ALL-FABNET: Acute Lymphocytic Leukemia Segmentation Using a Flipping Attention Block Decoder-Encoder Network

Ammar S. Al-Zubaidi [1], Mohammed Al-Mukhtar [1,*], and Ammar Awni Abbas Baghdadi [2]

[1] Computer Center, University of Baghdad, Baghdad, Iraq
[2] College of Mass Media Baghdad, University of Baghdad, Iraq
Email: ammar.sabah@uobaghdad.edu.iq (A.S.A.Z.); mohammed.abdul@cc.uobaghdad.edu.iq (M.A.M.); ammar.a@comc.uobaghdad.edu.iq (A.A.A.B.)
*Corresponding author

***Abstract*—Acute Lymphoblastic Leukemia (ALL) is a malignant neoplasm defined by the abnormal proliferation of immature lymphocytes in the hematopoietic system, specifically in the blood or bone marrow. The efficacy of ALL treatment is closely linked to its timely identification. Currently, the first diagnosis of ALL involves clinicians laboriously and fallibly examining stained blood smear microscopy images. Recently, deep learning techniques in biomedical diagnostics, focusing on human-centric approaches, have emerged as a potent tool to aid clinicians in their decision-making processes. As a result, researchers have devised a multitude of computer-aided diagnostic methods to detect ALL in blood images autonomously. However, most existing techniques for segmenting White Blood Cells (WBCs) do not consider the need for concurrent segmentation of the cytoplasm and nucleus. It is important to note that a significant drawback of the currently employed networks is their limited computational efficiency, which necessitates a substantial quantity of trainable parameters. The proposed deep learning model demonstrates favorable outcomes and can potentially be used to develop a dependable computer-aided detection system for leukemia malignancy. we suggest an Attention-Flipping Block (FAB) for the lightweight ALL-image segmentation model. It is evaluated using three publicly accessible datasets consisting of blood samples from individuals diagnosed with leukemia. These datasets are specifically referred to as C-NMC 2019, ALL_IDB1, and ALL_IDB2. With ALL-IDB2, the model's segmentation accuracy is 93.56%, and its classification accuracy is 97.94 %, with an F1-Score of 97.65%.**

***Keywords*—lymphoblastic, leukemia, segmentation, classification, Densely Connected Convolutional Networks (DenseNet), ALL-FABNET**

I. Introduction

Leukemia is a cancer of the hematopoietic system that causes White Blood Cells (WBCs) to replicate improperly in the bone marrow. Leukemia may be acute or chronic [1, 2]. Acute leukemia exhibits rapid onset and advances to an advanced stage, whereas chronic leukemia manifests a more protracted course of deterioration. Based on the classification method known as the French-American-British (FAB) system, acute leukemia can be categorized into two main types: Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML). Similarly, chronic leukemia can be classified into two distinct subgroups: chronic lymphocytic leukemia (CLL) and Chronic Myeloid Leukemia (CML). Acute Lymphoblastic Leukemia (ALL) is a blood cancer that spreads quickly and mostly affects lymphoid precursor cells in the bone marrow, bloodstream, and other parts of the body. Blood has three Essential Components: Platelets (PLT), Red Blood Cells (RBC), and White Blood Cells (WBC). Since bone marrow produces blood cells, ALL usually begins there [3–6]. A healthy white blood cell has a nucleus in the middle and a cytoplasm on the outside. Patients with acute lymphoblastic leukemia exhibit an absence of cytoplasm from the cell, with the nucleus entirely encasing it. Therefore, the WBC nucleus has a different shape and texture [7]. In addition, the cells diminish excessively in both size and number. WBC counts are considered important markers for diagnosing various infections in medicine. If treatment is not received or if it is not discovered in its early stages, ALL can potentially be deadly due to its rapid spread throughout the body. ALL is easily curable if identified in its early stages [8].

To detect acute lymphoblastic leukemia, hematologists require patients to have a full blood test since the disease alters WBC count and morphology. Manually counting white blood cells and assessing their morphology is a laborious, time-consuming, error-prone, and subjective process [9, 10]. Therefore, automated technologies based on artificial intelligence must supplement or replace manual processes. More accurate algorithms for diagnosing hematological abnormalities are being developed with the help of recent developments in artificial intelligence. AI image processing and automated detection are easy ways to cut down on human mistakes.

Manuscript received December 30, 2023; revised March 1, 2024; accepted March 8, 2024; published July 19, 2024.

The technique involves utilizing a pre-trained machine learning model on a fresh dataset with the expectation that the model's ability to distinguish between different classes will remain beneficial. Most high-performing models undergo extensive training, so they require less time to adapt to a new dataset. These factors enable the optimization of the model on a specific dataset with a smaller amount of data, making large datasets unnecessary for achieving excellent performance [11].

Currently, Convolutional Neural Networks (CNNs) are highly efficient for diagnosing and categorizing normal and blast cells in medical imaging applications [12]. Utilizing CNNs requires substantial data and computational resources for training. In numerous scenarios, available datasets may prove inadequate for training a CNN. Transfer learning is a precise strategy that involves utilizing CNN models for a specific case while reducing the computational burden [13].

In addition, segmentation plays a crucial role in both overall performance and disease diagnosis. The primary goal of this process is to isolate the targeted white blood cells by removing platelets and red blood cells and to separate any cells that overlap. The essential components of this system are three segmentation methods: signal and image processing techniques, machine learning techniques, and deep learning techniques. Typically, machine learning and deep learning-based techniques perform better than the former.

According to published research, identifying acute lymphoblastic leukemia requires segmenting the WBC nucleus. Reducing the presence of other elements, such as RBCs, platelets, and the background, is necessary for high WBC detection accuracy. However, most existing techniques for the segmentation of White Blood Cells (WBCs) do not consider the need for concurrent segmentation of the cytoplasm and nucleus [14]. It is important to note that a significant drawback of the currently employed networks is their limited computational efficiency, which necessitates a substantial quantity of trainable parameters. To address these problems, our proposed model employs attention mechanisms in conjunction with an encoder-decoder network to enhance models' capacity to focus their processing on certain regions of the input sequence or image. This, in turn, improves the efficiency of information processing. Attention blocks, such as transformers, have become an essential element of various deep learning systems. The flipped attention block structure relies on the memory unit to convert the input into a higher-dimensional space. As a result, the flipped attention block structure can provide a more comprehensive description of the overall feature information in the dataset. For training our network, we utilized openly accessible datasets, specifically the C_NMC_2019 and Acute Lymphoblastic Leukemia (All) databases. The C_NMC_2019 dataset comprises A total of 15,114 images of Acute Lymphoblastic Leukemia (ALL) were gathered from 118 patients. Each image has a resolution of 450×450 pixels and a black background. The ALL-IDB dataset consists of two subsets. One subset has 260 segmented lymphocytes, with 130 belonging to the leukemia class and the remaining 130 belonging to the non-leukemia class. The dataset is specifically designed for classification purposes. The second subset consists of around 108 unsegmented blood images that are categorized into leukemia and non-leukemia groups. A sample of datasets utilized in our work is shown in Fig. 1.

