

A deep learning approach in detection of malaria and acute lymphoblastic leukemia diseases utilising blood smear microscopic images

Quyen Hoang Vo^{1,2}, Xuan-Hieu Le^{1,2}, Thanh-Hai Le^{2,3}, Thi-Thu-Hien Pham^{1,2*}

¹School of Biomedical Engineering, International University, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

²Vietnam National University - Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

³Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, 268 Ly Thuong Kiet Street, Ward 14, District 10, Ho Chi Minh City, Vietnam

Received 30 December 2021; revised 14 February 2022; accepted 23 February 2022

Abstract:

The numerous rising infections and deaths of malaria and acute lymphoblastic leukaemia (ALL) highlights the urgent need for early, useful, and efficient diagnosis methods. Recently, the framework of artificial intelligence has been applied to minimize time-consuming tasks, to increase the accuracy and flexibility of clinical diagnoses, and to reduce the pressure on physicians, diagnosticians, and clinical experts. In this study, a detection system for malaria and ALL is proposed that utilizes blood smear microscopic images with the aid of deep learning algorithms to identify and classify these two diseases automatically. The blood smear microscopic images consist of 1503 ALL images, 891 malaria images, and 1503 normal images that were divided into a training, validation, and testing sets in ratios of 50, 25, and 25%, respectively. The proposed model was built into three stages including the first stage for segmentation-applied modified UNet pre-trained model, the second stage for classification based on the convolution neural network model, and the final stage for classification utilizing perceptron as the combining model. As a result, the proposed system provides an alternative and interpretable method to detect abnormal leukocytes for ALL and malaria-infected blood cells with a 93% overall accuracy including the detection rate for ALL of 95% and the detection rate for malaria of 92%.

Keywords: acute lymphoblastic leukaemia, blood smear microscopic image, deep learning, malaria.

Classification number: 3.6

1. Introduction

Malaria and ALL are two of the most dangerous blood diseases known today. According to the published data of the World Health Organization (WHO), there were an anticipated 241 million cases of malaria worldwide in 2020, with 627,000 deaths from this disease [1]. Whereas, ALL is a leading type of blood cancer that could grow rapidly to a deadly condition within weeks. Consequently, ALL contributed to the cause of 111,000 deaths globally and seriously impacted the lives of at least 876,000 patients in 2015 [2].

Malaria is a life-threatening disease that occurs due to the invasion of red blood cells (RBCs) caused by plasmodium parasites through the bites of infected mosquitoes. Consequently, the disease can be widely spread and affect the lives of millions of people in the world. Once bitten, the infection initially occurs in the

liver before the parasites re-enter the bloodstream and target RBCs. At this moment, RBCs be infected with plasmodium become easily “sticky”, and when they go through the small blood vessels inside the organs, they get stuck. Since the number of stuck RBCs enlarges, blood volume flows to the organ is reduced, which causes further complications such as kidney failure, coma, etc. If the sequential process happens inside the blood vessels in the brain, the result is clinically recognized as malaria disease - complications can emerge including impaired consciousness, coma, and even death [3].

Leukaemia is a term to describe a group of blood cancers that begin in the bone marrow, resulting in an excessive amount of abnormal white blood cells (WBCs). Four types of leukaemia cancer are ALL, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML), and chronic lymphocytic leukaemia (CLL). Among them, ALL occurs most frequently and it is a dangerous

*Corresponding author: Email: ptthien@hcmiu.edu.vn

disease that occurs when there is an uncontrollable proliferation of lymphoid precursors (i.e., lymphoblasts) in the bone marrow with restrained maturation. These overproduced leukemic cells cannot function properly and even suppress other normal WBCs from forming. These abnormal cells are the cause of fatigue, anaemia, fever, and bone pain because of the spread of these cells into the bone and joint surfaces, which further causes recurrent infections. The leukemic cells can then diffuse out of the bone marrow moving in the bloodstream and accumulate in various organs such as the lymph nodes (or lymph glands), spleen, liver, and central nervous system (brain and spinal cord) [4]. Consequently, patients diagnosed with ALL could die from various threats such as infection of bacteria, fungus, or even excessive bleeding in organs [3]. If left untreated, ALL disease can be fatal within weeks.

Among the current methods of examination and diagnosis, visual microscopical examination of Giemsa-stained blood smears is the most extensively used approach for assessing the development stage of malaria and leukaemia diseases [5]. However, the quality of the reagents, the image resolution of microscopes, and the experience of laboratory physicians/specialists all play a role in diagnosis. Another method is a polymerase chain reaction (PCR), which is also used to detect parasite nucleic acids of malaria or leukaemia. Although the PCR method is slightly more sensitive than smear microscopy, it has some limitations in the diagnosis of patients under restricted health care conditions [6]. Besides, a complete blood count (CBC) test is also commonly used to diagnose ALL. However, the sensitivity of this method is quite low because it depends on the number of white blood cells, for example, when the white blood cell count is not large enough and the symptoms may not be sufficiently obvious for the clinician to be able to confirm whether the patient has leukaemia [7]. In summary, these manual approaches may sometimes be time-consuming and expensive because all of the methods of detecting malaria and leukaemia are manual and completely rely on the expertise and knowledge of trained medical doctors [8].

