNN is a method developed from Faster

R-CNN by adding branches to predict mask objects parallel with the existing branch to detect the square limit [18]. The main point is to separate the classification task and pixel-level mask prediction.

Based on the Faster R-CNN framework [42], a third branch was added to predict object masks in parallel with the existing branch for classification and localization. The branch mask is a fully connected small network that is applied to each region of interest (RoI), predicting pixel-to-pixel mask segmentation. Through this branch mask it is possible to select pixels wisely on each object segment and extract them separately without any background (which is not possible in semantic segmentation).

Since pixel-level segmentation requires more fine-grained alignment than the boundaries box, the Mask R-CNN enhances the RoI pooling layer using RoIAlign so that it can be better and more precisely mapped to the original image area. The R-CNN model processes the input image provided to the network by performing a selective search. Then, it uses the selective search result area for feature extraction and classification using the trained CNN.

The Fast R-CNN model still uses a selective search algorithm to get the proposal region [43], but RoI pooling module is added to this model. The Fast R-CNN model can extract fixed-size windows from the feature maps, then use these features to get the final class label and boundary box. The advantage of the Fast R-CNN model is that the network can now be trained end-to-end. The Faster R-CNN model introduced the region proposal network (RPN), which builds region proposals of each object directly into the architecture to increase the need for selective search algorithms.

The Mask-RCNN implementation used in this study was built by Matterport [44], based on the Keras and TensorFlow frameworks. To train the Mask R-CNN model, we use transfer learning method to solve the problem of the small dataset. As feature extraction CNN, a residual neural network-101 (ResNet101) [45] pretrained with the Microsoft common objects in context (MS COCO) dataset was used. In addition, this implementation uses a modified version of ResNet with the feature pyramid network (FPN) architecture [46], which is a top-down approach that allows extracting features at different scales and gives better results in accuracy and speed. The training procedure was configured with a learning rate of 0.001, 20 epochs, 100 steps per epoch, and 0.9 minimum confidence. The selected optimizer was the stochastic gradient descent (SGD) with a momentum coefficient of 0.9 and the mini-batch size was set to 2 images. During the training process, the training dataset will be separated into two parts with a split validation ratio of 3:1. To assess model performance during the training process, we measure and compare the mean average precision (mAP) values. So, the result of the training process is a trained Mask R-CNN model that can perform instances segmentation on white blood cell microscopic images infected with ALL, it means the model is able to detect which cells belong to ALL and classify each lymphoblast cell to the ALL subtypes.

3.6 Testing of trained Mask R-CNN

Furthermore, the trained Mask R-CNN can be used to perform instances segmentation on the testing dataset. The output of Mask R-CNN is in the form of bounding boxes, predicted masks, and class probabilities for each lymphoblast cell in the image. Even for cells that touch or overlap each other, Mask R-CNN can handle them directly without requiring any additional processing. From these results, we can record every single lymphoblast cell in the image, which belongs to ALL subtypes, L1, L2, or L3. Then with these data we can determine the subtypes of the multi-cell image by using the majority vote.

3.7 Majority vote

To decide on the class of a multi-cell image, in this case consisting of several single-cell lymphoblasts, we need to vote on the majority class of each detected single-cell class. The rules we use are recommendations from pathologists. First, if there is only one or more single cells detected and predicted as class L3 then the conclusion is that the image is classified into ALL type L3. Next, if the number of single cells detected and predicted as class L2 is equal to or more than class L1 then the conclusion is that the image is classified into ALL type L2. Furthermore, if the number of single cells detected and predicted as class L1 is more than L2 class then the conclusion is that the image is classified into ALL type L1. Finally, after getting the prediction results from the multi-cell images, we can evaluate the performance of our proposed method using a confusion matrix to measure accuracy, precision, and sensitivity.