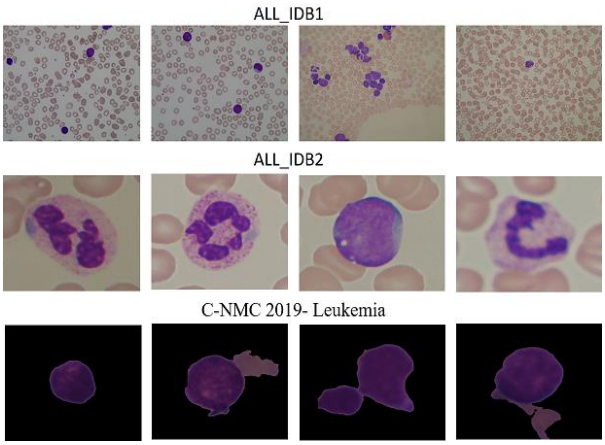

Fig. 1. Samples of three datasets. The first row illustrates the ALL_IDB1 dataset, the second row is ALL-IDB2, and the third row is C-NMC 2019-Leukemia.

The collection consists of 108 images gathered in September 2005. It comprises around 39,000 blood constituents, with the lymphocytes identified by experienced oncologists. The ALL-IDB2 dataset [15] consists of cropped images of normal and blast cells that are part of the ALL-IDB1 dataset. The collection consists of 260 photos, with 50% depicting lymphoblasts. images from ALL-IDB2 exhibit comparable gray-level characteristics to those of ALLIDB1, with the only difference being the size of the images. As the dataset is small amount, we used a data augmentation process and increased the lightning contrast to eliminate the darkness of each image. After this process Edge Detection (ED) has provided some advantages for feature extraction. Identifying key elements in images, including object borders or forms, is essential for tasks like object identification and image segmentation. The dimensionality of the data may be reduced with the aid of Edge Detection (ED), which in turn makes future processing activities more efficient.

In comparison to other activation functions, the use of sigmoid activation functions for blood cell image segmentation carries with it several benefits as well as some downsides. If the output has to be registered as a binary classification, sigmoid functions help by squashing it between 0 and 1. On the other hand, Because of their smoothness and continuous differentiability across their whole range, sigmoid functions are ideal for optimization techniques that rely on gradients, such as gradient descent. The biggest disadvantage of sigmoid is that the vanishing gradient issue is very common in Deep Neural Networks (DNN) that have several layers, and it is particularly harmful to sigmoid functions. Alternative activation functions, such as Rectified Linear Unit (ReLU) and its variants, such as Leaky ReLU and Parametric ReLU, are

often used in deep learning for image segmentation. This is because these functions result in faster convergence, eliminate the problem of vanishing gradients, and improve the overall performance of deeper neural networks.

II. Literature Review

This study presents a new and simple ALL-image segmentation method based on color image modification [16]. A set threshold was used to partition the preprocessed pictures. Next, numerous metrics were used to evaluate the suggested method's output. Finally, the suggested technique was compared with existing standards. The suggested method outperforms previous techniques in terms of accuracy, specificity, sensitivity, and time. Furthermore, the data suggests that the proposed method greatly increases performance metrics.

In this study, WBC segmentation uses color space analysis and Otsu thresholding to separate WBC regions [17]. Nevertheless, noise in segmented images must be removed, and this is done using a morphological filter with connected Component Labeling (CCL). This watershed approach must be used to separate overlapping WBCs to retrieve the individual cells. The cells' data is then gathered via feature extraction and sent into the classifier. A Support Vector Machine (SVM) classifier is employed in this system to categorize lymphoblasts.

The image has been preprocessed by varying its brightness, contrast, and size to prepare it for segmentation. During the segmentation process, Otsu's thresholding and mathematical operators separate the nucleus region into segments [18]. Mathematical morphological techniques simplify the nucleus region for feature extraction in the post-processing phase. Finally, the segmented regions are split into ALL-affected and normal cells using the recommended customized K-nearest neighbor classification algorithm. The trials in this study included more than 80 images from the ALL-IDB2 dataset, and the results showed accuracy rates of 96.25%, 95% sensitivity, and 97% specificity.

The developed model for the detection and diagnosis of ALL is discussed in [19]. Shyalika *et al*. [19] suggested a segmentation model that uses supervised learning-based K-Nearest Neighbor (KNN) classification with morphological operations. The results demonstrate a high diagnosis accuracy of 88.8% for ALL using the recommended technique. The end product is a QT GUI development package written in Python that handles the majority of the planned backend functionality and a PHP-based internet application that assists hematologists, physicians, and patients in carrying out useful activities.

Shafique and Tehsin [20] suggested a segmentation strategy based on the basic threshold method. The model is trained by utilizing bone marrow images. The authors implemented convolutional neural networks and deep learning approaches to produce precise categorization outcomes. Consequently, experimental outcomes were generated and compared to those of SVM, K-Nearest Neighbor (KNN), and Naive Bayesian classifiers. The experimental results showed that the proposed method achieved 97.78% accuracy.

Segmenting images into each pixel is suggested in [21]. A deep joint segmentation model may provide more segmentation power and far higher detection accuracy. In this method, a Taylor political Monarch Butterfly Optimization (PMBO-enabled Taylor) Deep Residual Neural Network (DRN) is designed for ALL classification. The important components—texture, statistical, and grid data—are extracted to enhance the categorization procedure. The Taylor PMBO trains a Deep Residual Neural Network (DRN) classifier for the ALL detection and classification process. This model fared better than earlier techniques (0.9656, 0.9666, 0.9709, 0.9280, and 0.9321) when accuracy, sensitivity, specificity, F1–Score, and precision were evaluated.

Jawahar and Gandomi proposed the ALNet model, which is based on a deep neural network and a depth-wise convolution method with varying dilation rates for classifying images of white blood cells [22]. They compared the model's performance with other pre-trained models, namely VGG16, ResNet-50, Google Net, and Alex Net. The evaluation was based on different metrics, including precision, accuracy, loss accuracy, recall, F1–Score, and Receiver Operating Characteristic (ROC) curves. The experimental findings reveal that the ALNet model had the maximum accuracy in classification, reaching 91.13%. Additionally, it obtained an F1-Score of 0.96 while exhibiting reduced computational complexity.

