Decoding Acute Lymphoblastic Leukemia: Insights from Convolutional Neural Networks and Pretrained Model

Jason Hendrawan
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia, 11530
jason.hendrawan@binus.ac.id

Jonathan Adrian
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia, 11530
jonathan.adrian001@binus.ac.id

Verrel Juanto Lukmana
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia, 11530
verrel.lukmana@binus.ac.id

Felix Indra Kurniadi
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia,11530
felix.indra@binus.ac.id

***Abstract*— Accurate detection of Acute Lymphoblastic Leukemia (ALL) is essential for prompt diagnosis and effective treatment. In this study, we introduce a deep learning method employing Convolutional Neural Networks (CNN) and pretrained models to enhance the precision of ALL detection. We leveraged a dataset comprising blood cell images for training and validation. Performance evaluations of three CNN models—ResNet-50, VGG16, and a bespoke CNN architecture—were conducted using metrics like accuracy, loss, and validation scores. Our findings reveal that VGG16 achieved a notable accuracy of 96.53%, while our custom CNN yielded 88.15%. Notably, VGG16, despite its high accuracy, exhibited signs of overfitting. In contrast, ResNet-50 underperformed, registering an accuracy of just 79.05%. This work underscores the potential of technological innovations in healthcare, especially in ALL diagnosis, and paves the way for more in-depth exploration of deep learning in cancer detection.**

Keywords*— Convolutional Neural Networks (CNN), Acute Lymphoblastic Leukemia (ALL), Image recognition, ResNet-50, VGG16

I. INTRODUCTION

Acute Lymphoblastic Leukemia (ALL) is a type of blood cancer characterized by an abundance of immature lymphocytes. In this condition, abnormal white blood cells, known as lymphoblasts, rapidly multiply in the bone marrow, progressively replacing healthy blood cells [1]. Common symptoms of ALL include weight loss, fatigue, bone and joint pain, and recurrent infections. The detection of ALL typically involves a blood test where a sample is taken and examined in a laboratory. However, automating the prognosis of ALL presents challenges due to the morphological similarities between cancerous and normal cells. The conventional approach to classifying ALL relies on the expertise of experienced medical professionals [2].

Therefore, the utilization of artificial intelligence (AI) techniques, such as the Convolutional Neural Network (CNN), offers significant potential for enhancing the efficiency and precision of diagnosing ALL. CNNs, advanced machine learning algorithms, have seen extensive applications in medical image analysis, especially in cancer detection. With CNNs, automated systems can learn to identify specific characteristics and patterns within blood cell images, which can indicate the presence of ALL [1].

Despite the widespread use of CNN technology in medical image processing, there is limited research on its application in detecting ALL. Thus, the objective of this study is to develop a machine for ALL detection based on CNN and to evaluate its diagnostic accuracy using blood cell images. The anticipated outcome of this research aims to further technological advancements in healthcare, particularly in improving the precision and efficacy of ALL diagnosis.

II. RELATED WORKS

Leukemia, a type of blood cancer affecting people of all ages, ranks among the leading causes of death worldwide. Traditional diagnostic techniques, such as blood and bone marrow examinations, are slow and painful. Consequently, there is a pressing need for non-invasive and rapid methods. In recent years, the application of AI in medicine has surged, with CNN being identified as a promising tool for the automated detection and diagnosis of various medical conditions, including ALL (Acute Lymphoblastic Leukemia). By leveraging CNNs, the detection of ALL can be rendered more precise and efficient, paving the way for earlier detection and optimized treatments. Consequently, numerous studies have explored the utility of various CNN types in ALL detection.

Safuan et al. (2020) determined that the VGG pretrained CNN model excelled in distinguishing between lymphoblast and non-lymphoblast cells for ALL detection, achieving 100% training accuracy and 99.13% testing accuracy [1]. Prellberg and Kramer combined pre-processing techniques, such as color normalization and data augmentation, with a CNN architecture to diagnose leukemia, achieving a weighted F1-score of 88.91% in the C-NMC online challenge [3]. Kumar et al. designed a model that accurately discerned cancer types with a 97.2% accuracy, surpassing other baseline methods. This model combined CNN with a Support Vector Machine

(SVM) to categorize various cancers, including ALL [2]. Jiang et al. introduced a diagnostic approach using the ViT-CNN ensemble model, achieving a classification accuracy of 99.03% in distinguishing between cancerous and normal cells in ALL diagnosis [4].