4 Result and analysis

In this section, we will discuss the results of the instances segmentation performed by the trained Mask R-CNN and the classification results after going through the majority vote process. We will analyze it based on quantitative and qualitative values. In addition, some related outputs will be presented in

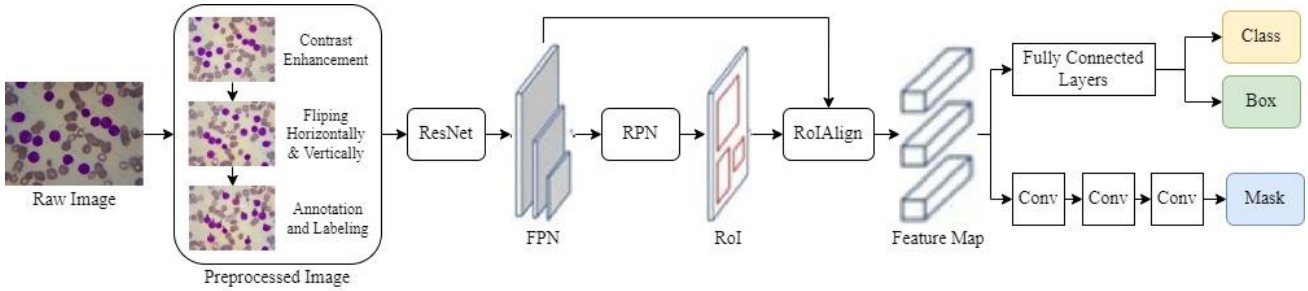
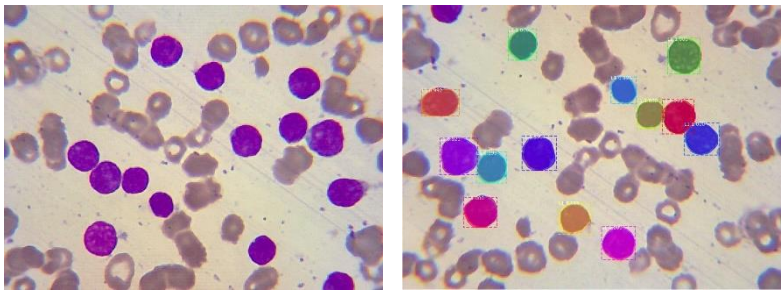
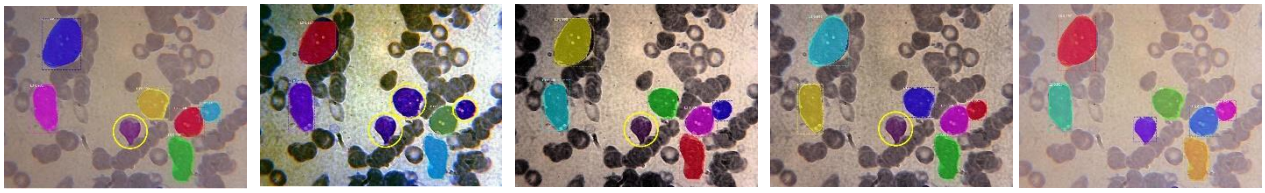


Figure. 5 The architecture of enhanced Mask R-CNN for lymphoblast instance segmentation



(a) (b)

Figure. 6 The result of lymphoblast instance segmentation: (a) input image and (b) instance segmentation result



(a) (b) (c) (d) (e)

Figure. 7 The result of improved lymphoblast instance segmentation: (a) without contrast enhancement, (b) HE, (c) DHE, (d) CLAHE, and (e) EFF

this section.

4.1 Lymphoblast instance segmentation

The lymphoblast instance segmentation stage using Mask R-CNN aims to get the class of ALL subtypes from every single cell that is predicted into ALL and then recorded for use in the classification stage. The multi-cell microscopic images of the patients with ALL are used as input. The images will be given a preprocessing before used in the lymphoblast instance segmentation stage. The proposed system introduces a contrast enhancement process in Mask R-CNN. The contrast enhancement process was added to solve the lighting problem in the lymphoblast instance segmentation stage using Mask R-CNN. Fig. 5 is the architecture of enhanced Mask R-CNN and the training process.

