An Investigation Of U-net For Image Segmentation For Leukemia Classification Problem

Zheng Jiawen , Hadi Affendy Dahlan

Faculty of Information and Technology, National University of Malaysia, 43600 UKM Bangi, Selangor, Malaysia

* Corresponding author : a184590@siswa.ukm.edu.my

Detecting leukemia, among other Microscopic medical image applications, necessitates the segmentation of images to distinguish normal white blood cells from their mutated counterparts. Traditionally, this task relies on the expertise of medical professionals to identify the distinct types of leukemia-causing white blood cells. The proposed solution involves the implementation of the U- Net architecture for image segmentation. The methodology starts with image augmentation techniques, grayscale conversion for preprocessing, followed by U-Net segmentation and Convolutional Neural Network (CNN) classification. The expected results include an end-to-end system capable of accurately detecting and classifying leukemia subtypes (AML or APML) alongside normal white blood cells. Furthermore, since many datasets for this use case are few and not available in the quantity normally required to train neural networks, the proposed model is designed to these constraints while delivering state of the art results, achieving 0.9 accuracy. Despite the existing repertoire of methods, such as image augmentation and the utilization of generative adversarial techniques (GAN), which have demonstrated good results, there remains ample scope for further refinement. The WBC Dataset was chosen for this project because it is sourced from the UKM medical lab, labeled with the help of pathologists.

Keywords: machine learning, medical image processing, microscopic image segmentation, UNET

1 INTRODUCTION

Leukemia is a type of blood cancer which generates malignant white blood cells (WBCs) in the human body. These abnormal blood cells cease to function as normal white blood cells, compromising the immune system. They can also limit the ability of bone marrow to generate red blood cells and platelets.[1] Acute lymphoblastic leukemia is usually diagnosed by performing a complete blood count test [10]. Another method called bone marrow aspiration followed by microscopic examining of blood smear4 is performed to confirm that patient is diagnosed with leukemia. After obtaining a sample of the patient's blood, the blood is dyed with a purple dye which enhances its visibility under a microscope. An expert observes this sample and detects the abnormal white blood cells present in the sample. As illustrated in Figure 1.0, promyelocytes serve as precursors to granulocytes, and their presence indicates Acute Myeloid Leukemia. Similarly, the detection of lymphoblasts, which are precursors to lymphocytes, signifies the presence of Acute Lymphoblastic Leukemia. Second and following paragraphs of the text have no initial indentation as well. Normal text of the paper. In this

paper we apply a machine learning approach to the detection of leukemia using instance segmentation. Microscope image segmentation is of great importance in many medical applications such as anomaly detection. These applications regularly require the segmentation of images into 3 categories: Neutrophil, Promyelocytes and Lymphoblast. In this paper a U net architecture is proposed to perform segmentation by extracting fine grained features, while controlling the number of trainable parameters to prevent overfitting. Then a Convolutional neural network is used to classify the images. The network is trained in 300 microscopic images, comprising of 3 classes in an imbalanced way. This dataset is of a manageable size to be created by a medical institution. The manual segmentation of medical images by human labor presents significant challenges, as it is not only laborious and time-consuming process but also prone to inaccuracies, particularly given the surge in medical imaging modalities and the overwhelming volume of images requiring scrutiny [2]. The sheer magnitude of this task renders the creation of a large and diverse database an imposing barrier, underscoring a recurring challenge when assessing the viability of neural networks for segmentation in the realm of medical imaging.

To address these inherent limitations, various studies have proposed computer-aided diagnosing methods for Acute Lymphoblastic Leukemia (ALL), leveraging microscopic blood image analysis for enhanced detection. These automated methods have demonstrated superior efficiency, speed, cost-effectiveness, and accuracy compared to manual approaches. For instance, Shafique, S., & Tehsin, S. implemented an automated method for White Blood Cell (WBC) segmentation and leukemia classification. They had used data augmentation such as image rotation and mirroring for preprocessing of microscopic blood images, then they a pretrained AlexNet was used for white cell classification. Upon reviewing existing literature, it becomes evident that limited attention has been devoted to the classification of ALL subtypes, specifically Acute Promyelocytic Leukemia (APML) and Acute Myeloid Leukemia (AML). The challenges in discerning these subtypes arise from their heightened intraclass variability and notable interclass similarity, making accurate detection and classification inherently difficult. Recognizing the critical role that subtype identification plays in guiding treatment plans; it becomes imperative to ensure correct diagnoses.

2 MATERIAL AND METHODS

The research recognizes the limited work done in the classification of specific ALL subtypes, such as Acute Promyelocytic Leukemia (APML) and AML, owing to their heightened intraclass variability and notable interclass similarity. Correctly identifying these subtypes is pivotal in guiding clinicians toward recommending tailored treatment plans, underscoring the significance of accurate classification in leukemia diagnosis.

