

Performance Evaluation of EfficientNetB3-Based Deep Learning Model for Classification of Acute Lymphoblastic Leukimia and Normal Blood Cells

Sayed Muchallil^{1,2*}, Maya Fitria², Ridha Arrahman² and Khairun Saddami²

¹ Department of Informatics, University of Oslo, Oslo, Norway

² Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia

Abstract

Acute Lymphoblastic Leukemia (ALL) is a rapidly progressing blood cancer that predominantly affects children and requires early and accurate diagnosis to improve patient survival rates. Traditional diagnostic methods rely heavily on manual examination of blood smear images by pathologists, which is not only time-consuming but also susceptible to human error and variability. To address this limitation, this study proposes an automated detection model based on deep learning, specifically employing the EfficientNetB3 convolutional neural network architecture. A publicly available dataset containing microscopic images of ALL and normal blood cells was used for training and evaluation. The images were preprocessed using normalization and augmentation techniques and resized to 300×300 pixels to align with the EfficientNetB3 input requirements. The model was trained using the Adam optimizer and monitored with EarlyStopping to prevent overfitting. Experimental results showed that the proposed model achieved an accuracy of 92.23%, precision of 92.75%, and recall of 95.57%, significantly outperforming conventional approaches such as Canberra distance, K-Nearest Neighbor, and ensemble CNN methods. In addition to the classification model, a web-based ALL detection system was developed to make the solution more accessible and user-friendly. The frontend was built using ReactJS, while the backend API, built with Flask, handles image input, model inference, and output delivery. The interface allows users to upload cell images, input patient names, and receive instant classification results along with confidence scores. This integrated system demonstrates a practical application of AI in medical diagnostics and holds potential for use in real-world, resource-limited clinical settings.

Paper History

Received May 02, 2025
Revised July 10, 2025
Accepted August 3, 2025
Published August 5, 2025

Keywords

Leukimia Classification;
CNN; Transfer Learning;
Image Classification;
ALL

Author Email

sayed.muchallil@usk.ac.id
mayafitria@usk.ac.id
ridharra2@gmail.com
ksaddami@usk.ac.id

1. Introduction

Leukemia, which affects blood and bone marrow and causes an excess of abnormal white blood cells (WBC), has become of common type of malignant blood cancer worldwide [1], [2]. In 2018, leukemia was identified as the fifteenth most frequent type of cancer globally. In Hong Kong, it is listed among the top ten deadliest cancers, with approximately 500 new cases diagnosed annually [3]. As a life-threatening disease, it is also among the most common cancers and ranks ninth in Indonesia [4].

In general, leukemia is broadly categorized into two types, namely acute and chronic, which are further classified into four subtypes: Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myelogenous Leukemia (CML) [5]-[7]. One major subtype, Acute Lymphoblastic Leukemia (ALL), primarily affects children and originates from immature lymphocytes [8]-[10]. Patients diagnosed with ALL are commonly characterized by the overproduction of bone marrow and lymphoblasts, resulting in immune system deterioration and could eventually cause death to the patients [11]-[13]. It is reported that roughly 25% of all juvenile malignancies and

75% of all leukemia cases in people under 20 are caused by ALL, where the highest prevalence occurs among children aged 2 to 5 [5]. Thus, an early diagnosis of ALL is essential for immediate treatment, such as medication, radiation therapy, and chemotherapy, to increase the survival rate.

Commonly, ALL diagnosis and early screening are conducted manually by analyzing the microscopic blood smear [14]-[16]. The evaluation is considered a time-consuming process as it requires trained laboratorians or pathologists role to identify the presence of abnormal patterns in the WBCs [17]-[18]. Moreover, the conventional screening is also expensive [15], [19]. So, the advancement of pattern recognition and machine learning technology to detect variations in the white blood cell count that can point to the presence of ALL can increase the precision and speed of diagnosis and surpass the traditional method.

The use of machine learning, particularly deep learning, in the classification of ALL has been extensively explored in recent studies. Previous approaches to ALL classification have utilized color-based object recognition methods employing distance metrics such as Euclidean

Corresponding author: Sayed Muchallil, sayed.muchallil@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.

DOI: <https://doi.org/10.35882/ijeemi.v7i3.113>

Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Distance, Manhattan, Canberra, and Chebyshev, achieving maximum accuracies ranging from 76.92% to 82.31%. When combined with classical classifiers, including Naive Bayes, Support Vector Machine, K-Nearest Neighbor, and Decision Tree, the highest accuracy reported was 88.25% [5].

In recent years, deep learning methods, especially convolutional neural networks (CNNs), have demonstrated superior performance in image-based diagnosis tasks. CNN architectures like AlexNet, VGG16, ResNet50, and Inception-ResNet-v2 have been widely used in ALL detection, often combined with transfer learning to handle the limited availability of annotated medical datasets [20]-[23]. These approaches have reported remarkable results, frequently exceeding 95% accuracy, with some studies claiming up to 99.7% accuracy using optimized pipelines.