The lymphoblast instance segmentation stage proposed are applied to 60 multi-cell images of training dataset and 301 multi-cell images of testing dataset. Ground truth data for the training of Mask R-CNN is made from the training dataset by the annotation and labeling process. The results of lymphoblast instance segmentation are in the form of bounding boxes, predicted masks, and class probabilities for each lymphoblast in the image. An example from the result of lymphoblast instance segmentation is presented in Fig. 6.

Table 1. The mAP values for evaluate the lymphoblast instance segmentation

Contrast Enhancement	mAP (%)
without	81.96
HE	60.59
DHE	75.82
CLAHE	78.53
EFF	**83.65**

In this study, we measure the performance of lymphoblast instance segmentation quantitatively. For the performance evaluation, we use the mean average precision (mAP) value to measure how accurately the lymphoblast instance segmentation. Thus, the higher the mAP value, the more accurate the lymphoblast instance segmentation results. Table 1 is the mAP values of the experiments conducted without contrast enhancement and the four contrast enhancement methods according to the scenario. The table states that the best contrast enhancement

method for multi-cell white blood microscopic images in lymphoblast instance segmentation is the method of exposure fusion framework (EFF).

Fig. 7 is an example of the results of qualitatively improved lymphoblast instance segmentation from the experiments conducted. From these results, without the contrast enhancement process, lymphoblast cells that were difficult to detect were marked by yellow circles in Fig. 7 (a). Meanwhile, with some commonly used contrast enhancement methods, such as histogram equalization (HE), dynamic histogram equalization (DHE), and contrast limited adaptive histogram equalization (CLAHE), still difficult to detect the lymphoblast cells which marked by yellow circles in Fig. 7 (b, c, and d).

4.2 ALL subtypes classification

After getting the predicted class results for each lymphoblast in multi-cell images, these classes will be used by majority vote rule for classify the ALL subtypes belonging to the multi-cell images. Fig. 8 explained the ALL subtypes classification process performs by majority vote. First, we record the detected lymphoblasts and count the number of each lymphoblast class. After that, by using the majority vote rule, we can get the results of the ALL subtypes classification stage for multi-cell images.

Quantitatively for the performance evaluation of the ALL subtypes classification stage, we used a confusion matrix. There are several performance values that can be obtained from the confusion matrix, such as accuracy, precision, and sensitivity. Accuracy is a measure of experimental results that calculates the truth ratio of all data, the number of correct predictions divided by the total of data. Precision provides information on how many cases that were predicted to be positive turn out to be positive in fact. Sensitivity provides information on how many positive cases are in fact that can be correctly predicted positive by the classification model. Table 2 is the performance values of the ALL subtypes classification stage by using confusion matrix for each experiment scenario.

Once again, the table states that the best contrast enhancement method for multi-cell white blood microscopic images in ALL subtypes classification is the method of exposure fusion framework (EFF). Therefore, the detail of the confusion matrix for performance evaluation using the additional contrast enhancement process of the EFF method is presented in Table 3.

From the confusion matrix table, we can see that the proposed method is able to perform robust classification for the L1 and L3 subtypes. However,

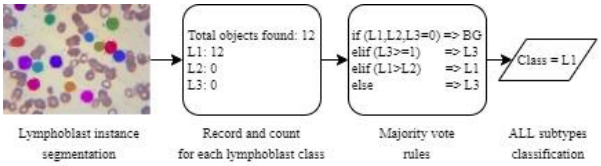


Figure. 8 The ALL subtypes classification process

Table 2. The performance values of the ALL subtypes classification stage by using confusion matrix

Contrast Enhancement	Accuracy (%)	Precision (%)	Sensitivity (%)
without	80.07	79.91	79.41
HE	77.08	77.88	79.48
DHE	74.09	73.68	72.90
CLAHE	74.09	73.04	71.10
EFF	**83.72**	**85.17**	**81.61**

Table 3. Confusion matrix result of ALL subtypes classification using the proposed method

		Prediction			
		L1	**L2**	**L3**	**FN**
Actual	**L1**	**114**	10	1	11
	L2	25	**30**	0	25
	L3	1	0	**108**	1
	FP	26	10	1	

for the L2 subtype, the proposed method still seems difficult to overcome, because the characteristics of L1 and L2 are almost the same and even difficult to distinguish manually by pathologists. The difficulty of the L2 subtype classification can also be observed based on the sensitivity values for each subtype class, 91.20 % for L1, 54.55 % for L2, and 99.08 % for L3.