In the realm of medical image processing, the stages pre-processing, and feature extraction are pivotal to successful segmentation and classification in the analysis of contaminated blood cell images. Abd et al. [6] proposed an automated approach to identify the count of leukemia-infected cells. They employed hue, saturation, and value (HSV) segmentation to eliminate background white blood cells, followed by morphological erosion to address overlapping cells. The segmentation process involved utilizing the H component of the hue, saturation, and intensity (HSI) color space on color smears of microscopic images. This method has achieved a 97.8% accuracy in counting of blast cells. Similarly, in [7] the authors employed a pipeline consisting of color transformation to translate the image into HSV color space and then adjust the S component of the as preprocessing

step before using a fixed thresholding method to segment the cells. The accuracy they obtained from this method was 95.82%.

In a study by Victor et al. [7], the authors employ the use of U-net architecture to segment the cell nucleus, then up sampling to output the final image. This image is then further preprocessed Convolutional Neural network is then implemented for classification of the segmented image, and the CNN's model was found to be 99.6%. Alagu et al. [3] utilized a combined strategy that involved the integration of UNET and deep Convolutional Neural Network (CNN), and this method allowed them to reach an accuracy of 98.2%.

Effective machine learning models demand input features that possess relevance and significance for predicting outcomes. It is essential to recognize that not all features carry equal importance in the prediction task, and certain features may even introduce undesirable noise into the model. To address this issue, two methods, namely feature selection and feature extraction, are employed. In [3] the authors address the selection of distinguishing features from smear images, performing feature extraction using 3 different CNN models AlexNet, GoogleNet, and SqueezeNet. 3000 features are obtained, and the best features are selected using recursive feature elimination, mutual information, and the minimum recursive maximal relevance algorithms. Other Examples of feature selection techniques, used for leukemia classification in blood images, are the principal component analysis (PCA) in [4], gain ratio in [5]. Emulating the process outlined in [3], we aim to mitigate model overfitting, a prevalent concern with small datasets, by reducing the dimensionality of our data. This strategic approach aligns with the overarching goal of enhancing model generalization and performance in the challenging landscape of medical image classification.

Drawing from the existing literature, it is evident that numerous studies have extensively utilized datasets sourced from the ALL-IDB2 repository. This dataset does not contain any training instances for AML and thus models

trained from such datasets can only classify ALL. Traditional methodologies such as image processing have achieved accuracies comparable to state-of-the-art models. Nevertheless, the superiority of deep learning models persists, attributed to an increased number of hidden layers and strategic hyperparameter selection. Consistently, the incorporation of techniques like data augmentation, dropout regularization, and ensemble approaches has demonstrated improved accuracies, solidifying the deep Convolutional Neural Network (CNN) as the standard for medical image analysis.

2.1 Architectural Design

The proposed system extracts potential features from the segmented nuclei of both healthy and leukemia causing cells. The system's architecture is illustrated in Figure 1.0. Initially the images from the dataset undergo image augmentation to increase the number of samples. The applied augmentations are as shown in fig 1.0. Following that two preprocessing steps are applied which are Greyscale conversion from RBG color space and subsequently Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm applied to the grayscale image for contrast enhancement. These images are then fed into our UNET architecture which takes the preprocessed images as input and

outputs the input mask of the image. This effectively segments and separates the nucleus from the rest of the image, and this is a key feature we want for classification of the cell. A CNN performs classification task from the output and predicts the two sub types of Leukemia: AML or APML, or Normal white blood cells.