Despite these promising results, critical limitations remain. Many of these models are computationally expensive and prone to overfitting, especially when trained on small or imbalanced datasets. Publicly available datasets like ALL-IDB and C-NMC 2019 are commonly used, and to address data imbalance, studies frequently adopt data augmentation techniques such as rotation, flipping, and SMOTE [24], [25]. However, these models still face challenges in terms of scalability, efficiency, and deployment in resource-constrained settings.

Most existing research relies on classical CNN architectures, with limited exploration of more efficient deep learning models. One such model is EfficientNet, which introduces a compound scaling method that balances network depth, width, and resolution, providing improved accuracy with significantly fewer parameters and lower computational cost [26]-[28]. Among the EfficientNet family, EfficientNetB3 offers a strong trade-off between performance and efficiency, making it suitable for real-time applications and mobile deployment in healthcare systems [29], [30]. Moreover, transfer learning plays a vital role in medical imaging applications, where large labeled datasets are scarce [31]. Utilizing pre-trained weights from large-scale datasets, EfficientNetB3 can learn high-level visual features and adapt to the subtle morphological differences in leukemic cells, thereby improving classification accuracy and generalizability with limited training data.

This research aims to address the current gap in the literature by developing an acute lymphoblastic leukemia classification model using the EfficientNetB3 architecture with transfer learning. The objective is to create a robust and computationally efficient model capable of achieving high accuracy while being suitable for deployment in real-world, low-resource clinical environments. Implementing EfficientNetB3 in medical image classification, our research aims to contribute a novel and impactful solution to the ongoing efforts in early leukemia detection. Moreover, a web-based system was also developed in this work, aiming to effectively demonstrate the capability of the model in classifying the ALL results.

The rest of this paper is structured as follows: Section 2 describes the material and methods used in this work, Section 3 presents the experimental results, Section 4 highlights the discussion based on the results, and Section 5 provides the conclusion of this work.

II. Materials and Methods

Fig. 1 shows the processes undertaken in this study, beginning with data collection, followed by data pre-processing, dataset preparation, model development, web-based development, and concluding with model evaluation. The following subsection provides a detailed explanation of each process.

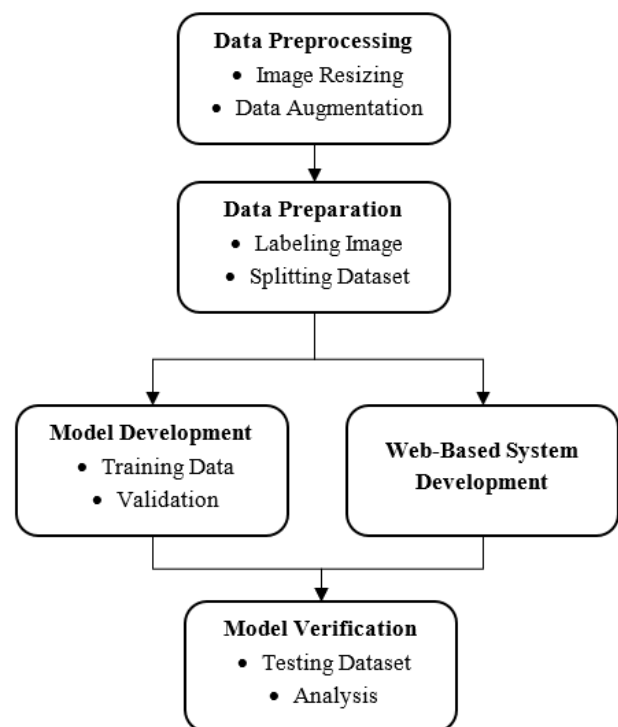


Fig. 1. The research methodology flowchart illustrates the key stages of this study, from data preprocessing to result analysis.

A. Data Processing

Approximately 25% of pediatric cancers constitute ALL [5]. Distinguishing the microscopic images of immature leukemic cells from normal cells remains challenging due to the similarity of their morphology. Fig. 2 depicts the sample images of normal cells (a) and ALL cells (b). As seen in the images, that normal lymphocyte is typically small to medium-sized, approximately 7-10 μm in diameter, normally possesses a dense, round nucleus occupying about 80-90% of the cell volume, and the cytoplasm appears as a thin rim around the [32], [33]. On the other hand, the ALL cell's size is generally larger, ranging around 10-20 μm in diameter, and it contains a high nucleus-to-cytoplasm (N:C) ratio with a large and irregularly shaped nucleus. Moreover, the cytoplasm displays a more abundant and deeply basophilic cytoplasm [32].