Talaat *et al*. [23] proposed a model for segmenting WCs by applying histogram equalization and threshold estimation using the Zack method. The researchers derived a set of properties from the segmented cells, encompassing color, shape, texture, and hybrid characteristics. They employed the social spider optimization algorithm to determine the optimal features effectively. In addition, they employed a variety of classifiers to assess the efficacy of the presented methods using the ALL-IDB2 dataset. The suggested model demonstrated notable outcomes, achieving a classification accuracy of 95.23%.

A novel Bayesian-based optimized CNN is introduced for the identification of ALL in microscopic images in [24]. To enhance the classification performance, the architecture of this study's CNN is tailored to accommodate input data using the Bayesian optimization approach for hyperparameter tuning. The Bayesian optimization strategy utilizes an educated iterative approach to explore the hyperparameter space in pursuit of the optimal configuration of network hyperparameters that minimizes an objective error function. The CNN under consideration is trained and validated using a hybrid dataset. This hybrid dataset is created by merging two Publicly Available (ALL) datasets. The findings of this study reveal the superiority of the proposed Bayesian-optimized CNN over other optimized deep learning ALL classification models.

The Deep Convolutional Neural Network (DCNN) model is elaborately designed with several layers, diverse filter sizes, and a minimized number of filters and parameters to improve its efficacy and efficiency. The proposed model is called the Classification of Skin Lesions Network (CSLNet) which utilizes 68 convolutional

layers [25]. CSLNet was used to classify skin lesions using dermoscopic images. The experiment was conducted using ISIC-17, ISIC-18, and ISIC-19 datasets and benchmarked against the latest techniques. A specific Deep Convolutional Neural Network (DCNN) architecture is suggested for accurate detection of human gastrointestinal disorders using endoscopic images [26]. The DCNN architecture is designed with many routes, each examining the same endoscopic image at different resolutions. The proposed architecture is limited to two routes because of computational constraints, despite its potential to accommodate more routes and resolutions.

Hybrid Transfer Learning Extreme Gradient Boosting (HTL-XGB) is presented in [27]. It leverages transfer learning and incorporates advanced CNNs to extract features effectively. The algorithm also utilizes the classification and detection capabilities of extreme gradient boosting for ALL identification. In addition, the article proposes an algorithm for detecting activated and non-activated cells in complete blood smear images. The proposed methodology utilizes image processing techniques in conjunction with the HTL-XGB architecture. A series of rigorous experiments on various cutting-edge Convolutional Neural Networks (CNNs) found that Xception achieved the highest level of accuracy, reaching 89.7%.

III. Materials and Methods

The spatial interactions of the input feature maps are the primary focus of the position attention mechanism. DenseNet architecture is based on an encoder-decoder structure as the basic architecture and dense convolutional connections. Through the use of the DenseNet structure, the parameters of the model may be reduced, which in turn speeds up the training process. At the same time as the convolution operation extracts the low-level characteristics of medical pictures, the feature information is communicated to the position attention and channel attention mechanism modules.

The following are the main developments described in this manuscript:

- The present study uses the DenseNet 169 module as the principal network architecture to collect feature information from lymphoblastic leukemia images to improve the model's effectiveness and reduce the number of parameters. The DenseNet 169 module can potentially improve medical image analysis efficiency, segmentation performance, and precision.
- In the process, the image's fundamental visual features are extracted and passed on to the position attention and flipped attention mechanism modules. A more comprehensive feature map may be generated by joining the tensors produced by these modules. The network's architecture is reminiscent of the classic Convolutional Neural Network (U-Net) network structure previously used for segmentation.

The decoder module performs numerous actions, including up-sampling and image transposition, using the received image feature information. Applying the sigmoid function produces the desired segmentation result. Low-level image features are extracted from all images using the convolution operation as part of transferring feature information to the position attention and flipped attention modules. Convolutional layers are initially used to extract the significant visual aspects of the input image. Following that, the aforementioned features are transformed or encoded into a spatial grid, where each grid position represents a unique area or location within the image.

The location attention module can efficiently focus on important areas within an image for further processing or analysis using an attention matrix. It makes it easier for models to focus on essential visual inputs while ignoring unimportant or less critical areas. However, the suggested flipped attention module successfully captures and characterizes the feature information throughout the entire dataset using a dual memory unit technique consisting of two Conv1D procedures that share parameters. This can improve information sharing between samples and help identify the properties of the entire dataset. As seen in Fig. 2, we suggest an Attention-Flipping Block (FAB) for the lightweight ALL image segmentation model.

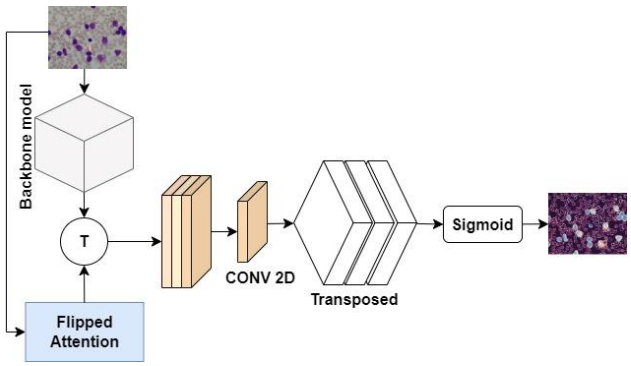

Fig. 2. The proposed segmentation networks.

The attention model is frequently used for a variety of jobs. The Squeeze-and-Excitation Networks (SEN) technique, as detailed in [28], has been proven to increase a network's representational capacity by including attention mechanisms that simulate channel-wise interactions. According to the attention model's issue statement, as the feature map grows, memory and computing overheads increase by a factor of four. The idea of crisscross attention was suggested by Hu *et al.* [29] to reduce computational and memory costs. For this method to capture the overall context successfully, row and column attention must be considered sequentially.

Each layer in a DenseNet design is connected to every other layer in a feed-forward fashion. This connectivity arrangement makes it possible to reuse features and encourages gradients to spread throughout the network, which mitigates the problem of vanishing gradients. Fig. 3 shows a dense unit is made up of many layers, each of which receives the feature maps from all of the unit's antecedent layers as inputs.

After dense blocks, transition layers, which include convolutional and pooling layers to reduce the spatial dimensionality of the feature maps, are frequently used. Observable characteristics of DenseNet [30] include parameter and processing efficiency. The network's dense

connections make it easier to reuse features gathered at different depths, encouraging improved feature reuse. Additionally, by offering shorter pathways for gradient propagation, DenseNet's dense connectedness helps to alleviate the vanishing gradient problem.

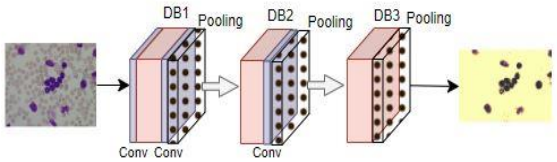

Fig. 3. Backbone of DenseNet.