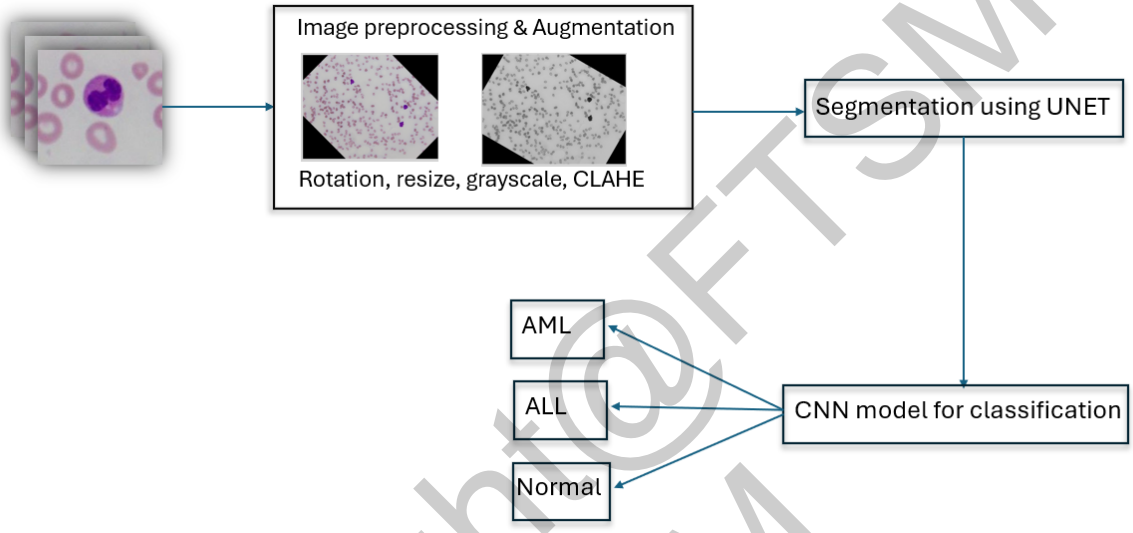


Fig 1.0 System Architecture design

Second and following paragraphs are not indented either, normal text of the paper (Cambria 11pt, block-aligned)

2.1.1 Image Augmentation

A prevalent challenge encountered in the implementation of machine learning in the field of medical imaging is the insufficient availability of data. This arises due to the time-consuming and costly nature of data collection. This is especially true since one of our classes is heavily imbalanced, namely the Promyelocytes, as evident in figure 2.1. To address this limitation, a set of techniques collectively known as "data augmentation" was employed to augment the dataset. These strategies involve generating additional data points based on the existing collected data. As evident in figure 3.3 we rotate the image by 30 and 45 degrees. This helps our model classify objects even when the camera or subjects are not perfectly aligned.

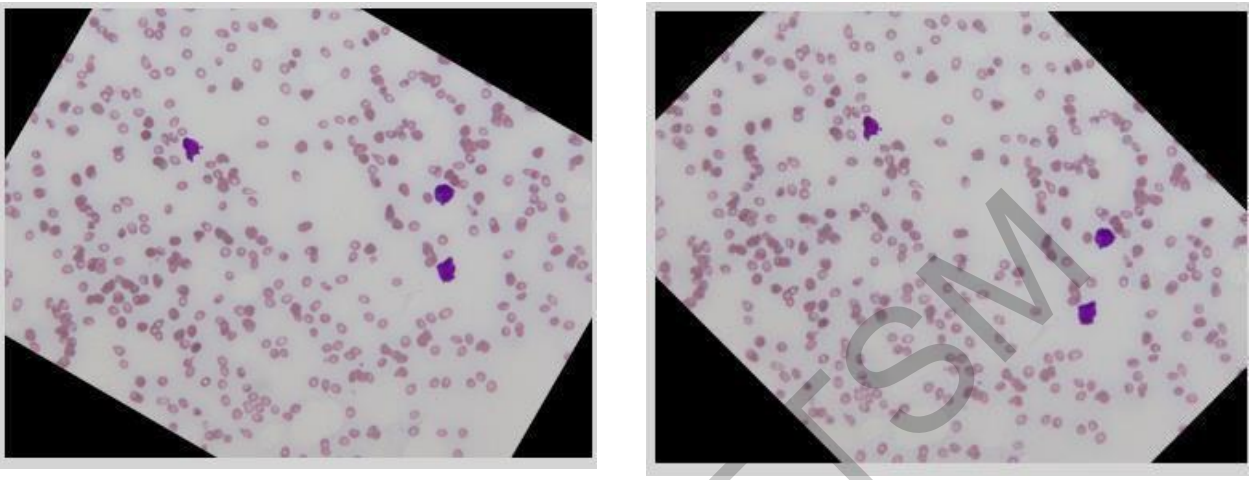


Fig 2.0 rotation variations

Method	Default	Adjusted
Auto Orient	None	applied
Zoom range	-	0.25
Rotation (Degrees)	-	30, 45, & 60
Resize	1024 x 1024	224 x 224

Table 2.0 image augmentations

fig 2.1 dataset composition and class distribution

2.1.2 UNET segmentation

The UNet model, renowned for its effectiveness in computer vision applications, plays a pivotal role in precisely segmenting white blood cell nuclei, particularly crucial in detecting alterations associated with Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML). AML and ALL causes changes in the shape, texture, and size of the nucleus. Segmentation of the nucleus from these White blood cells can help in classifying normal neutropihls from Promyelocytes and lymphoblast, precursors to AML and ALL respectively.

The architecture of the UNet, illustrated in Figure 1.0, features a unique structure comprising a narrowing path and an expanding path. The inducing route involves convolutions followed by ReLU and maxpooling operations, contributing to a reduction in spatial space and an augmentation in

available features. High-resolution features from the contracting path seamlessly integrate into the expansive pathway via a series of up sampling, preserving both feature and spatial information. The microscopic images of normal and masked cells serve as inputs for UNet, with the respective binary masks of nuclei obtained through a process involving contraction and expansion.