4.3 Approach comparison

The proposed system aims to do the ALL subtypes classification automatically based on the lymphoblast instance segmentation, in which most studies still require manual detection and segmentation of lymphoblast cells. Another contribution is the use of enhanced Mask R-CNN to improve the accuracy and efficiency. From the evaluation results, our approach was able to detect previously undetectable lymphoblast cells on white blood cell microscopy images. A comparison of the related approaches for the detection and classification of ALL is presented in Table 4.

For performance comparison, it would be fair to choose a previous study that used same dataset. The dataset used in this study is the dataset of previous study by Nenden Siti Fatonah which was obtained from dr. Soetomo hospital of Surabaya [32]. Therefore, we will compare the performance of the proposed method with the previous study by Nenden Siti Fatonah. The performance comparison of the

Table 4. Comparison with existing approaches

Type of approach	Method	Lymphoblast detection	Instance segmentation	Provide mask on lymphoblast	ALL subtypes classification
Conventional	COBE and DoPI [29]	Yes	No	No	No
	The color-based hybrid modeling [30]	Yes	No	No	No
	Image segmentation and data mining [31]	Yes	Yes	No	Yes
	Enhanced edge detection and SVM [32]	Yes	Yes	No	Yes
Deep learning	Data augmentation and CNN (VGG-16) [33]	Yes	No	No	No
	Attention-based CNN (ECA and VGG-16) [34]	Yes	No	No	No
	Efficient deep CNN (MobilenetV2 and ResNet18) [35]	Yes	No	No	No
	Pretrained deep CNN (AlexNet) [36]	Yes	No	No	Yes
	Semantic segmentation and CNN classifier [37]	Yes	No	No	Yes
	Proposed Method (Enhanced Mask R-CNN)	**Yes**	**Yes**	**Yes**	**Yes**

Table 5. The performance comparison under the same dataset

Method	Accuracy (%)	Precision (%)	Sensitivity (%)
Enhanced edge detection and SVM [32]	75.42	80.55	74.92
Proposed Method (Enhanced Mask R-CNN)	**83.72**	**85.17**	**81.61**

related approach under the same dataset is presented in Table 5. The proposed automated system can support the manual diagnosis process by the pathologist. As a result, our approach can be applied directly to white blood cell microscopy images without the costly time and effort.

5 Conclusion

This study developed an automatically classification of acute lymphoblastic leukemia (ALL) based on instance segmentation. The instance segmentation is done by Mask R-CNN architecture on deep learning method. This system has two main stages, the first stage is lymphoblast instance segmentation on multi-cell images. In this step, we propose an enhanced Mask R-CNN. The contrast enhancement process is added using the exposure fusion framework (EFF) method, which aims to solve bad lighting problems, especially in stained white blood cell microscopic images. The second stage is ALL subtypes classification, which used the majority vote rule. We record the detected lymphoblasts and count the number of each lymphoblast class from the first stage. After that, by using the majority vote rule, we classify the ALL subtypes for multi-cell images. The proposed method was able to get 83.72 % accuracy, 85.17 % precision, and 81.61 % sensitivity.