Machine learning algorithms paired with image processing have emerged as indispensable tools for early leukemia diagnosis. Rezayi et al. compared the validation accuracies of ResNet-50, VGG-16, and a proposed convolutional network, as well as machine learning methods like multilayer perceptron and random forest, with VGG-16 achieving the highest accuracy of 84.62\% [5]. Almadhor et al. introduced an ensemble prediction approach employing algorithms like K-Nearest Neighbor (KNN), SVM, Random Forest (RF), and Naive Bayes (NB) on the C-NMC leukemia dataset. Features were extracted using pretrained CNN-based Deep Neural Network (DNN) architectures, and feature selection techniques like ANOVA, Recursive Feature Elimination (RFE), and RF were employed. The results revealed SVM outperforming other algorithms with 90.0\% accuracy [6]. Ullah et al. (2021) showcased a non-invasive CNN-based technique using medical images for ALL diagnosis. Their solution integrated the Efficient Channel Attention (ECA) module with the VGG16 model to derive high-quality deep features from the image dataset [7]. Oliveira and Dantas proposed a CNN methodology for classifying lymphocyte images as malignant or benign, achieving an F1-score of 92.60%. Although not definitive for disease diagnosis, their method offers valuable assistance to oncologists [8]. Sampathila et al. introduced a machine learning methodology for ALL diagnosis using a peripheral blood smear image dataset. Their method employed CNN models like VGG16, ResNet50, and InceptionV3, achieving an impressive 98.72% accuracy in image classification [9].

Thanh et al. used CNNs to detect ALL for early diagnosis, ensuring timely treatment for ALL patients. Their proposed method achieved an impressive accuracy of 96.6% [10]. Mondal et al conducted research on automated ALL recognition employing deep CNN. They explored four distinct methods, including VGG-16, MobileNet, InceptionResNet-V2, DenseNet-121, and Xception. The Xception model outperformed other individual CNN models, delivering results of 85.9%-ACC, 86.0%-WFS, and 85.9%-BA [11]. Another VGG-16 based study by Muntasa and Yusuf integrated the method with SVM and MLP classification algorithms to detect ALL using color image principal object characteristics. Their results indicated a maximum accuracy of 81.54\% for various methods [12]. Albeeshi and Alshanbari also employed the VGG16 pretrained model with SVM and MLP classification algorithms. They targeted a rapid ALL diagnostic approach since timely diagnosis and treatment initiation are paramount. Their results showcased SVM classifier MLP's 77% accuracy at a 0.001 [13].

III. Methodology

The CNN method is used to detect ALL in blood cells. The steps involved in this methodology are: Dataset Collection, Data Pre-processing, Machine Training, Machine Testing, and Result Analysis.

A. Dataset

The dataset used in this study is the publicly available C-NMC-2019 dataset from the Kaggle repository. It consists of images from subjects with ALL and from healthy bone marrow (Hem) patients. Each image has a resolution of 450 × 450 pixels [11]. The leukemia dataset comprises a total of 10,661 images, organized into three main folders: fold0, fold1, and fold2. Each of these folders contains two sub-folders, labeled "all" and "hem" [6].

B. Data Preprocessing

The blood cell images will undergo preprocessing, which includes resizing them to a consistent size. Each image will then be labeled according to its status (1 or 0) to train the detection system. Finally, the labeled blood cell images will be split into two sets: training data and testing data, at a ratio of 60:40. This division facilitates training the CNN and evaluating its performance. These steps ensure proper processing of the blood cell images before they are used in the training and testing of the model [13].

C. Modeling

1) Convolutional Neural Network: The model comprises convolutional layers, activation layers, pooling layers, and fully connected layers [2]. In this research, we augmented the generic CNN model by adding another layer to the architecture. This addition enhances the model's capacity, depth, and performance compared to the generic model.

To train the model, the training data (xTrain and yTrain) is fed into the fit() method. The number of epochs is predetermined, dictating how many times the model will iterate through the training data [11]. During training, the model's performance is also monitored using the validation data.

2) VGG 16: The model is composed of convolutional layers from the VGG16 architecture, followed by a flatten layer to convert the feature maps into a vector, a dense layer for feature extraction, and a final dense layer for classification. The model is pretrained on the "imagenet" dataset and trained on the provided data for 35 epochs, with evaluation performed on the test data to assess its performance [7].