The segmentation accuracy is significantly enhanced through the concatenation operation, which fills the positional data gaps caused during convolution. The segmentation results are meticulously validated using accuracy and Receiver Operating Characteristic (ROC) measures alongside corresponding ground truth, ensuring the reliability of the outcomes. Furthermore, the emphasis on the precision of segmentation is evident, with both DICE and Intersection over Union (IoU) performance measures utilized for validation against expert-recommended ground truth masks. The results, expressed as values ranging from 0 to 1, exemplify the effectiveness of the UNet architecture in accurately delineating white blood cell nuclei, crucial for the identification of leukemia-induced changes.

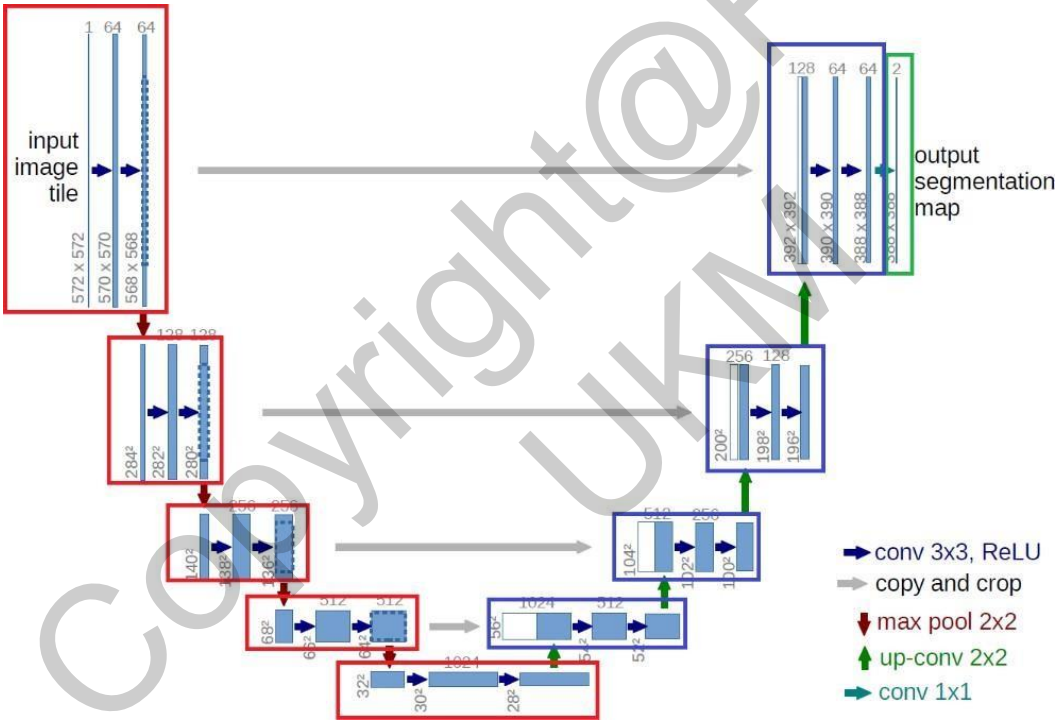


Fig 2.2 Unet Architecture design

2.1.3 Classifying on Segmented Images

A CNN model for leukemia type categorization (such as ALL and AML) was implemented. It employs a minimal amount of pre-processing in contrast to other image categorization methods. The model takes an input image (in this case the segmented image) and predicts the Leukemia type as an output. CNN Typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

2.1.4 Convolutional Layer

The fundamental component of a CNN is the convolution layer, responsible for the majority of the network's computational workload. In this layer, a dot product is performed between two matrices: one matrix consists of learnable parameters, known as a kernel, while the other matrix represents a restricted section of the receptive field. Despite the spatially smaller dimensions of the kernel compared to an image, it possesses greater depth. In the case of an image with three (RGB) channels, the kernel's height and width are spatially compact, but its depth extends across all three channels.

During the forward pass, the kernel slides across the height and width of the image, generating an image representation for each receptive region. This process results in a two-dimensional activation map, indicating the kernel's response at each spatial position in the image. The size of the sliding kernel is referred to as the stride.

For an input of size W x W x D and Dout kernels with a spatial size of F, a stride of S, and padding amount P, the output volume's size, denoted as Wout x Wout x Dout, can be determined using the

2.1.5 Pooling Layer

The pooling layer functions by substituting specific locations in the network's output with a summary statistic derived from nearby outputs. This process aids in diminishing the spatial size of the representation, thereby reducing the computational workload, and required weights. The pooling operation is applied individually to each slice of the representation.

Various pooling functions exist, including the average of the rectangular neighborhood, L2 norm of the rectangular neighborhood, and a weighted average based on the distance from the central pixel. However, the most widely adopted approach is max pooling, which extracts the maximum output from the neighborhood.