DenseNets resolve common degradation problems and training breaks by providing direct connections between all succeeding layers as shown in Eq. (1).

$$x_A = H_A([X_0, ..., X_{A-1}]) \tag{1}$$

The Ath layer receives as input the feature maps of all preceding layers, $[X_0, \ldots, X_{A-1}]$:

Concatenation of the feature maps created in layers is referred to as X_0. Compared to other network topologies, densely connected layers show fewer output dimensions. This trait can successfully lessen the danger of overlearning features and, as a result, cut down on the computing time needed for training. The ALL-training data were scaled to meet the input layer's required size of 256×256×3. Fig. 3 describes the proposed network of segmentation and integrated flip attention block.

A novel attention mechanism is provided in [31], referred to as External Attention (EA). This mechanism uses two memory units, which are two Conv1D algorithms that exchange parameters, to describe the feature information of the whole dataset. The purpose of this study is to provide a flipped attention block. The block diagram used memory unit was the same model of the inverted External Attention Block (IEA) [32], but we expanded the reshaped layers for two branches. Before the conv operation, all information is with a memory unit.

For small and straightforward datasets, using just one occurrence of down- and up-sampling may be sufficient to get satisfactory segmentation results. Multiple up- and down-sampling strategies must be used to extract semantic feature maps for various regions from huge and complex datasets due to the difficulty in obtaining complete global information from a network on a small scale. Zhou *et al.* [33] proposed a change to the skip connections to overcome the problems.

Let F_0 be a matrix with the dimensions F0 CHW that represents the local picture features. Let F_1 and F_2 also be two new local features produced by convolution, where F1 and F_2 are both matrices with dimensions C, H, and W. The most recently created local features, F_1 and F_2, are molded into a format labeled CN, where N is equal to the product of the image's height (H) and width (W) or the total number of pixel points. Before being fed into the SoftMax layer, the local features F_1 and F_2 are treated to a matrix multiplication process. The location attention features are produced by this process and can be illustrated in Eq. (2).

$$P_{ij} = \frac{e^{F_{1i} \cdot F_{2j}}}{\sum_{i=1}^{n} e^{F_{1i} \cdot F_{2j}}} \tag{2}$$

The influence of the location of pixel points i on position j is denoted by the variable P_{ij}. While, F_{1j} and F_{2j} represent the two novel feature mappings generated by the convolution layer from the local features as shown in Eq. (3). When two-pixel spots' features resemble one another, there is likely a strong relationship between them. After performing the convolution operation on the local feature F_0, a feature map F_3 is generated. This feature map belongs to the tensor space CHW.

$$E_j = \theta_1 \sum_i^n P_{ij} . F_{3i} + F_{1j} \tag{3}$$

Subsequently, the feature map undergoes a conversion process to become the CN tensor. After obtaining the new tensor, both the position attention feature S and the new feature map D undergo a transposition operation, followed by matrix multiplication. Pixels representing the same object or having similar properties are assigned to a common category as part of classifying pixels with similar semantic attributes into a single class. The neural network can recognize and distinguish different items or areas inside a picture thanks to this categorization.

The ultimate result is generated by executing a summing operation between the position attention map and the unique input image feature map F_1. This allows for the completion of the process. A random number generator provides F_1's starting values. A contextual link exists between the location of the output feature information E and how it is displayed. During the functioning of the network, pixels that share semantic qualities may be grouped into a single class. On the other hand, pixels with separate semantic features can be assigned to different classes depending on how they relate to the context.

It is generally accepted that an image's pixels have varied degrees of value concerning the picture as a whole. Because of this, each pixel is given a unique weight value throughout the image segmentation process to outline the image precisely. The suggested method efficiently captures and characterizes the feature information over the whole dataset using a dual memory unit consisting of two Conv1D procedures that share parameters.

The size of the data is changed to $X \in L^{(CH \times W)}$ with the use of convolution 1×1 and reshaping techniques. In the structure the features text ruinous information, two flipping layers are added. In Memory Unit 1, the feature map FM is multiplied by four to get the result of attention $\in L^{(4CH \times W)}$. A reshaping process then recovers the original feature map dimensions after Memory Unit 2 restores the dimensionality. The convolution process will distribute features on two layers to increase the transfer speed.

One benefit of the memory unit is its ability to transform the input into a space with higher dimensions. This results in a more comprehensive representation of the dataset's feature information within the memory unit. Patterns, connections, and data changes of a more complex kind may be captured in the higher-dimensional space. Next, the distributed data gathered by the flipping layers and the resultant feature map can be acquired using a 1×1 convolution and subsequently combined with the residual data. The output corresponds to the processed feature representation of the input. As a function of the input, the output is complicated, and the network's goal during training is to learn this mapping. The input and learned mapping are combined to produce the output in a residual block.

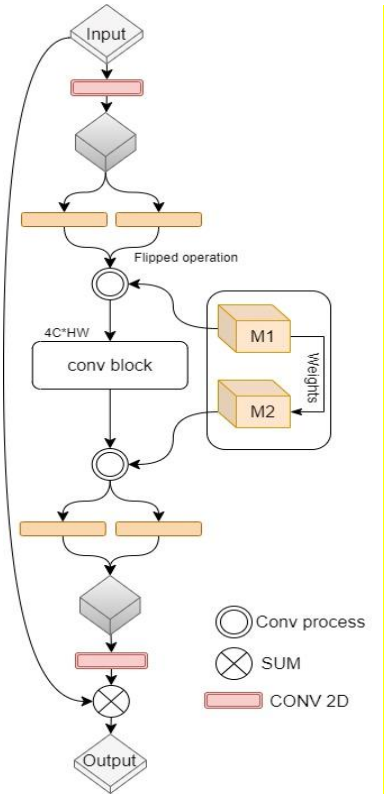

Fig. 4. The mechanism of the flipped attention block structure.

If the memory unit can learn to encode important information more generically and abstractly, it will be able to identify constant patterns across all examples in the dataset. Hence, the generalizability of a model may be enhanced if its memory units are mapped to a higher-dimensional space. Fig. 4 illustrates the mechanism of the flipped attention block structure with the memory unit.

However, the final output Feature Map (FM) will be obtained through the last layer with a 1×1 convolution, and it will directly add to the residual information skip connection. However, the number of channels will not be fixed as much as we set the processed four-time expansion. The biggest advantage of ALL-FABNET is that uses a Memory Unit (MU) which is responsible for mapping all the input features to highly dimensional space. This will be able to describe all features.