Recently, blood smear microscopic image analysis has allowed pathologists to examine thin blood smear samples under optical microscopes to identify the presence of infected red blood cells for malaria cases or cancerous lymphocytes, which indicate the existence of malaria or ALL [9, 10]. It must be said that the analysis of

smear films remains a challenge for pathologists not only due to the time-consuming assessment but also because of the requirement for specialized knowledge of image evaluation and for the massive workload that they must go through every day [10]. Obviously, improvements in image analysis, especially the applications of artificial intelligence (AI), are becoming increasingly popular and studied more widely. These advanced methods have the potential of automatic detecting processes that encourage faster and more precise diagnosis in the clinical microbiology field for infectious diseases [11]. Computer-aided diagnosis becomes more powerful and reliable when machine learning/deep learning approaches are combined with image processing techniques. The benefit of the machine learning/deep learning methods is to reduce human interaction while also improving the quality of the diagnostic results. Deep learning, a subset of machine learning, provides a better end-to-end approach by automatically extracting relevant features and learning how to employ those features in performing given tasks on their own in an incremental manner. Deep learning deals with studies that require more precision, more mathematics, and more computation, which are being used in various fields to improve the reliability of automated systems. Thus, biomedical image analysis with deep learning techniques has become a popular area in recent years as it eliminates the need for specialized expertise and complicated manual feature extraction [12].

Among the available studies on leukaemia detection, one of the most impressive is the work of T.T.P. Thanh, et al. (2018) [13]. Thanh's group investigated the use of CNN-based models to detect leukaemia. Although having an impressive accuracy of 96.6%, the stability of the model and the results of this work should take into account the training set contained only 108 original images. However, the authors suggested data augmentation methods such as rotation and flip/flop could overcome the fact of insufficiency of the training set data in [13]. In other studies, P. Jagadev, et al. (2017) [14] and L. Putzu, et al. (2014) [15] used traditional computer vision techniques combined with deep learning algorithms to detect leukaemia. P. Jagadev, et al. (2017) [14] enhanced the image content by applying different traditional methods including Otsu's algorithm and Kmeans clustering to segment the potential WBCs. However, the best performance of the ensemble model achieved only 90% precision and 89% recall. L. Putzu, et al. (2014) [15] proposed a mathematical morphology approach of mathematical expressions of sample features

such as shape, bending, and roundness ratio. Then, chain codes are extracted and fed to a support vector machine (SVM) classifier. As a result, their work reached a recall of 93%, which is promising.

For malaria detection, J. Hung, et al. (2017) [16] presented the Faster R-CNN model and pre-trained AlexNet model to detect each RBC in the microscopic blood smear images. Utilizing both transfer learning and object detection methods, their research reached an accuracy of 96% compared to the one-stage classifier. Another work, by N.E. Ross, et al. (2006) [17], proposed the identification of infected malaria cells by segmentation of RBCs utilizing thresholding techniques. They extracted the sample features by feeding images to an SVM classifier. The results achieved an accuracy of 81-85%. Later, M. Poostchi, et al. (2018) [18] presented a method that blended a multiscale Laplacian of Gaussian filter and an approximate centroid for the marker-controlled watershed to detect and segment RBCs individually. They applied a voting-based detection system that reached a precision of 88%. Whereas, based only on manual extraction features, M. Sheikhhosseini, et al. (2013) [19] applied a rule-based system that reached a considerable sensitivity (recall of disease class) of 82.88%.

In the proposed study, a deep learning model is applied to analyse microscopic images of blood smear to determine the existence of ALL and malaria with the aid of deep learning. Firstly, a convolutional neural network (CNN) is trained to segment the components of blood smear images, which are WBCs and RBCs. The dataset of blood smear microscopic images were collected from several different sources that are validated by scientific research community. Then, for malaria detection, each segmented RBC component is fed to another trained network to determine whether the RBC is healthy or has a high-risk of being infected with malaria. Whereas for ALL detection, the same process is applied to WBCs to examine if they are normal or leukemic. As a result, the evaluation metrics of the achieved results show that this proposed study is promising, highly accurate, and can be extended further into practical application.

2. Materials and methods

Three major tasks of a deep learning model are detection, classification, and segmentation. Detection is concentrated on detecting objects in an image and marking them with a rectangle around the object, for example, a person or an animal. On the other hand,

classification is defined by categorizing the whole image into a predefined class such as “people”, “animals”, or “furniture”. Meanwhile, segmentation deals with the association of each pixel in an image with a class label, such as WBC, RBC, background, etc. It should be noted that segmentation models provide an exact contour of the classified object in the image meanwhile classification models identify what is in the image, and the detection models place a bounding box around a specific object. To improve the accuracy and usefulness of this study, both segmentation and classification models were applied in this approach.

2.1. Loss functions

Loss function is an important part of artificial neural networks, which is used to measure the inconsistency between the predicted value and actual value. The loss function has a non-negative value, which means that the robustness of the model increases along with a decrease of the value of the loss function. Besides accuracy, the loss function can also be used to evaluate whether a model is overfitting or underfitting. Even though there are several available loss functions, each function is only suitable for a particular task. In the scope of the proposed study, 2 loss functions are explained, which are soft Dice and categorical cross-entropy [20].