3) Resnet50: This model is composed of the ResNet50 architecture, which is a deep convolutional neural network known for its ability to capture intricate image features [6]. By leveraging pre-trained weights from the "imagenet" dataset, the model benefits from learned representations. The global average pooling layer reduces the spatial dimensions of the extracted features, enabling the model to capture global information efficiently. The compiled model utilizes the Adam optimizer and binary cross-entropy loss, and a summary is displayed to provide a comprehensive overview of the model's structure and parameter count.

D. Machine Training

After training the models, the model's performance is evaluated using the testing data (xTest and yTest). The evaluate() method is used to calculate the accuracy and loss of the model based on the testing data. The results of the evaluation, including the accuracy stored in the score variable as score[1]. Additionally, the trained model can be used to predict the classes or labels of the testing data using the predict() method. The predicted results are stored in the yPred variable.

E. *Machine Training*

During the results analysis phase, we begin by visualizing the model's accuracy during training. This involves plotting the accuracy history on both the training data and validation data. This graph helps observe how the model's accuracy improves with the number of epochs. Additionally, it is important to visualize the model's loss, which helps in evaluating its performance. By examining the loss history on the training data and validation data, we can analyze the changes in the model's loss and assess the effectiveness of the architecture and parameters used. Additionally, a graph that combines loss and accuracy can provide a more comprehensive understanding of the model's overall performance.

IV. Result and Analysis

The performance of CNN models for the detection of Acute ALL is evaluated using a dataset consisting of training and validation data. We are using python and keras as the library for building the models. The models were trained for 35 epochs with the following results