The accurate classification of ALL is required to be developed in clinical applications to support the diagnosis process by pathologists efficiently and

effectively. It is still challenging to find the best method for this task, especially in the process of detection, segmentation, and classification of lymphoblast. Even the condition of the dataset is also challenging in this task because the condition of the original image obtained from the hospital can be different from the condition of the image obtained from the open-source dataset on the internet. Therefore, it is often necessary to add several different preprocesses to adjust the condition of the dataset. For the next studies, we took the initiative to carry out more experiments and modifications on the CNN architecture that can reliably performing the task of instance segmentation on white blood cell microscopic images. In addition, the future studies can be extended to classify another type of leukemia such as acute myeloid leukemia (AML) or even in other cases of red blood cell cancer such as polycythaemia vera (PV).

Conflicts of interest

The authors declare have no conflict of interest.

Author contributions

Conceptualization, Aldinata Rizky Revanda, Chastine Fatichah, Nanik Suciati; methodology, Aldinata Rizky Revanda, Chastine Fatichah, Nanik Suciati; software, Aldinata Rizky Revanda; validation, Aldinata Rizky Revanda, Chastine Fatichah, Nanik Suciati; formal analysis, Aldinata Rizky Revanda, Chastine Fatichah, Nanik Suciati; investigation, Aldinata Rizky Revanda, Chastine Fatichah, Nanik Suciati; resources, Aldinata Rizky Revanda, Chastine Fatichah; data curation, Aldinata Rizky Revanda, Nanik Suciati; writing—original draft preparation, Aldinata Rizky Revanda; writing—review and editing; Chastine Fatichah, Nanik Suciati; visualization, Aldinata Rizky Revanda; supervision, Chastine Fatichah, Nanik Suciati.

Acknowledgments

The authors would like to thank Institut Teknologi Sepuluh Nopember (ITS) for supporting this study through fresh graduate scholarship and research grant.

References

[1] A. M. Patil, M. D. Patil, and G. K. Birajdar, "White Blood Cells Image Classification Using Deep Learning with Canonical Correlation Analysis", *Innovation and Research in BioMedical Engineering* (*IRBM)*, Vol. 42, No. 5, pp. 378–389, 2021.

[2] H. Kutlu, E. Avci, and F. Özyurt, "White Blood Cells Detection and Classification Based on Regional Convolutional Neural Networks", *Medical Hypotheses*, Vol. 135, 2020.

[3] T. Terwilliger and M. A. Hay, "Acute Lymphoblastic Leukemia: A Comprehensive Review and 2017 Update", *Blood Cancer Journal*, Vol. 7, No. 6, 2017.

[4] M. Z. Othman, T. S. Mohammed, and A. B. Ali, "Neural Network Classification of White Blood Cell Using Microscopic Images", *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 5, pp. 99–104, 2017.

[5] T. Pansombut, S. Wikaisuksakul, K. Khongkraphan, and A. Phon-on, "Convolutional Neural Networks for Recognition of Lymphoblast Cell Images", *Computational Intelligence and Neuroscience*, Vol. 2019, 2019.

[6] C. D. Ruberto and L. Putzu, "White Blood Cells Identification and Classification from Leukemic Blood Image", In: *Proc. of International Work-Conf. on Bioinformatics and Biomedical Engineering*, Granada, Spain, pp. 99–106, 2013.

[7] R. D. Labati, V. Piuri, and F. Scotti, "All-IDB: The Acute Lymphoblastic Leukemia Image Database for Image Processing", In: *Proc. of 18th IEEE International Conf. on Image Processing*, Brussels, Belgium, pp. 2045–2048, 2011.

[8] J. Laosai and K. Chamnongthai, "Classification of Acute Leukemia Using Medical-Knowledge-Based Morphology and CD Marker", *Biomedical Signal Processing and Control*, Vol. 44, pp. 127–137, 2018.

[9] Y. Lu, X. Qin, H. Fan, T. Lai, and Z. Li, "WBC-Net: A White Blood Cell Segmentation Network Based on UNet++ and ResNet", *Applied Soft Computing*, Vol. 101, 2021.

[10] S. N. M. Safuan, M. R. M. Tomari, and W. N. W. Zakaria, "White Blood Cell (WBC) Counting Analysis in Blood Smear Images Using Various Color Segmentation Methods", *Measurement*, Vol. 116, pp. 543–555, 2018.