Categorical cross entropy (CCE) is the most fundamental loss function and it is commonly used in classification tasks. The cross entropy can be formulated as follows [20]: $CE = -\sum_i^C t_i \log(s_i)$ (1)

where t_i and s_i are the ground truth and the CNN score from class i to class C , respectively. For example, given $C=2$ (where a classification only has two classes, i.e., positive and negative), Eq. (1) becomes [20]:

$$CE = -\sum_{i=1}^{C=2} t_i \log(s_i) = -t_1 \log(s_1) - (1-t_1) \log(1-s_1) \quad (2)$$

where $t_2=1-t_1$ and $s_2=1-s_1$ are the ground truth and the score for C_2 . Equation (2) is often called binary cross entropy (BCE). Similarly, CCE only differs in the way of defining scores. The CCE requires the scores to be probabilities of one class on C classes, that is then scaled by a softmax function. Specifically [20],

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \quad (3)$$

$$CE = -\sum_{i=1}^{C=2} t_i \log(f(s)_i) \quad (4)$$

where t is a one-hot vector and p is the correct label. If there is only one element in t , where $t=tp$, that forms the CE non-zero. Then, Eq. (4) becomes [20]:

$$CE = -\log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right) \quad (5)$$

where p represents the true class (i.e., positive class). It is noted that the CCE will go down as the prediction (i.e., the predicted score for the true class) becomes more accurate. CCE equals zero if the prediction is perfect. Therefore, cross-entropy can be a loss function for training a classification model.

In semantic segmentation tasks, soft Dice loss is a commonly used object function. It was inspired by the Dice coefficient formula suggested by T.J. Sørensen (1948) [21]. The Dice coefficient measures the similarity of 2 samples. In the context of semantic segmentation, it measures the overlapped region between the predicted mask and the ground truth. Simply, the larger the coefficient, the better the performance of the prediction model. The Dice coefficient is given by:

$$Dice = \frac{2 \times |mask \cap prediction|}{|mask| + |prediction|} \quad (6)$$

In contrast, soft Dice loss is defined as the subtraction of the Dice coefficient from 1. The larger the loss values, the worse the performance of the model. Given the fact that any image could be treated as a 2-dimensional matrix, where each element is equivalent to a pixel of that image, $|A \cap B|$ can be approximated as the element-wise multiplication between the prediction and mask. Then, sum the resulting matrix.

To quantify $|A|$ and $|B|$, one can apply the squared sum of each matrix. Therefore, the mathematical expression of soft Dice loss is defined as follows:

$$Soft - dice = 1 - \frac{2 \sum_{pixels} y_{true} y_{predicted}}{\sum_{pixels} (y_{true}^2 + y_{predicted}^2)} \quad (7)$$

Metrics

To evaluate the model on a dataset, different metrics are applied. Some commonly used metrics for classification tasks are the confusion matrix, sensitivity, precision, and accuracy. The confusion matrix illustrates the performance of a model with two or more classes. A confusion matrix contains 4 components including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) where TP is defined as a total number of accurate predictions that are “positive”, FP is defined as the total number of inaccurate predictions

that are “positive”, TN is defined as the total number of accurate predictions that are “negative” and FN is defined by the total number of inaccurate predictions that are “negative”.

Precision (also called positive predictive value) is the probability that subjects having a positive test precisely have the disease. Meanwhile, the recall of a test (also called the true positive rate) is defined as the proportion of subjects having the disease that have a positive result. In other words, the recall is a high value when a model is highly sensitive. The precision and recall can be defined as follows:

$$Precision = \frac{TP}{(TP + FP)} \quad (8)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (9)$$

Classification accuracy is the number of correct predictions made as a ratio of all predictions made. This is the most common evaluation metric for classification problems. However, it is only suitable when there are an equal number of observations in each class (which is rarely the case), while most of the models often face an imbalanced dataset. When a dataset is imbalanced, the dominance in a proportion of the majority class can make classification accuracy a misleading evaluation.

Semantic segmentation uses different metrics for evaluation compared to classification tasks. The most widely used metrics for segmentation models are the Dice coefficient and intersection over union (IoU). IoU is a metric that allows the evaluation of how much overlap a predicted bounding box has on the ground truth bounding box. The IoU compares the ratio of the area where the two boxes overlap to the total combined area of the two boxes.

Assuming the predicted mask is bounded by a rectangle and the ground truth is also placed within another rectangle, the coordinates of these two rectangles are used to simplify the calculation and implementation of the IoU algorithm as described in Fig. 1.

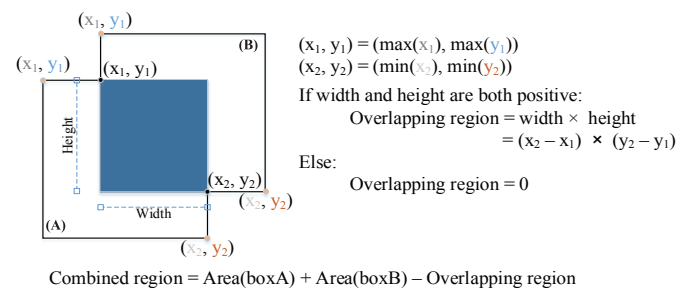


Fig. 1. Implementation of IoU calculation.