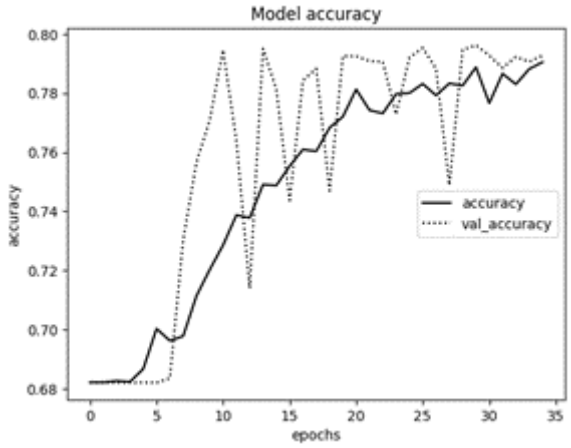

Fig. 1. Resnet50 training accuracy and validation accuracy

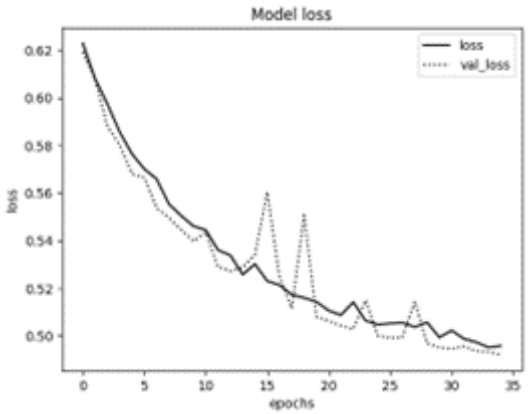

Fig. 2. Resnet50 training loss and validation loss

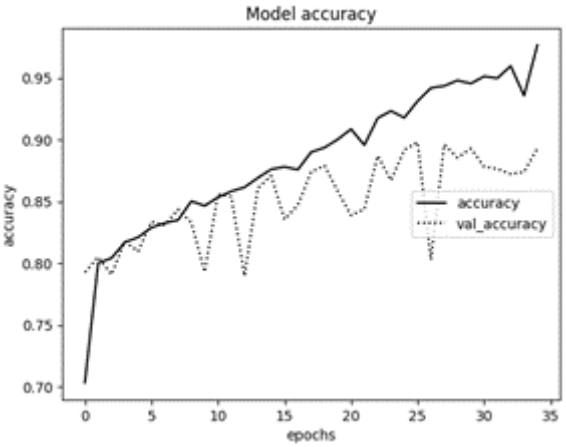

Fig. 3. VGG16 training accuracy and validation accuracy

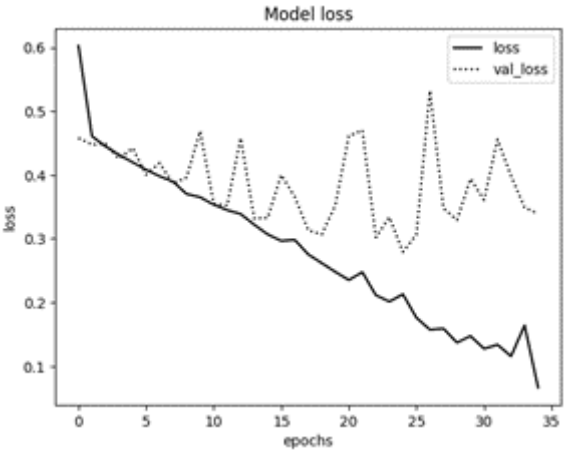

Fig. 4. VGG16 training loss and validation loss

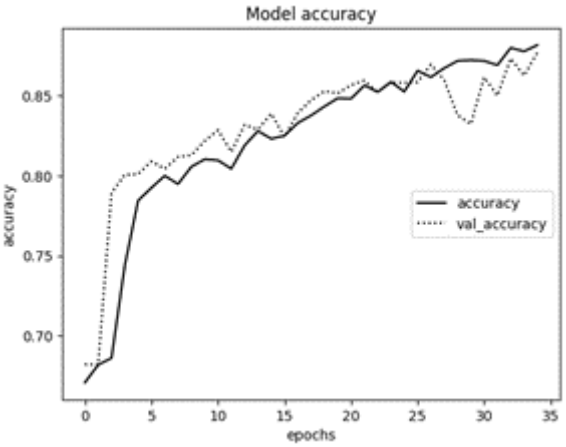

Fig. 5. proposed CNN training accuracy and validation accuracy

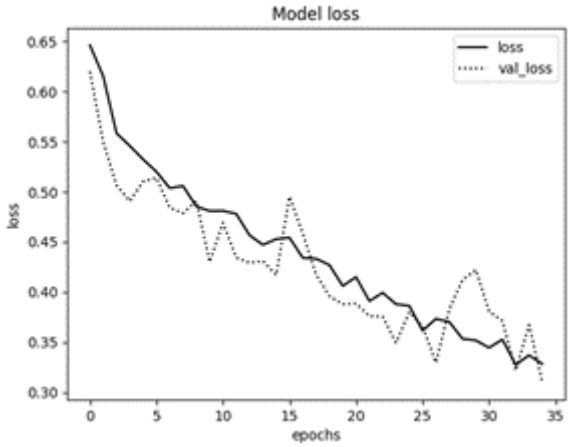

Fig. 6. proposed CNN training loss and validation loss

TABLE I. ACCURACY AND LOSS OF THREE DIFFERENT MODEL

	Resnet	VGG	CNN
Acc	0.7905	0.9653	0.8815
Loss	0.4957	0.0944	0.3282
Val_acc	0.7927	0.8966	0.8821
Val_loss	0.4919	0.4045	0.3096

The graphs and comparison tables presented above offer a comprehensive overview of the performance metrics, namely accuracy, loss, validation accuracy, and validation loss, across multiple models. The graph shows the trend of the loss and accuracies of each model over 35 epochs, while the table shows us the summary of performance metrics of each model, particularly in the context of accuracy, loss, validation accuracy, and validation loss. Upon analyzing the results of each model, we conclude that the best performing model is CNN. Both the CNN and VGG16 models consistently have similar accuracy results in the range of 0.85 to 0.89 across each iteration with VGG generally having the higher accuracy, but for the loss VGG16 outperforms CNN. But after analyzing the graph and performance metrics even further, the VGG model shows signs of overfitting.

V. CONCLUSION

In this study, we aimed to develop a deep learning-based detection system for ALL using CNN and pretrained models. The dataset used for training and testing was the C-NMC-2019 dataset, which consists of images from both ALL and healthy bone marrow patients. This dataset underwent preprocessing, including resizing and labeling, to ensure proper processing.

Three different models—CNN, VGG16, and ResNet50—were trained and evaluated. The CNN model consistently achieved accuracy results ranging from 0.85 to 0.89, while VGG16 demonstrated even higher accuracy. However, the VGG16 model exhibited signs of overfitting. In comparison, the ResNet50 model performed poorly, with average accuracies around 0.79. During the results analysis, we visualized accuracy and loss trends over 35 epochs. The CNN model displayed strong performance, boasting high accuracy and relatively low loss. Overall, the CNN model emerged as the most reliable performer among the evaluated models.