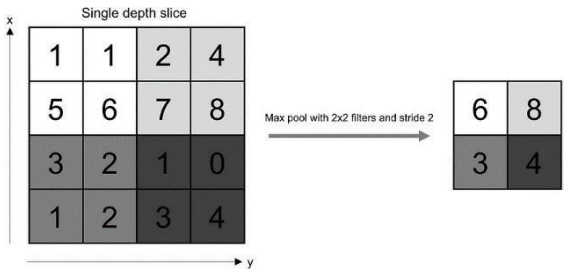


Fig 2.3 output from max pool layer

For an activation map with dimensions W x W x D, a pooling kernel of spatial size F, and a stride of S

2.1.6 Fully Connected Layer

The term "fully connected" is coined because a fully linked layer establishes connections between every neuron and another. These layers operate akin to traditional neural networks, where they engage in image classification based on convolved features. During this stage, loss functions are calculated and subsequently backpropagated. The suggested model consists of five interlinked layers, with the fifth serving as the output layer. All four interconnecting layers employ SoftMax as their activation function. The sigmoid activation function yields a probability ranging from 0 to 1 for each distinct class label that the model tries to predict.

2.2 Interface Design

Users can access the system without needing to login. The system guides the process by only displaying the available functionality, which is segmentation. After segmentation is successful only then will the option to classify appear for the user. A dynamic day/light theme toggle is added, and its usefulness will be evident in later steps.