2.3. Transfer learning with VGG16 & UNET

The transfer learning technique allows pre-trained models to achieve the expected accuracy with fewer training epochs on a new dataset. Especially when the dataset is small, reusing knowledge from pre-trained models will help the trained models make better predictions because the models are learning from both knowledge sources, that is, new data and old data. Due to the advantages of using pre-trained models and the goals of this study, two popular models were selected including VGG16 [22] for the classification task and UNet [23] for the segmentation task. VGG16 is a CNN model that took first and second place in the localization and classification tasks respectively in the 2014 ImageNet challenge. The pretrained VGG16 has the ability of learning and extracting sufficient features for the classification task. UNet was presented by O. Ronneberger, et al. (2015) [23] to segment neural structures in human brain in 2012. The UNet architecture consists of two U-shaped symmetrical paths: the contraction (also called the encoder) is on the left and the extension (also called the decoder) is on the right. The contraction path is used to do the feature extraction task of capturing the context of the image. This path is called contraction because the size of the layers are decreasing. Simultaneously, the depth of layers increases gradually from 3 to 512. The extension path contains symmetrical layers corresponding to the layers of the contraction path. The decoder process is applied to help increase the layer size gradually. Finally, a segmented mask is obtained to show the prediction label of each pixel. The unique feature of UNet architecture is that it applies a symmetric skip connection between the encoder and decoder path.

2.4. Data collection

Table 1 presents two segmentation datasets that were collected and utilized to train RBC and WBC segmentation models. The combined datasets are called the WBC Image Dataset, which was published by X. Zheng (2018) [24] and “erythrocytesIDB” proposed by Universidad de Oriente de Cuba [25], respectively, for WBCs and RBCs. For WBCs, the dataset contains three hundred 120×120 px sub-images of a single WBC and one hundred 300×300 px colour images (including neutrophils, eosinophils, basophils, monocytes, and lymphocytes). These samples were taken by a motorized auto-focus microscope, then were processed with a developed haematology reagent for rapid WBC staining. For RBCs, the erythrocytesIDB dataset contains images of peripheral blood smear samples taken from patients with sickle cell disease that contain full-field images and individual cells classified as circular, elongated, or other. These samples were fixed with absolute alcohol and stained with Giemsa in a proportion of 2% of

reagent to 1 ml of distilled water. The RBC segmentation model treated each sample as a full-sized blood smear microscopic image with a binary mask including red blood cell pixels that were coloured white. After segmentation, the number of samples in the datasets changed in size, as shown in Table 2.

Table 1. Segmentation database.

Dataset name	Number of images/Cell type	Release year
ErythrocytesIDB	320/RBCs	2017
WBC image dataset	400/WBCs	2018

Table 2. Distribution of samples for classification.

Class name	Sample size
Normal	1503
ALL	1503
Malaria	891

Table 3 shows the sample size for the training, validation, and testing subsets. The percentages for training, validation, and testing were chosen as 50%, 25%, and 25%, respectively. The dataset, in fact, is made up of blood smear microscopic pictures gathered from different sources. Table 4 shows the distribution of samples organized by database name.

Table 3. Sample size of each data subset.

Data set	Sample size
Training set	1948 (50%)
Validation set	974 (25%)
Testing set	975 (25%)

Table 4. Distribution of dataset by sources.

Database name	Disease	Number of images
ALL-IDB	Leukaemia	1503
MLSRC	Malaria	668
MP-IDP	Malaria	223
ALL-IDB	Normal	1503

3. Implementation of model

3.1. System overview

Figure 2 illustrates the three-stage classifier system that was created to detect ALL or malaria on blood smear microscopic images. The system took each fed-in block to predict the mask for each block, and then combined all masks to generate the binary mask for the entire image in the first step. Then, the system segregated each blood cell (i.e., RBC and WBC) into every single image using contour finding and the watershed method.

In the second stage, each WBC was isolated and sent into a CNN classifier, which verifies whether the WBC is leukemic or malignant. Meanwhile, each RBC was loaded

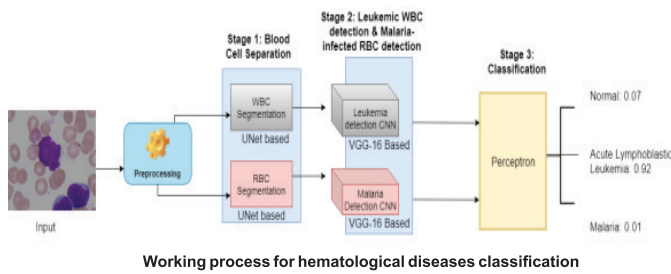


Fig. 2. The process of classifier.

into a second CNN model to determine the probability of an RBC becoming infected with malaria parasites. Only the maximum predicted probability for each type of RBC/WBC was reserved because an image might contain both numerous RBCs and WBCs. The two probabilities were concatenated and given to a perceptron network in the last stage of the system, which determines whether the studied image was a normal cell, infected with malaria, or ALL.