The purpose of this study is to provide a flipped attention block. The block stracture for memory unit was the same model of the inverted External Attention Block (IEA), but we expanded the reshaped layers for two branches. Before the conv operation, all information is with a memory unit. Finally, for future techniqe we will excute weakly supervised learning methods [34] that was achieved more accurate results.

IV. Result and Discussion

We trained the model on three lymphoblastic leukemia datasets: ALL_IDB1, ALL-IDB2, and C-NMC 2019-Leukemia. This section describes the performance evaluation of the proposed ALL-FABNET. The preprocessing step is the main factor in achieving good accuracy. For this reason, we enhanced the training set data to raise the sample size. We used procedures like rotating, cropping, color transformation, and flipping as part of our data augmentation process. Assigning varied weight information to the model's input image based on location and channel is important for image segmentation since distinct pixels in different places have different responsibilities for the total image. The HarDNet model is the backbone. Although HarDNet is efficient, its architecture and unique harmonic kernels could make it difficult to comprehend and use. Compared to more straightforward designs, users may have to put in more effort to grasp the model's complexities. The basic encoder-decoder of our model is DenseNet. Reusing features improves the efficiency of parameters and computations and helps with the issue of disappearing gradients. Feature reuse also has the potential to enhance the network's information flow.

To evaluate our trained models' segmentation and classification results, we utilized evaluation metrics, including accuracy, precision, recall, and F1-Score.

Listed below are the terms used to describe how metrics are calculated.

True Positive (TP) denotes an accurately detected lesion.

False Positive (FP) denotes an incorrect prediction of a lesion.

True Negative (TN) denotes an accurately detected background.

False Negative (FN) denotes a pixel incorrectly assigned as a background pixel.

Accuracy illustrates the proportion of predictions that were accurate as shown in Eq. (4).

$$Accuracy=\frac{TP}{TP+TN+FP+FN} \tag{4}$$

Recall quantifies the fraction of accurately detected anticipated as shown in Eq. (5).

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Precision is determined by dividing the number of actual positives by the number of predicted positives as shown in Eq. (6).

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

The Receiver Operating Characteristic (ROC) curve is a graph that shows how well a classification model works at all thresholds for classification.

For evaluation, different baselines are implemented to analyze the significance of the proposed segmentation model. The findings of our model are collected from various in-depth experiments. First, the accuracy, precision, recall, and F-Score of the proposed ALL-FABNET are compared to the other pre-trained baseline models, as shown in Table I. It reveals that our model performs exceptionally well on the ALL-IDB1, ALL-IDB2, and C-NMC 2019 datasets [35].

TABLE I. EVALUATION OF THE PROPOSED SEGMENTATION MODEL WITH BASELINE MODELS

Model	Datasets	Accuracy (%)	Precision (%)	Recall (%)	F1–Score (%)
ResNet18	C-NMC 2019	81.01	86.70	84.65	81.60
DenseNet121	C-NMC 2019	94.53	93.15	92.75	95.95
ResNet50	C-NMC 2019	93.50	92.73	93.10	94.94
DenseNet121	ALL-IDB2	84.20	94.85	93.90	96.95
VGG19	ALL-IDB2	86.92	85.45	88.75	83.16
Proposed model	ALL-IDB1,108 images	91.34	93.54	90.67	94.69
Proposed model	ALL-IDB2,260 images	93.56	90.63	89.54	93.64
Proposed model	C-NMC 2019	96.67	94.06	91.27	95.54

It is observed that the proposed segmentation model outperforms other trained models in accuracy, precision, recall, and F1-Score when utilizing the C-NMC 2019 dataset. The ALL-FABNET model achieves 96.67%, 94.06%, 91.27%, and 95.54% in terms of accuracy, precision, recall, and F1–Score, respectively. The trade-off is that the proposed model has a greater number of parameters and model size than the baseline model. It is obvious from the results that the proposed model performs better than other trained models. In image segmentation, the Receiver Operating Characteristic (ROC) curve is a graphical representation of the performance of a binary classifier, as the threshold for classifying a pixel as foreground or background varies. The ROC curve plots the true Positive Rate (TPR) against the False Positive Rate (FPR) at different threshold settings. The area under the ROC Curve (AUC) measures the classifier's overall performance and compares the performance of different classifiers. The ROC curve for the segmentation model is shown in Fig. 5 It shows the robustness of our segmentation in term of accuracy.

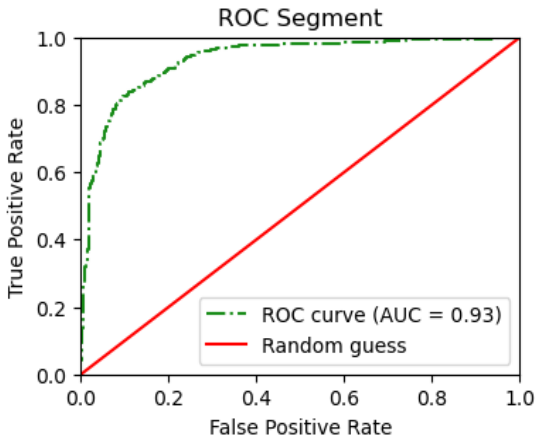

Fig. 5. Accuracy of the ALL-FABNET segmentation model.

In addition, our segmentation model gives a good result in terms of precision and recall as shown in Fig. 6.

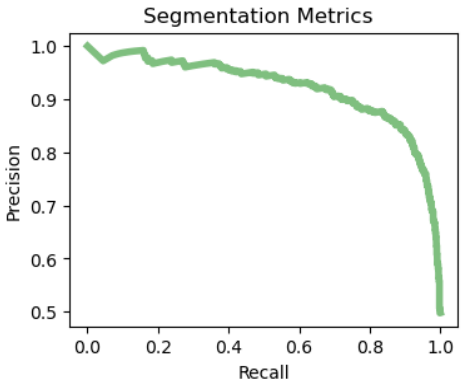

Fig. 6. Precision and recall of the segmentation model.

The performance is measured by the rate of positive (absent) classifications that the model has correctly identified. It is how many of the positive items were predicted correctly out of the number of such items in the data, while the precision accuracy of such a model is under consideration. It is a number where the balance is between the correct positive predictions made by the model to the overall number of positive predictions.

However, this section assesses the efficacy of the ALL-FABNET classifier for ALL classification. This study utilizes multiple baseline model classifiers and evaluation matrices to evaluate and compare the performance of the proposed ALL-FABNET model and baseline classifiers in comparison with ResNet50, ResNet18, DensNet-121, and VGG16, as shown in Table II.

TABLE II. EVALUATION OF THE PROPOSED CLASSIFICATION MODEL WITH BASELINE MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet50	89.46	78.55	90.84	79.64
DenseNet121	92.38	90.45	91.82	93.80
DenseNet169	**96.667**	**94.667**	**95.667**	**96.667**

Table II illustrates the best results in bold for our comparative training. Three different datasets were

utilized for the study to emphasize the best classification of lesions.