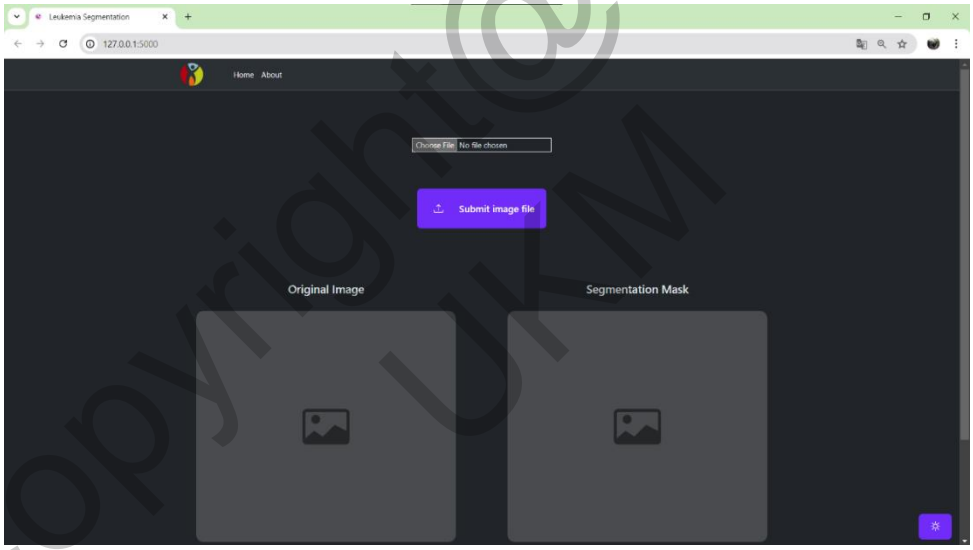


Fig 2.4 Landing page with dark theme

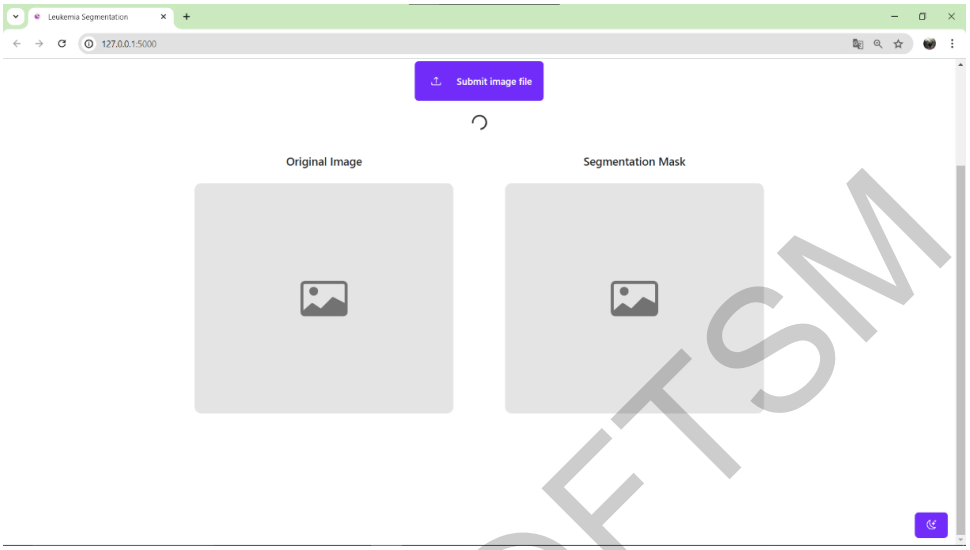


Fig 2.5 loading spinner indicating backend processing

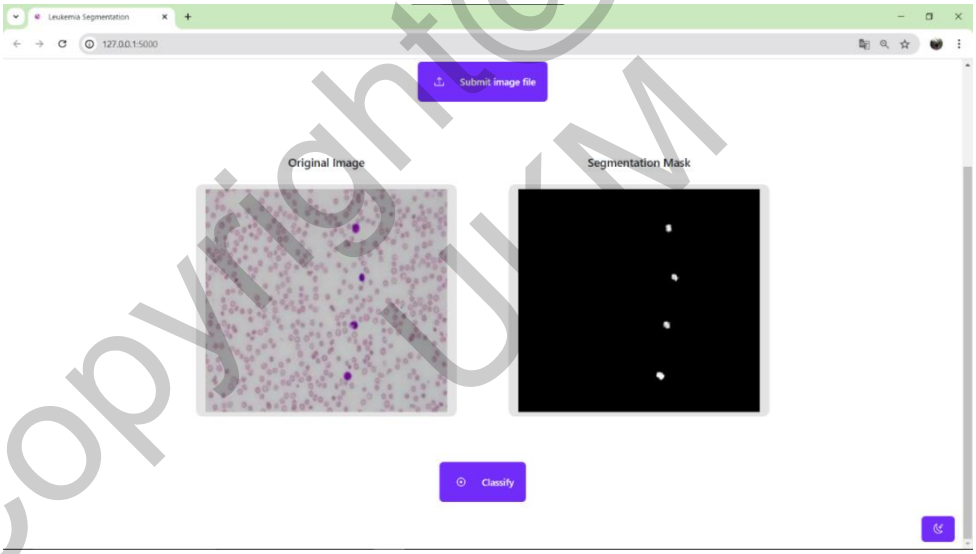


Fig 2.6 segmented results

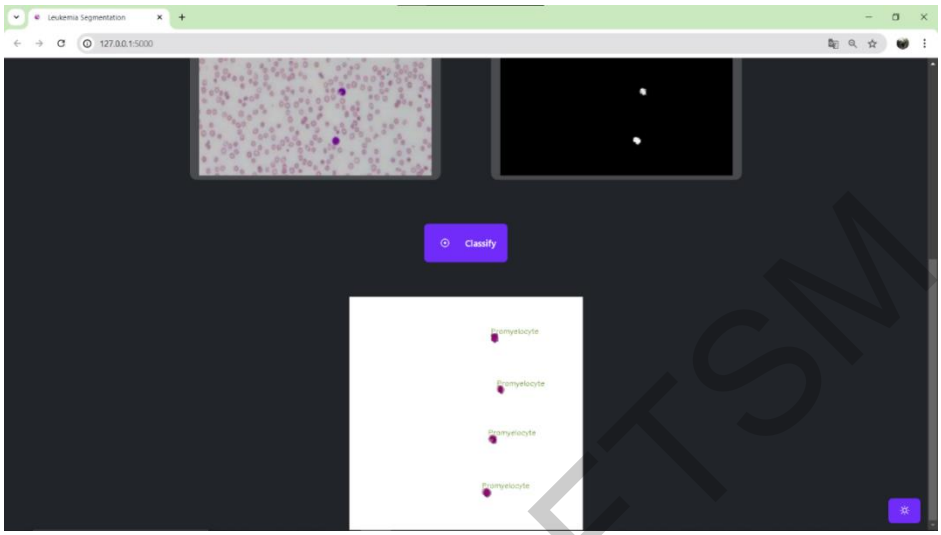


Fig 2.7 classification results in contrasting theme

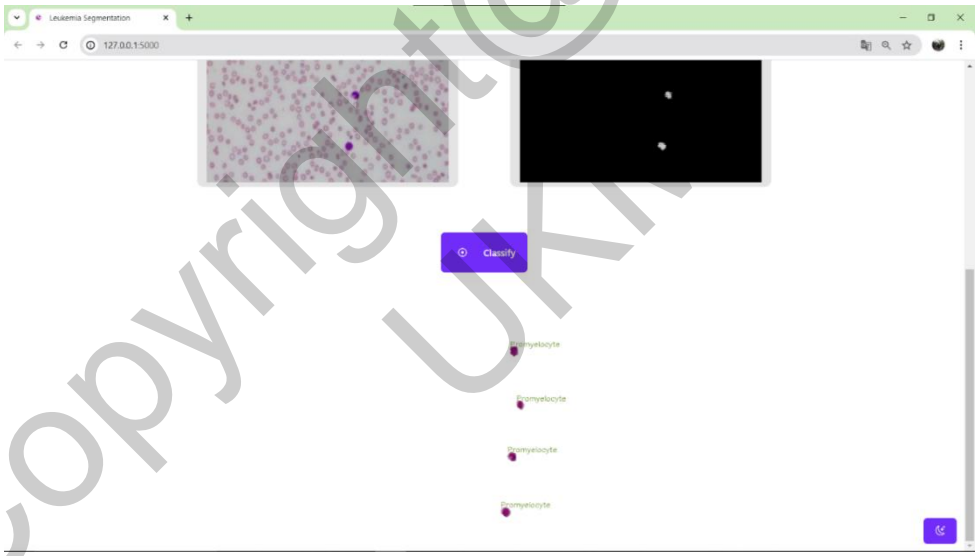
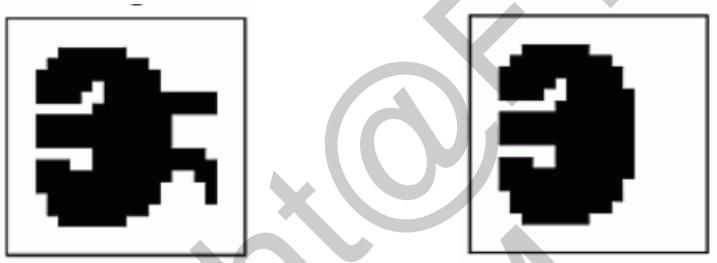


Fig 2.8 classification results in default light theme

2.3 Image Morphological Operations

Due to the small number of training images, the model is shown to be overfitting. This problem is addressed in section 5 test results by introducing early stopping and cross validation. Additionally, the model faced challenges when cells in the images were clumped together, causing the segmentation masks to merge incorrectly. To address the issue of clumped cells in the segmentation masks, we applied opening image morphological operation with a 5x5 kernel to effectively separate weakly connected masks. This technique involves a combination of erosion followed by dilation, which helps in breaking apart loosely connected regions within the masks, thereby improving the accuracy of individual cell segmentation.



Original image Opening

Fig 2.9 Example of image Opening morphological operation

3.0 Results and Discussion

Model testing is the process of evaluating and validating the performance of machine learning models to ensure their accuracy, robustness and functionality. The model was tested on google colab, a cloud service with minimal service and access to GPUs, with Intel Xeon CPU with 2 vCPUs (virtual CPUs) and T4 GPU with a memory clock rate 1250 MHz.