3.2. Implementation

Stage 1: Blood cell separation: two segmentation models were developed, which were the WBC segmentation model for WBCs and the RBC segmentation model for RBCs. Every image was divided into numerous 128x128 px images to save time and prevent system crash owing to the large size of the images. Each sub-image was then supplied to one of the two segmentation models based on the UNet pre-trained model. The predicted binary masks of all sub-images were then concatenated together for each segmentation model to generate the total binary masking for the original image. The whole process for the first stage is visualized in Fig. 3.

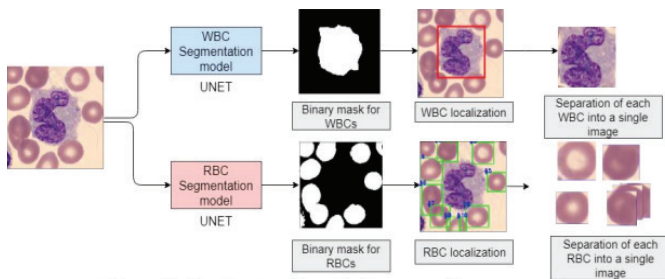


Fig. 3. Visualization the first stage of the system.

The results of the UNet training process were the RBC mask, in which positive pixels (i.e., white pixels) correspond to RBCs, and the WBC mask, in which each white pixel corresponded to WBCs. These masks are commonly used in fully convolutional networks for segmenting biological molecules. After that, each binary mask was subjected to contour finding and minimum enclosing rectangle to produce rectangle images with only one cell of interest in the centre. The models were trained using the soft Dice loss function and the customized IoU measure, which was adapted from the original UNet.

Stage 2: Blood cell type classification: following Stage 1, each isolated cell was fed into one of two models in Stage 2 as shown in Fig. 4. The first classifier determines the likelihood of a malaria-infected RBC. The second model identified the likelihood that a WBC is leukemic. Both models were trained in 30 epochs on the same training dataset and then validated on the same validation dataset using the same VGG16 structures with the same modifications. The last layer was removed and replaced with a sigmoid-activated layer of 2 units. Stage 2 generates a list of probabilities for RBCs being infected with malaria and a list of probabilities for WBCs being leukemic lymphoblasts from an image containing numerous RBCs and WBCs. Nonetheless, only the highest probability of each vector was preserved for processing in the next stage.

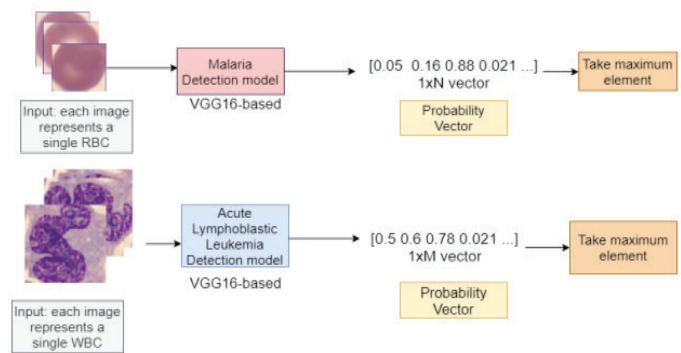


Fig. 4. Visualization of Stage 2 of the proposed model.

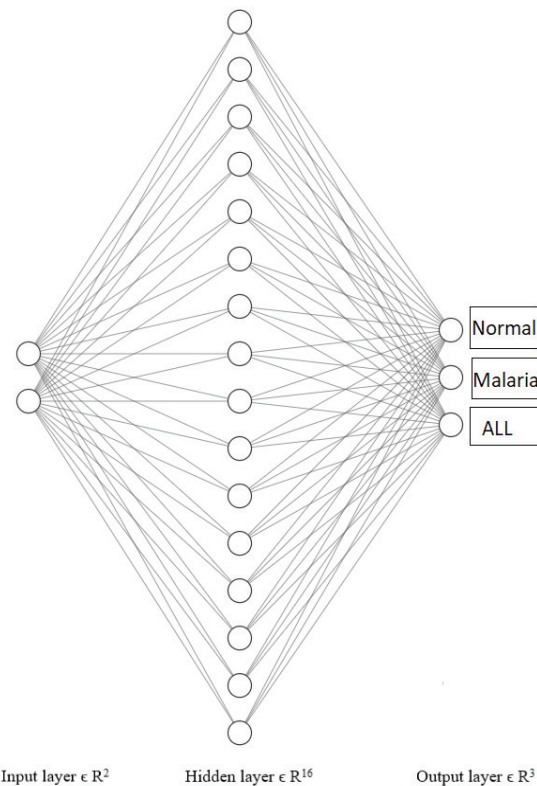


Fig. 5. Visualization of the perceptron architecture.