The ROC curve in Fig. 7 shows the effectiveness of our model, in which the suggested classifier increased the classification accuracy to 97.94%. In addition, it significantly improved the detection rate as compared to the baseline classifiers in terms of precision (97.54%), as shown in Fig. 8 ALL-FABNET also achieved 98.65% in terms of recall and a maximum F1-Score of 97.65%. The positive discovery elucidates the efficacy of our segmentation approach, which gives us the robustness of our classification model.

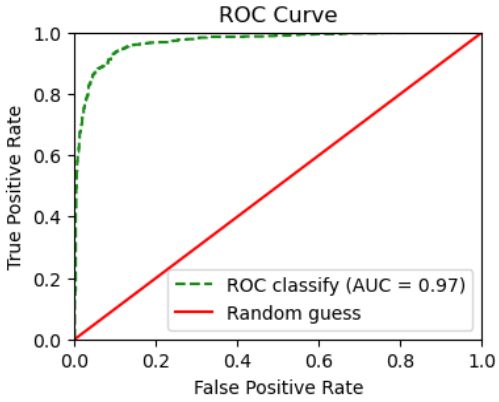

Fig. 7. ROC of the ALL-FABNET classification model.

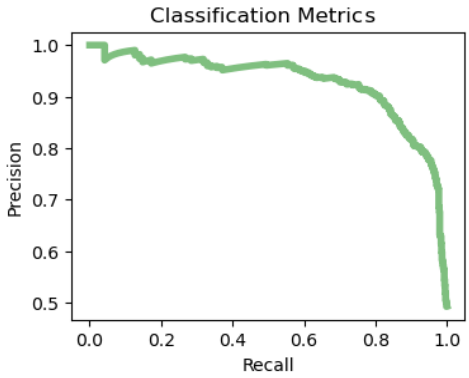

Fig. 8. Precision and recall of the classification model.

Furthermore, during our comparison study, we evaluated ALL-FABNET's approach regarding Unet and Unet++ by utilizing the ALL-IDB2 dataset. ALLFABNET achieved a segmentation performance of 93.56% and a classification performance of 97.94%. Classification accuracy reached a high level in the 50th epoch. Fig. 9 illustrates the comparative accuracy of Unet, Unet++, and ALL-FABNET. Fig. 10 displays a comparative analysis of the losses across all models.

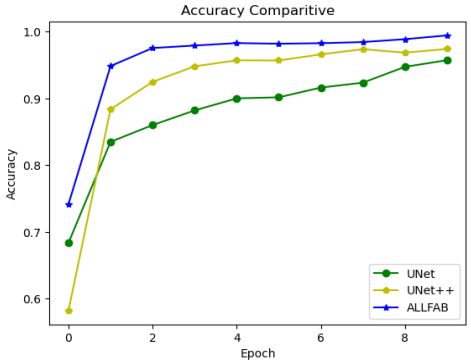

Fig. 9. Accuracy comparison for three models.

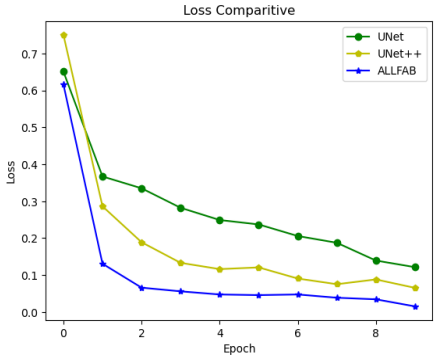

Fig. 10. Loss comparison for three models.

We train the backbone model without Appling FAB block attention to compare with other training models' results as registered in Table III.

The table shows the results comparative between the three models without using the FAB block. We train the models with transfer learning techniques.

The initial image input in the first row is shown in Fig. 11. To remove extraneous features, the second row illustrates the procedure for the edge detection approach. ALL-FABNET Segmenting is the third-row image for multi-cell blood microscopic studies. For the first row (a), Fig. 12 shows the original image input. To remove extraneous features, the second row displays the procedure for the edge detection approach. Class III ALL-FABNET segmentation for imaging of individual blood cells.

TABLE III. EVALUATION OF THE PROPOSED SEGMENTATION MODEL WITHOUT USING FAB BLOCK MODELS

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1–Score (%)
ResNet50	C-NMC 2019	90.85	90.53	90.53	95.15
DenseNet121	C-NMC 2019	90.34	89.60	88.49	94.65
ResNet50	ALL-IDB2	89.46	78.55	91.84	79.64
DenseNet121	ALL-IDB2	92.38	90.45	91.82	93.80
ResNet 18	ISBI	91.78	85.83	87.68	92.64
VGG16	ISBI	86.92	95.67	90.66	93.50
Proposed model	ALL-IDB1, 108 image	95.73	93.65	94.62	95.23
Proposed model	C-NMC 2019	97.50	95.32	96.76	97.03
Proposed model	ALL-IDB2 260 images	97.94	96.54	98.65	97.65

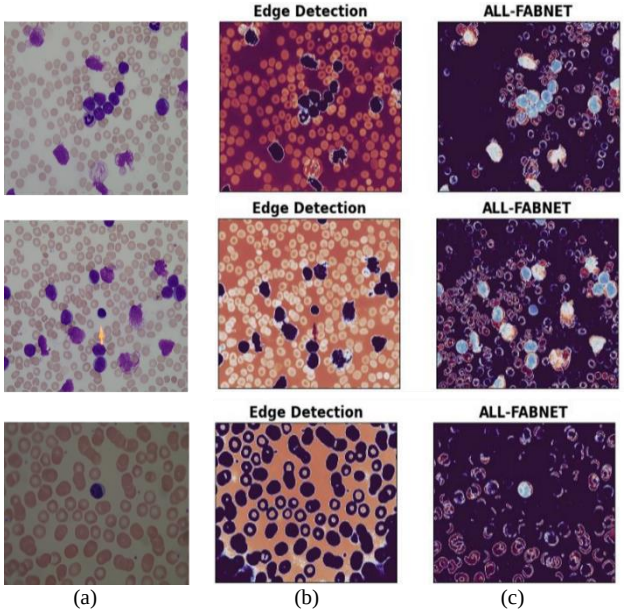

Fig. 11. Comparative results of edge detection and ALL-FABNET for ALL_IDB1; (a) original input image, (b) Edge detection process, (c) ALL-FABNET segmentation for multi-cell blood microscopic images.