[11] S. N. M. Safuan, M. R. M. Tomari, W. N. W. Zakaria, N. Othman, and N. S. Suriani, "Computer Aided System (CAS) of Lymphoblast Classification for Acute Lymphoblastic Leukemia (ALL) Detection Using Various Pre-Trained Models", In: *Proc. of IEEE Student Conf. on Research and Development (SCOReD)*, Batu Pahat, Malaysia, pp. 411–415, 2020.

[12] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Comparison of Traditional Image Processing and Deep Learning Approaches for

Classification of White Blood Cells in Peripheral Blood Smear Images", *Biocybernetics and Biomedical Engineering*, Vol. 39, No. 2, pp. 382–392, 2019.

[13] A. Bodzas, P. Kodytek, and J. Zidek, "Automated Detection of Acute Lymphoblastic Leukemia From Microscopic Images Based on Human Visual Perception", *Frontiers in Bioengineering and Biotechnology*, Vol. 8, 2020.

[14] S. Khan, M. Sajjad, T. Hussain, A. Ullah, and A. S. Imran, "A Review on Traditional Machine Learning and Deep Learning Models for WBCs Classification in Blood Smear Images", In: *Proc. of IEEE Access*, Vol. 9, pp. 10657–10673, 2021.

[15] Y. Y. Baydilli and Ü. Atila, "Classification of White Blood Cells Using Capsule Networks", *Computerized Medical Imaging and Graphics*, Vol. 80, 2020.

[16] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent Advances in Convolutional Neural Networks", *Pattern Recognition*, Vol. 77, pp. 354–377, 2018.

[17] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects", In: *Proc. of IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN", In: *Proc. of IEEE International Conf. on Computer Vision (ICCV)*, Venice, Italy, pp. 2980–2988, 2017.

[19] C. Huang, A. Yu, Y. Wang, and H. He, "Skin Lesion Segmentation Based on Mask R-CNN", In: *Proc. of International Conf. on Virtual Reality and Visualization (ICVRV)*, Recife, Brazil, pp. 63–67, 2020.

[20] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit Detection for Strawberry Harvesting Robot in Non-Structural Environment Based on Mask R-CNN", *Computers and Electronics in Agriculture*, Vol. 163, 2019.

[21] J. R. Santaquiteria, G. Bueno, O. Deniz, N. Vallez, and G. Cristobal, "Semantic versus Instance Segmentation in Microscopic Algae Detection", *Engineering Applications of Artificial Intelligence*, Vol. 87, 2020.

[22] S. Hou, B. Dong, H. Wang, and G. Wu, "Inspection of Surface Defects on Stay Cables Using a Robot and Transfer Learning", *Automation in Construction*, Vol. 119, 2020.

[23] A. A. Naufal, C. Fatichah, and N. Suciati, "Preprocessed Mask R-CNN for Parking Space Detection in Smart Parking Systems", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 6, pp. 255–265, 2020, doi: 10.22266/ijies2020.1231.23.

[24] M. A. Khan, T. Akram, Y. D. Zhang, and M. Sharif, "Attributes Based Skin Lesion Detection and Recognition: A Mask R-CNN and Transfer Learning-Based Deep Learning Framework", *Pattern Recognition Letters*, Vol. 143, pp. 58–66, 2021.

[25] J. H. Shu, F. D. Nian, M. H. Yu, and X. Li, "An Improved Mask R-CNN Model for Multiorgan Segmentation", *Mathematical Problems in Engineering*, Vol. 2020, 2020.

[26] M. A. Khan, M. A. Khan, F. Ahmed, M. Mittal, L. M. Goyal, D. J. Hemanth, and S. C. Satapathy, "Gastrointestinal Diseases Segmentation and Classification Based on Duo-Deep Architectures", *Pattern Recognition Letters*, Vol. 131, pp. 193–204, 2020.