Stage 3: Classification utilizing perceptron as the combining model: the two maximum probabilities from Stage 2 were used as the two input neurons for a perceptron model in the third stage of this technique. The perceptron model was made up of three layers as illustrated in Fig. 5. The first layer is the input layer, which has two neurons that correspond to the two maximum probabilities mentioned. The middle layer is a hidden layer of neurons. The last layer is the output of 16 nodes, which have a softmax activation function and indicate the probabilities of the original input classified as “malaria”, “ALL”, or “normal”. As a combined classifier, the perceptron model is used. On the training set, the perceptron was trained for 20 epochs.

4. Results and discussion

Figures 6 and 7 show the results of the two WBC and RBC segmentation models, respectively. The training loss and validation loss for the WBC segmentation model significantly reduced throughout epochs indicating that the model is accurately and effectively trained. After 85 epochs, the model learns and begins to converge at a mean IoU of 0.91. A 0.91 mean IoU score simply means that the projected segmentation region overlaps the ground truth by 91%, which is a good and promising precision for the segmentation challenge. With a mean IoU of 0.94, the RBC segmentation model’s training and validation losses saturate after 90 epochs.



Fig. 6. Training and validation losses / IoUs of WBC segmentation model.

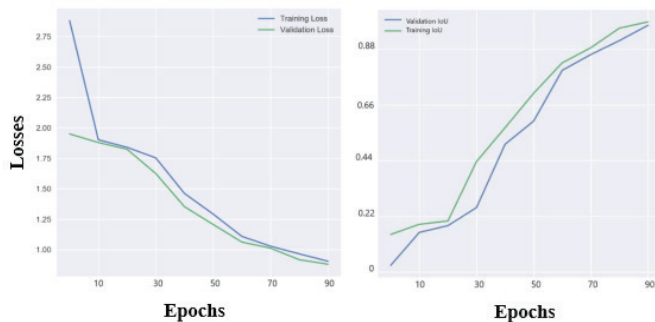


Fig. 7. Training and validation losses/IoUs of RBC segmentation model.

In the second stage, two CNN classification models are built. Fig. 8 shows the loss and accuracy values of the malaria detection model over epochs. Fig. 9 illustrates the same parameters for ALL. Both models achieved a high accuracy of validation when training.

A confusion matrix is developed to infer the meaning of precision, recall, and accuracy to evaluate the performance of an entire system on the testing dataset. Fig. 10 and Table 5 provide more information on these metrics. Fig. 10 depicts the confusion matrix’s insight, which provides readers with a numerical expression of the model’s performance. Table 5 summarizes the overall performance of the system on the testing dataset.

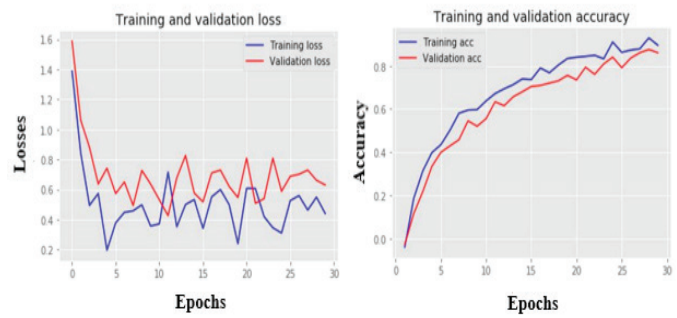


Fig. 8. Training and validation losses/accuracies of the malaria detection model.

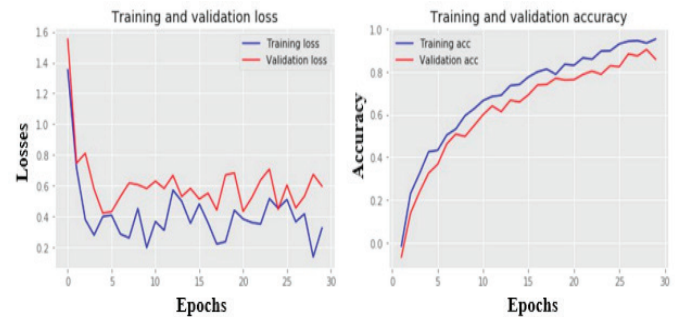


Fig. 9. Training and validation losses/accuracies for the ALL detection model.

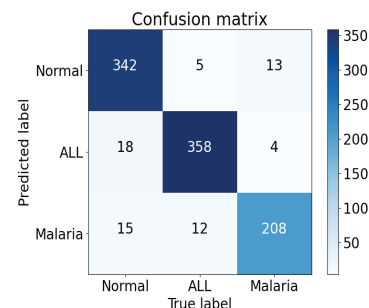


Fig. 10. Confusion matrix for evaluation of whole system on testing set.

Table 5. The results of precision, recall for each class and whole accuracy of the system.

Class name	Metrics (%)		
	Normal	ALL	Malaria
Precision	95	94	88
Recall	91	95	92
Whole accuracy of the system	93		

With the help of ensemble learning, the overall performance of the classification system reaches an accuracy of 93%. Individually, each sub-performance model is rather good and acceptable. For the segmentation task, the average IoU is over 0.9 indicating that the system can forecast masks with over 90% overlapped area with true ground truths. Both CNN detection models were found to have an accuracy of more than 90%. The accuracy of the malaria detection model is less than that of the model for detecting acute lymphoblastic leukaemia (89 vs 92%). Furthermore, the loss plots of all the models show that the models are correctly trained without overfitting or underfitting. Tables 6 and 7 compare the precision and recall of classification among the proposed system and other publications.