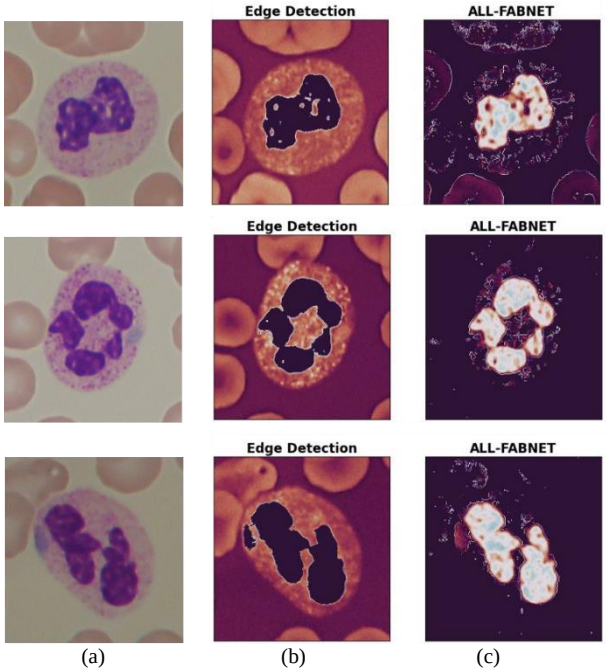

Fig. 12. Comparative results of edge detection and ALL-FABNET for ALL_IDB2; (a) Original input image (b) Edge detection process, (c) ALL-FABNET segmentation for single-cell blood microscopic image.

V. Conclucion and Feature Work

Our model used the DenseNet 169 network as the backbone structure, and medical image feature information was extracted using four null space convolutional pooling pyramids. The model's parameter count can be drastically reduced by using a structured method, which can also speed up the training process. The suggested model incorporates multiscale semantic information at each level, enabling the preservation of more detailed feature maps in the decoder block, unlike U-Net. The modules in charge of position and floated attention mechanisms receive the feature information. A module called the dual-path flipped attention mechanism selectively amplifies important features while suppressing unimportant or distracting ones. However, the flipped attention block structure uses a memory unit to increase the efficiency of the network. The Dense Net 169 structure may significantly decrease the number of parameters in the model and enhance the speed of model training. A sequence of transpose convolution and up-sampling techniques are used to facilitate concatenating the two kinds of features. The method was evaluated and examined using many medical imaging datasets, including the ALL_IDB1, ALL-IDB2, and C-NMC 2019-Leukemia datasets.

The strengths and weaknesses of ALLFAB are that memory units have improved representational capacity because they can map incoming data to a higher dimensional space, which allows them to store and understand data with more complex relationships and patterns. The rationale for this is that the ability to encode intricate traits and variations in the dataset is enhanced with more dimensions. Higher dimensional representations may be quite beneficial for accurately distinguishing the various groups or categories included in the dataset. This strategy is highly advantageous, especially in classification tasks when the goal is to distinguish between many groups based on the provided attributes. However, Transforming the input data into a higher-dimensional space allows for non-linear transformations. The importance of this rests in the fact that several real-world datasets show non-linear relationships between their attributes and that more dimensions allow for greater flexibility in capturing these non-linearities. The higher dimensional representation has the potential to enhance performance by prioritizing essential characteristics and disregarding irrelevant ones. By focusing on the most distinctive characteristics of the data, this approach for extracting features may simplify the learning process.

The amount of memory and processing resources required to increase when mapping information to a higher dimensional domain. This leads to an increase in the duration of the model's training and inference, which will ultimately lead to a decrease in its efficiency. This is especially likely to occur in situations when time or resources are limited. The availability of data is diminishing, which poses a greater difficulty in effectively modeling and extracting insights from it. In high-dimensional spaces, the data points tend to be more widely spread out, which requires a larger amount of training data to adequately cover the whole space. The likelihood of overfitting increases when the input is mapped to a higher-dimensional space since the model is then better able to replicate the training data. When a model learns and uses irrelevant or unnecessary patterns or noise from its training data, a phenomenon known as overfitting occurs, and the model's ability to predict new data is impaired. Moreover, the model in this phase gets more difficult to comprehend and make sense of higher dimensional representations than lower-level ones. Utilizing higher-dimensional spaces for

storing representations in large-scale applications with substantial datasets may need more memory and storage resources. For future research, Using Grad-CAM for lesion localization in weakly supervised learning training seems to be an effective approach. Grad-CAM is a method that highlights and shows the crucial regions in an image that impact the model's decision, particularly useful in medical imaging tasks like lesion detection.

Secondly, Comparing the results using the YOLO algorithm is a great approach to evaluating the effectiveness of our model. YOLO is well-known for its real-time object detection capabilities and has been widely used in several applications, including medical imaging. We will assess our model's accuracy, speed, and suitability for lesion detection tasks by comparing it with YOLO.

Future research will focus on creating novel algorithms to enhance the segmentation task, testing the proposed approach with increased computational resources, conducting experiments on higher resolution and larger datasets, and adapting the model for use in other imaging fields.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Ammar S. Al-Zubaidi, Mohammed Al-Mukhtar study concepts study design and analysis with interpretation; Mohammed Al-Mukhtar, provide diagrams and literature results of SOTA. Ammar Awni Abbas experimental studies, statistical analysis and manuscript editing. All authors had approved the final version.

REFERENCES

[1] P. K. Das, S. Meher, R. Panda, and A. Abraham, "An efficient blood-cell segmentation for the detection of hematological disorders," *IEEE Trans Cybern*, vol. 52, no. 10, pp. 10615–10626, Oct. 2022. doi: 10.1109/TCYB.2021.3062152

[2] H. M. Jumaah, J. H. Yenzeel, and M. G. Mehdi, "Evaluation of some biochemical parameters and hormones in patients with acute myeloid leukemia in Iraq," *Iraqi Journal of Science*, vol. 62, no. 5, pp. 1460–1466, May 2021. doi: 10.24996/ijs.2021.62.5.9

[3] J. M. Bennett *et al.*, "Proposals for the classification of the acute leukaemias french-american-british (FAB) co-operative group," *Br J. Haematol*, vol. 33, no. 4, 1976. doi: 10.1111/j.1365-2141. 1976. tb03563.x

[4] A. Rehman, N. Abbas, T. Saba, S. Ijaz Ur Rahman, Z. Mehmood, and H. Kolivand. Classification of acute lymphoblastic leukemia using deep learning. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30351463/

[5] P. K. Das, V. A. Diya, S. Meher, R. Panda, and A. Abraham, "A systematic review on recent advancements in deep and machine learning based detection and classification of acute Lymphoblastic leukemia," *IEEE Access*, vol. 10, pp. 81741–81763, 2022. doi: 10.1109/ACCESS.2022.3196037