To begin, we utilized a train-test split of 70/30. The test set, comprising 30% of the data, includes both the images and their corresponding ground truth masks, which are stored in a separate test directory. This ensures that the test data remains unseen by the model during training, providing an unbiased evaluation of the model's performance. The test images and masks are loaded into sorted lists to maintain a consistent pairing between each image and its corresponding mask. Each image and mask are resized to 256x256 pixels to standardize the input dimensions for the model. Additionally, the pixel values of the images are normalized to the range [0, 1], ensuring that the model receives input data with uniform scaling.

Table 2 : Performance metrics

Binary Accuracy	Dice Coeff	Precision
0.998	0.348	0.795

In our first training attempt, training loss is 60% lower than validation loss. This signifies that the model is overfitting. To address this issue, we introduce early stopping and implement cross validation. The resulting plot shows significant improvement.

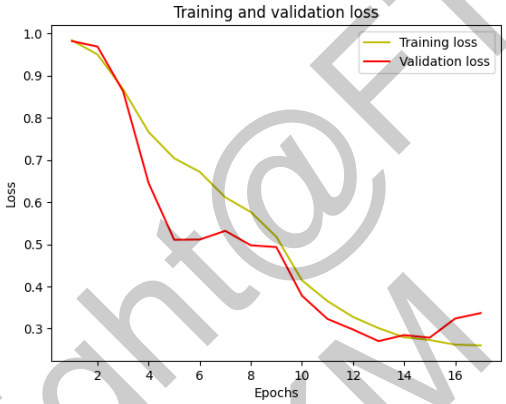


Fig 3 training and validation loss graphed

One potential issue that can be addressed in the future is the time required for model inferencing. Although the inference time itself is usually very fast, the web app takes considerably longer to display the output image. API testing suites such as postman may be used to identify the bottleneck in the system.

Cropping accurate ROIs require cells to be somewhat centered within the image. When a cell is right on the edge of the image, The crop goes outside the image and an error is returned. This is especially a problem when multiple cells instances are at the edge like in figure 1 as this impedes the ability to get an accurate crop. To fix this problem, The crop ROI is adjusted when a cell is detected at the edge of the image and the ROI coordinates are outside the image. For example, In the case that the cell is detected at the upper edge of the image, the ROI is moved down until it is contained within the images.