Table 6. Comparison of the proposed method in malaria detection with other studies by precision and recall.

Study	Malaria detection precision (%)	Malaria detection recall (%)
Our method	88	92
J. Hung, et al. (2017) [16]	98	97
N.E. Ross, et al. (2006) [17]	81	85
M. Poostchi, et al. (2018) [18]	88	91
M. Sheikhsosseini, et al. (2013) [19]	Not mentioned	82.28

Table 7. Comparison of the proposed method in ALL detection with other studies by precision and recall.

Study	ALL detection precision (%)	ALL detection recall (%)
Our method	94	95
H. Chowdhury, et al. (2018) [9]	90	89
T.T.P. Thanh, et al. (2018) [13]	97	97
L. Putzu, et al. (2014) [15]	Not mentioned	93

Compared to other studies, the proposed study outperforms H. Chowdhury, et al. (2018) [9] and L. Putzu, et al. (2014) [15], but lags behind T.T.P. Thanh, et al. (2018) [13] in detecting ALL in a blood smear microscopic image. The algorithm of Thanh, et al., on the other hand, appears to be overfitted, as their study used only 108 unique images for the dataset. Even though augmentation can expand datasets, it is not recommended to use augmented images on testing, as group of Thanh did, because this might lead to overfitting issues. In terms of malaria detection, the proposed study outperforms M.

Poostchi, et al. (2018) [18], M. Sheikhsosseini, et al. (2013) [19], and N.E. Ross, et al. (2006) [17], but falls short of J. Hung, et al. (2017) [16]. J. Hung’s group built a two-stage detection approach with Faster-RCNN, which is comparable to our proposed solution, however, their method requires hard-core training and does not directly flag up problematic cells.

Although the proposed study is not among the top performances, the suggested method still has several advantages than other studies in some respects. Firstly, the proposed method can perform both abnormal leukocytes for ALL and malaria-infected blood cells identification while other studies only detected one type of disease. Secondly, compared to related works that suggested only one CNN classifier, the proposed system is more interpretable as it can indicate the locations of not only ALL in white blood cells but also malaria-infected blood cells. Obviously, the detecting result of multiple diseases from blood smear microscopic images will provide more information for the histopathologist to evaluate the sample. Due to the proposed study indicating the potential subjects for detection, haematologists can have a quick look through those potential objects to verify the results faster and more precisely. In addition, if diagnosed with a disease, haematologists can also see why the proposed system classified the disease and in which cell(s) the model recognized the disease(s).

The system, in fact, has a number of challenges. The method takes longer to process an image than other alternatives, which is the most evident shortcoming. This is owing to the time-consuming nature of merging multiple models and processes at once. However, if parallel-computation techniques are used, this problem can be solved. Another issue that has been brought up is a lack of training data. Ideally, one would want hundreds of thousands of images to train a deep learning system. This difficulty can be solved by collecting more data or using the synthetic minority over-sampling technique (SMOTE).

5. Conclusions

This study has proposed a learning-based method that consists of three stages to segment white blood cells and red blood cells in blood smear microscopic images by the UNet model. The probability of acute lymphoblastic leukaemia and malaria is calculated by the VGG16 model and then those diseases are classified by the perceptron model. The performance of the proposed study achieves an average IoU of 90% for the segmentation task and an overall classification accuracy of 93% for the three types of conditions (malaria, acute lymphoblastic leukaemia, and normal). Overall, the achieved results indicate that the proposed system has promising potential for

the development of learning-based approaches for automated diagnosis of blood diseases utilizing blood smear microscopic images.

CRedit author statement

Quyen Hoang Vo, Xuan-Hieu Le, Thanh-Hai Le: Data curation, Software, Visualisation, Investigation, Writing original draft preparation; Thi-Thu-Hien Pham: Conceptualisation, Methodology, Supervision.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support provided to this study by Vietnam National University, Ho Chi Minh city under Grant No. DS2020-28-02.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