This study underscores the potential of CNN models and deep learning techniques for accurately detecting ALL. The findings emphasize the advantages of leveraging artificial intelligence in medical image analysis, recommending its integration into clinical settings for enhanced ALL detection. Future research could focus on refining these models and enlarging datasets to boost performance. In sum, this study advances ALL diagnostic technology and reinforces the significance of deep learning in medical imaging.

REFERENCES

[1] S. Nabilah Mohd Safuan, M. Razali Md Tomari, W. Nurshazwani Wan Zakaria, M. Norzali Hj Mohd, and N. S. Suriani, "Investigation of white blood cell biomaker model for acute lymphoblastic leukemia detection based on convolutional neural network," *Bull. Electr. Eng. Inform.*, vol. 9, no. 2, Apr. 2020, doi: 10.11591/eei.v9i2.1857.

[2] D. Kumar *et al.*, "Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 142521–142531, 2020, doi: 10.1109/ACCESS.2020.3012292.

[3] J. Prellberg and O. Kramer, "Acute Lymphoblastic Leukemia Classification from Microscopic Images Using Convolutional Neural Networks," in *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, A. Gupta and R. Gupta, Eds., in Lecture Notes in Bioengineering. Singapore: Springer Singapore, 2019, pp. 53–61. doi: 10.1007/978-981-15-0798-4_6.

[4] Z. Jiang, Z. Dong, L. Wang, and W. Jiang, "Method for Diagnosis of Acute Lymphoblastic Leukemia Based on ViT-CNN Ensemble Model," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, Aug. 2021, doi: 10.1155/2021/7529893.

[5] S. Rezayi, N. Mohammadzadeh, H. Bouraghi, S. Saeedi, and A. Mohammadpour, "Timely Diagnosis of Acute Lymphoblastic Leukemia Using Artificial Intelligence-Oriented Deep Learning Methods," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, Nov. 2021, doi: 10.1155/2021/5478157.

[6] A. Almadhor, U. Sattar, A. Al Hejaili, U. Ghulam Mohammad, U. Tariq, and H. Ben Chikha, "An efficient computer vision-based approach for acute lymphoblastic leukemia prediction," *Front. Comput. Neurosci.*, vol. 16, p. 1083649, Nov. 2022, doi: 10.3389/fncom.2022.1083649.

[7] M. Zakir Ullah *et al.*, "An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification," *Appl. Sci.*, vol. 11, no. 22, p. 10662, Nov. 2021, doi: 10.3390/app112210662.

[8] J. Maurício de Oliveira and D. Dantas, "Classification of Normal versus Leukemic Cells with Data Augmentation and Convolutional Neural Networks:," in *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Online Streaming, --- Select a Country --- : SCITEPRESS - Science and Technology Publications, 2021, pp. 685–692. doi: 10.5220/0010257406850692.

[9] N. Sampathila *et al.*, "Customized Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia Using Blood Smear Images," *Healthcare*, vol. 10, no. 10, p. 1812, Sep. 2022, doi: 10.3390/healthcare10101812.

[10] T. T. P. Thanh, C. Vununu, S. Atoev, S.-H. Lee, and K.-R. Kwon, "Leukemia Blood Cell Image Classification Using Convolutional Neural Network," *Int. J. Comput. Theory Eng.*, vol. 10, no. 2, pp. 54–58, 2018, doi: 10.7763/IJCTE.2018.V10.1198.

[11] C. Mondal *et al.*, "Ensemble of Convolutional Neural Networks to diagnose Acute Lymphoblastic Leukemia from microscopic images," *Inform. Med. Unlocked*, vol. 27, p. 100794, 2021, doi: 10.1016/j.imu.2021.100794.

[12] A. Muntasa and M. Yusuf, "Modeling of the Acute Lymphoblastic Leukemia Detection based on the Principal Object Characteristics of the Color Image," *Procedia Comput. Sci.*, vol. 157, pp. 87–98, 2019, doi: 10.1016/j.procs.2019.08.145.

[13] A. A. Albeeshi and H. S. Alshanbari, "Modeling of the Acute Lymphoblastic Leukemia Detection by ConvolutionalNeural Networks (CNNs)," *Curr. Med. Imaging Rev.*, vol. 19, no. 7, p. e141022210032, Jun. 2023, doi: 10.2174/1573405619666221014113907.