[27] K. Long, L. Tang, X. Pu, Y. Ren, M. Zheng, L. Gao, C. Song, S. Han, M. Zhou, and F. Deng, "Probability-Based Mask R-CNN for Pulmonary Embolism Detection", *Neurocomputing*, Vol. 422, pp. 345–353, 2021.

[28] H. Jung, B. Lodhi, and J. Kang, "An Automatic Nuclei Segmentation Method Based on Deep Convolutional Neural Networks for Histopathology Images", *BMC Biomedical Engineering*, Vol. 1, 2019.

[29] A. Muntasa and M. Yusuf, "A Novel Approach to Detect the Acute Lymphoblastic Leukemia Based on the Color Orthonormal Basis Entropy (COBE) and the Distribution of the Pixel Intensity (DoPI)", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 1, pp. 124–134, 2020, doi: 10.22266/ijies2020.0229.12.

[30] A. Muntasa and M. Yusuf, "Color-Based Hybrid Modeling to Classify the Acute Lymphoblastic Leukemia", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, pp. 408–422, 2020, doi: 10.22266/ijies2020.0831.36.

[31] V. Acharya and P. Kumar, "Detection of Acute Lymphoblastic Leukemia Using Image Segmentation and Data Mining Algorithms", *Medical & Biological Engineering & Computing*, Vol. 57, pp. 1783–1811, 2019.

[32] N. S. Fatonah, H. Tjandrasa, and C. Fatichah, "Identification of Acute Lymphoblastic Leukemia Subtypes in Touching Cells Based on Enhanced Edge Detection", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, pp. 204–215, 2020, doi: 10.22266/ijies2020.0831.18.

[33] J. M. D. Oliveira and D. Dantas, "Classification

of Normal versus Leukemic Cells with Data Augmentation and Convolutional Neural Networks", In: *Proc. of the 16th International Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Setúbal, Portugal, pp. 685–692, 2021.

[34] M. Z. Ullah, Y. Zheng, J. Song, S. Aslam, C. Xu, G. D. Kiazolu, and L. Wang, "An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification", *Applied Sciences*, Vol. 11, No. 22, 2021.

[35] P. K. Das and S. Meher, "An Efficient Deep Convolutional Neural Network Based Detection and Classification of Acute Lymphoblastic Leukemia", *Expert Systems with Applications*, Vol. 183, 2021.

[36] S. Shafique and S. Tehsin, "Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks", *Technology in Cancer Research & Treatment*, Vol. 17, 2018.

[37] A. Rehman, N. Abbas, T. Saba, S. I. U. Rahman, Z. Mehmood, and H. Kolivand, "Classification of Acute Lymphoblastic Leukemia Using Deep Learning", *Microscopy Research & Technique*, Vol. 81, No. 11, pp. 1310–1317, 2018.

[38] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A New Image Contrast Enhancement Algorithm Using Exposure Fusion Framework", In: *Proc. of International Conf. on Computer Analysis of Images and Patterns*, pp. 36–46, 2017.

[39] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A New Low-Light Image Enhancement Algorithm Using Camera Response Model", In: *Proc. of IEEE International Conf. on Computer Vision Workshops (ICCVW)*, Venice, Italy, pp. 3015–3022, 2017.

[40] X. Guo, Y. Li, and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation", *IEEE Transactions on Image Processing*, Vol. 26, No. 2, pp. 982–993, 2017.

[41] A. Dutta and A. Zisserman, "The VIA Annotation Software for Images, Audio and Video", In: *Proc. of the 27th ACM International Conf. on Multimedia*, Nice, France, pp. 2276–2279, 2019.

[42] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137–1149, 2017.

[43] R. Girshick, "Fast R-CNN", In: *Proc. of IEEE International Conf. on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440–1448, 2015.

[44] W. Abdulla, "Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow", 2017. [Online]. Available: https://github.com/matterport/Mask_RCNN.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770–778, 2016.

[46] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection", In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936–944, 2017.