Due to the class imbalance problem where there is one segment/class that dominates the samples, the model might find it easier to output all pixels as belonging to the dominant class/segment. This constant prediction, although not informative, can still yield high accuracy and small loss. This is the case when we trained the model initially, where it scored 99% accuracy rate but only returned the background as the segmentation mask. It did this for all testing images. Using a different loss function called dice loss effectively addresses this challenge of class imbalance. Unlike traditional

loss functions, such as cross-entropy, which heavily penalize misclassifications and thus encourage the model to prioritize the dominant class, the Dice loss offers a different approach. By focusing on the spatial overlap between predicted and ground truth segmentation masks, the Dice loss accounts for class imbalance, incentivizing the model to accurately capture the boundaries of all classes, irrespective of their prevalence in the dataset. This characteristic proves invaluable in cases where the dominant class tends to overshadow the minority classes, as it ensures that the model's performance is not disproportionately skewed towards favoring the majority class.

4 CONCLUSION

In summary, the development and deployment of the proposed system involved overcoming numerous challenges, from initial design constraints to unexpected issues during model training. Despite these obstacles, the system successfully segments and classifies white blood cells into three categories: AML, ALL, and Neutrophils. The use of UNet for segmentation and a CNN for classification proved effective, achieving high accuracy and precision in tests. The user interface was designed with simplicity and usability in mind, incorporating dynamic widgets, progress indicators, and options for users to download segmentation masks and classification outputs.

However, several weaknesses remain. The system's performance on images with clumped cells needs improvement, and the time taken to display results on the website is considerable. Additionally, the system could benefit from more advanced image processing techniques to provide users with more informative outputs.

Future improvements include utilizing UNet for semantic segmentation to streamline the process and potentially increase efficiency. Increasing the training data and introducing more variations is also crucial to reduce overfitting and improve model robustness. Given the scarcity of suitable labelled medical datasets, innovative data augmentation techniques and leveraging transfer learning from pre-trained models could be beneficial. Addressing the issue of incorrectly merged segmentation masks through adaptive thresholding techniques can also enhance the accuracy of predictions if training data cannot be increased.

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to Dr Hadi Affendy Dahlan, for their invaluable contributions and collaborative spirit in mentoring me throughout this research project. Without his expert advice I could not have possibly completed this project.

I am also thankful to Dr Afzan Binti Adam for providing me with the vital information and connections that were important to undertaking this project.

REFERENCES

1. Helping hematologists conquer blood diseases worldwide. (n.d.). AmericanSocietyOfHemotology. http://www.hematology.org/
2. Patil, D. (n.d.-b). Medical Image Segmentation: A Review. International Journal of Computer Science and Mobile Computing. https://ijcsmc.com/docs/papers/january2013/V2I1201306.pdf
3. Alagu, S., Priyanka, A., Kavitha, G., & K, B. B. (2021). Automatic detection of acute lymphoblastic leukemia using UNET based segmentation and statistical analysis of fused deep features. Applied Artificial Intelligence, 35(15), 1952–1969. https://doi.org/10.1080/08839514.2021.1995974
4. Alagu, S., Priyanka, A., Kavitha, G., & K, B. B. (2021). Automatic detection of acute lymphoblastic leukemia using UNET based segmentation and statistical analysis of fused deep features. Applied Artificial Intelligence, 35(15), 1952–1969. https://doi.org/10.1080/08839514.2021.1995974
5. Saleem, S., Amin, J., Sharif, M., Anjum, M.A., Iqbal, M., & Wang, S. (2021). A deep network designed for segmentation and classification of leukemia using fusion of the transfer learning models. Complex & Intelligent Systems, 8, 3105 - 3120.
6. Vogado, L. H. S., Veras, R., Araújo, F. H. D., Silva, R., & Aires, K. R. T. (2018). Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. Engineering Applications of Artificial Intelligence, 72, 415–422. https://doi.org/10.1016/j.engappai.2018.04.024
7. Halim, N.H., Mashor, M.Y., & Hassan, R. Automatic Blasts Counting for Acute Leukemia Based on Blood Samples.
8. Najjar, F. H., Khudhair, K. T., Khudhair, Z. N., Alwan, H. H., & Al-Khaykan, A. (2023). Acute lymphoblastic leukemia image segmentation based on modified HSV model. Journal of Physics, 2432(1), 012020. https://doi.org/10.1088/1742-6596/2432/1/012020
9. malaysia fact sheets. (n.d.). Blobal Cancer Observatory. https://gco.iarc.fr/today/data/factsheets/populations/458-malaysia-fact-sheets.pdf
10. Eldridge. (n.d.). What Is a Blood Smear Test? verywellHealth. Retrieved October 5, 2023, from https://www.verywellhealth.com/blood-smear-uses-and-results-4586471

11. Tak Ng, Y. (n.d.). *Building a data pipeline*. Cs320. https://cs230.stanford.edu/blog/datapipeline/#best-practices

12. Team, M. (2024, March 26). White blood cells (WBCs). My Hematology. https://myhematology.com/white-blood-cells/white-blood-cells-wbcs-the-defenders-of-the-body/