- [1] World Health Organization (2020), *World Malaria Report 2020*, <https://www.who.int/publications/i/item/9789240015791>, accessed 27 March 2021.
- [2] Global Burden of Disease 2015 Disease and Injury Incidence and Prevalence Collaborators (2016), “Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: A systematic analysis for the global burden of disease study 2015”, *Lancet*, **388**, pp.1545-1602, DOI: 10.1016/S0140-6736(16)31678-6.
- [3] K.N. Suh, K.C. Kain, J.S. Keystone (2004), “Malaria”, *CMAJ*, **170(11)**, pp.1693-1702, DOI: 10.1503/cmaj.1030418.
- [4] P. Anderlini, M. Luna, H.M. Kantarjian, et al. (1996), “Causes of initial remission induction failure in patients with acute myeloid leukaemia and myelodysplastic syndromes”, *Leukemia*, **10(4)**, pp.600-608.
- [5] M.T. Makler, C.J. Palmer, A.L. Alger (1998), “A review of practical techniques for the diagnosis of malaria”, *Annals of Tropical Medicine and Parasitology*, **92(4)**, pp.419-433, DOI: 10.1080/00034989859401.
- [6] Division of Parasitic Diseases and Malaria (2020), *Malaria Diagnostic Tests*, https://www.cdc.gov/malaria/diagnosis_treatment/diagnostic_tools.html#tabs-1-1, accessed 27 March 2021.
- [7] American Cancer Society, *Tests for Acute Myeloid Leukemia (AML)*, <https://www.cancer.org/cancer/acute-myeloid-leukemia/detection-diagnosis-staging/how-diagnosed.html>, accessed 27 March 2021.
- [8] S. Shafique, S. Tehsin (2018), “Acute lymphoblastic leukaemia detection and classification of its subtypes using pretrained deep convolutional neural networks”, *Technology in Cancer Research and Treatment*, **17**, DOI: 10.1177/1533033818802789.
- [9] H. Chowdhury, S. Banik, A. Hossain, et al. (2018), *Detection of Acute Lymphocytic Leukaemia (ALL) and Its Type by Image Processing and Machine Learning*, Thesis, BRAC University, 80pp.
- [10] M. Poostchi, I. Ersoy, K. McMenamin, et al. (2018), “Malaria parasite detection and cell counting for human and mouse using thin blood smear microscopy”, *Journal of Medical Imaging*, **5(4)**, DOI: 10.1117/1.JMI.5.4.044506.
- [11] K.P. Smith, J.E. Kirby (2020), “Image analysis and artificial intelligence in infectious disease diagnostics”, *Clinical Microbiology and Infection*, **26(10)**, pp.1318-1323, DOI: 10.1016/j.cmi.2020.03.012.
- [12] S. Chakraborty, K. Mali (2020), “An overview of biomedical image analysis from the deep learning perspective”, *Applications of Advanced Machine Intelligence in Computer Vision and Object Recognition: Emerging Research and Opportunities*, pp.197-218, DOI: 10.4018/978-1-7998-2736-8.ch008.
- [13] T.T.P. Thanh, C. Vununu, S. Atoev, et al. (2018), “Leukemia blood cell image classification using convolutional neural network”, *International Journal of Computer Theory and Engineering*, **10(2)**, pp.54-58, DOI: 10.7763/IJCTE.2018.V10.1198.
- [14] P. Jagadev, H. Virani (2017), “Detection of leukaemia and its types using image processing and machine learning”, *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, pp.522-526, DOI: 10.1109/ICOEI.2017.8300983.
- [15] L. Putzu, G. Caocci, C.D. Ruberto (2014), “Leucocyte classification for leukaemia detection using image processing techniques”, *Artificial Intelligence in Medicine*, **62(3)**, pp.179-191, DOI: 10.1016/j.artmed.2014.09.002.
- [16] J. Hung, S.C.P. Lopes, O.A. Nery, et al. (2017), “Applying faster R-CNN for object detection on malaria images”, *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.808-813, DOI: 10.1109/cvprw.2017.112.
- [17] N.E. Ross, C.J. Pritchard, D.M. Rubin, et al. (2006), “Automated image processing method for the diagnosis and classification of malaria on thin blood smears”, *Medical and Biological Engineering and Computing*, **44(5)**, pp.427-436, DOI: 10.1007/s11517-006-0044-2.
- [18] M. Poostchi, K. Silamut, R.J. Maude, et al. (2018), “Image analysis and machine learning for detecting malaria”, *Translational Research*, **194**, pp.36-55, DOI: 10.1016/j.trsl.2017.12.004.
- [19] M. Sheikhsosseini, H. Rabbani, M. Zekri, et al. (2013), “Automatic diagnosis of malaria based on complete circle-ellipse fitting search algorithm”, *Journal of Microscopy*, **252(3)**, pp.189-203, DOI: 10.1111/jmi.12081.
- [20] D.R. Cox (1958), “The regression analysis of binary sequences”, *Journal of The Royal Statistical Society, Series B (Methodological)*, **20(2)**, pp.215-242.
- [21] T.J. Sørensen (1948), “A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons”, *København I Kommission hos E. Munksgaard*, 34pp.
- [22] K. Simonyan, A. Zisserman (2015), “Very deep convolutional networks for large-scale image recognition”, *International Conference on Learning Representations (ICLR) 2015*, DOI: 10.48550/arXiv.1409.1556.
- [23] O. Ronneberger, P. Fischer, T. Brox (2015), “U-net: Convolutional networks for biomedical image segmentation”, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, Springer, **9351**, DOI: 10.1007/978-3-319-24574-4_28
- [24] X. Zheng (2018), “Data for: Fast and robust segmentation of cell images by self-supervised learning”, *Mendeley Data*, DOI: 10.17632/w7cvmnm4c5.1.
- [25] J.B.Z. Alfonso (2017), “ErythrocytesIDB”, *Special Hematology Department of The General Hospital*, <http://erythrocytesidb.uib.es/>, accessed 27 March, 2021.