[6] S. A. Enad and W. A. Al-Amili, "Investigation of secondary Acute Lymphoblastic Leukemia (sALL) among acute lymphoblastic leukemia (ALL) Iraqi patients," *Iraqi Journal of Science*, vol. 60, no. 2, pp. 223–227, Feb. 2019. doi: 10.24996/ijs.2019.60.2.3

[7] S. Alagu, N. A.Priyanka, G. Kavitha, and K. B. Bagan, "Automatic detection of acute lymphoblastic leukemia using UNET based segmentation and statistical analysis of fused deep features," *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 1952–1969, 2021. doi: 10.1080/08839514.2021.1995974

[8] A. Muntasa and M. Yusuf, "Modeling of the acute lymphoblastic leukemia detection based on the principal object characteristics of the color image," *Procedia Computer Science*, pp. 87–98, 2019. doi: 10.1016/j.procs.2019.08.145

[9] R. R. Maaliw *et al.*, "A multistage transfer learning approach for acute lymphoblastic leukemia classification," in *Proc. 2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 488–495. doi: 10.1109/UEMCON54665.2022.9965679

[10] G. J. M. Mahdi, B. A. Kalaf, and M. A. Khaleel, "Enhanced supervised principal component analysis for cancer classification," *Iraqi Journal of Science*, vol. 62, no. 4, pp. 1321–1333, Apr. 2021. doi: 10.24996/ijs.2021.62.4.28

[11] A. Saeed *et al.*, "A deep learning-based approach for the diagnosis of acute lymphoblastic leukemia," *Electronics*, vol. 11, no. 19, Oct. 2022. doi: 10.3390/electronics11193168

[12] M. Claro *et al.*, "Convolution neural network models for acute leukemia diagnosis," in *Proc. International Conference on Systems, Signals, and Image Processing*, IEEE Computer Society, Jul. 2020, pp. 63–68. doi: 10.1109/IWSSIP48289.2020.9145406

[13] L. T. Duong, P. T. Nguyen, C. D. Sipio, and D. D. Ruscio, "Automated fruit recognition using EfficientNet and mixnet," *Comput Electron Agric*, vol. 171, Apr. 2020. doi: 10.1016/j.compag.2020.105326

[14] N. Sampathila *et al.*, "Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images," *Healthcare (Switzerland)*, vol. 10, no. 10, Oct. 2022. doi: 10.3390/healthcare10101812

[15] ALL-IDB, IEEE DataPort. [Online]. Available: https://ieee-dataport.org/documents/all-idb-acute-lymphoblastic-leukaemia-international-database#files

[16] F. H. Najjar, K. T. Khudhair, Z. N. Khudhair, H. H. Alwan, and A. Al-Khaykan, "Acute lymphoblastic leukemia image segmentation based on modified HSV model," *Journal of Physics: Conference Series*, vol. 1, 2023. doi: 10.1088/1742-6596/2432/1/012020

[17] S. N. M. Safuan, M. R. M. Tomari, W. N. W. Zakaria, M. N. H. Mohd, and N. S. Suriani, "Computer aided system for lymphoblast classification to detectacute lymphoblastic leukemia," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 2, pp. 597–607, May 2019, doi: 10.11591/ijeecs.v14.i2.pp597-607.

[18] D. Umamaheswari and S. Geetha, "A framework for efficient recognition and classification of Acute Lymphoblastic Leukemia with a novel customized-KNN classifier," *Journal of* Computing *and Information Technology*, vol. 26, no. 2, pp. 131–140, 2018, doi: 10.20532/cit.2018.1004123.

[19] C. Shyalika, N. Pathirage, D. Kottahachchi, J. Shyalika, P. Kumara, and D. U. Kottahachchi, "Statistical morphological analysis based supervised classification algorithm for diagnosing acute lymphoblastic leukemia an image processing application," *Article in Journal of Theoretical and Applied Information Technology*, vol. 98, no. 18, 2020.

[20] S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technol Cancer Res Treat*, vol. 17, pp. 1–7, Jan. 2018. doi: 10.1177/1533033818802789

[21] G. P. Mercy Bai and V. Perumal, "Taylor political monarch butterfly optimization driven deep learning model for acute lymphoblastic leukemia detection and severity classification using blood smear images," *Concurr Comput*, vol. 35, no. 2, 2023. doi: 10.1002/cpe.7455

[22] M. Jawahar, S. H, J. A. L, and A. H. Gandomi, "ALNett: A cluster layer deep convolutional neural network for acute lymphoblastic leukemia classification," *Comput Biol Med*, vol. 148, 105894, 2022. doi: https://doi.org/10.1016/j.compbiomed.2022.105894

[23] A. Talaat, A. Abdeldaim, and A. E. Hassanien, "Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm," *Soft Comput.*, vol. 23, pp. 1–16, Aug. 2019. doi: 10.1007/s00500-018-3288-5

[24] G. Atteia, A. A. Alhussan, and N. A. Samee, "BO-ALLCNN: Bayesian-based optimized CNN for acute lymphoblastic leukemia detection in microscopic blood smear images," *Sensors*, vol. 22, no. 15, Aug. 2022. doi: 10.3390/s22155520

[25] I. Iqbal *et al.*, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic

images," *Computerized Medical Imaging and Graphics*, 101843, 2021.
[26] I. Iqbal *et al.*, "Automated identification of human gastrointestinal tract abnormalities based on deep convolutional neural network with endoscopic images," *Intelligent Systems with Applications*, 200149, 2022.
[27] A. J. Hau, N. Hameed, A. Walker, and M. M. Hasan, "A hybrid transfer learning and segmentation approach for the detection of acute lymphoblastic leukemia," Lecture Notes in Networks and Systems, 2023. doi: 10.1007/978-981-99-1916-1_14
[28] T. Shen et al., "HarDNet and dual-code attention mechanism-based model for medical images segmentation," *IEEE Access*, vol. 1, 2023.
[29] J. Hu, L. Shen, and G. Sun, "Queeze-and-excitation networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.
[30] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 4700–4708.
[31] M. H. Guo *et al.*, "Beyond self-attention: External attention using two linear layers for visual tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 5436–5447, 2022.
[32] J. Ruan *et al.*, "MALUNet: A multi-attention and light-weight unet for skin lesion segmentation," in *Proc. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022, pp. 1150–1156.
[33] Z. Zhou *et al.*, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, pp. 1856–1867, 2019.
[34] M. A. Mukhtar, A. H. Morad, M. Albadri, and M. S. Islam, "Weakly supervised sensitive heatmap framework to classify and localize diabetic retinopathy lesions," *Scientific Reports*, vol. 11, no. 1, 23631, 2021.
[35] A. Gupta and R. Gupta. (2019). All challenge dataset of ISBI 2019 [Online]. Available: https://doi.org/10.7937/tcia.2019.dc64i46r