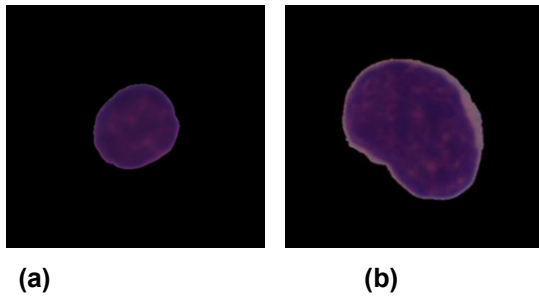


Fig. 2. Representative microscopic images showing morphological characteristics, (a) normal white blood cells, (b) all cells.

The dataset used in this study is sourced from Kaggle [34] and includes 12,528 microscopic images divided into two classes: 8,491 ALL images and 4,037 normal blood cell images. A pre-processing step was conducted to generate proper images for the model development. The stage was divided into two main tasks, namely, image resizing and image augmentation. The images were adjusted to a size of 300×300 pixels. This dimension was chosen to align with the input requirements of the EfficientNetB3 architecture, ensuring that the morphological features of the cells were maintained [35]. Subsequently, the resized images were augmented, aiming to provide dataset variability [36]. Augmentation included random rotation, flipping, and zooming. The process was carried out utilizing TensorFlow's ImageDataGenerator to increase model generalization. In addition, a normalization process was performed by scaling the RGB values of the images, aiming to enhance the stability of the training process. Pixel values were standardized using channel-wise ImageNet mean (RGB: [0.485, 0.456, 0.406]) and std ([0.229, 0.224, 0.225]) to match EfficientNetB3's pretraining. The normalized images of cell samples can be seen in Fig. 3.

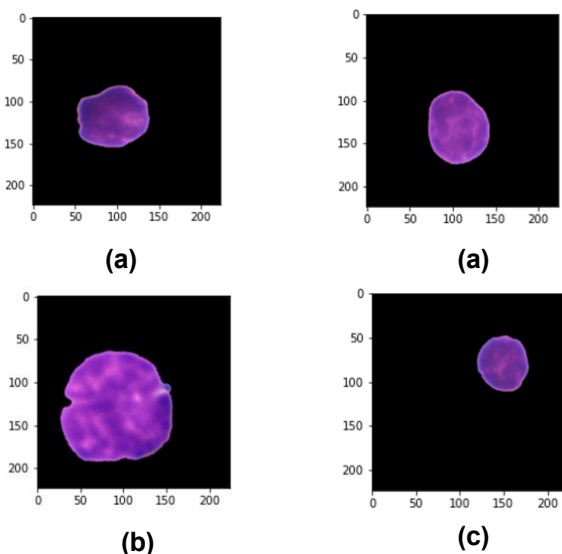


Fig. 3. Example of ALL cell images after implementing the augmentation method, (a) flipping, (b) random rotation, (c) zoom in, and (d) zoom out.

B. Dataset Preparation

The dataset used in this was blood smear images labeled as either Normal or ALL blood cells. The dataset is divided using stratified sampling (preserving class ratios) into three sub-datasets, namely the training, validation, and testing datasets, which were divided into 70%, 15%, and 15%, respectively. This process confirms that the model was evaluated on unseen data to provide a reliable assessment of its generalization performance. Table 1 exhibits the distribution of the dataset consisting of a number of training and testing data for both normal and ALL blood smear images. Additionally, to avoid data leakage to occur, all images from the same patient were grouped in one subset. The augmented versions of dataset were only generated from training set, and normalization states were solely computed from training data.

Table 1. Overview of total dataset used in the model development, including the number of normal and ALL samples.

Data	ALL	Normal	Total
Training	5.818	2.711	8.529
Validation	1.454	678	2.132
Testing	1.219	648	1.867
Total	8.491	4.037	12.528

C. Model Development

The model was developed on the provided prepared dataset employing EfficientNetB3 architecture and using the TensorFlow framework. EfficientNetB3 was selected in this work due to its optimal trade-off between accuracy, computational cost, and resource efficiency [37], [38]. The implementation of EfficientNetB3 is considered suitable for classification tasks that involve limited data and practical deployment needs, especially in medical image classification [39].

Unlike the conventional scaling methods (Fig. 4. a-c), which scale the image width, increase the number of layers, or accept higher image resolution, the EfficientNet architecture has a convolutional layer that enables feature maps extraction and accepts image resolution up to 300x300 (Fig. 4. d). The base EfficientNetB3 layers were frozen initially. After convergence, top 50% layers were unfrozen for fine-tuning. On top of that, the EfficientNet family is able to scale all three dimensions simultaneously, allowing for deeper and more efficient machine learning [39]. The architecture of EfficientNetB3 was modified by adding a final sigmoid activation layer, enabling the model to perform binary classification tasks effectively. Sigmoid activation is defined as Eq. (1), where f_x is sigmoid activation function and x is the value from previous layer.

$$f_x = \frac{1}{1 + e^{-x}} \quad (1)$$

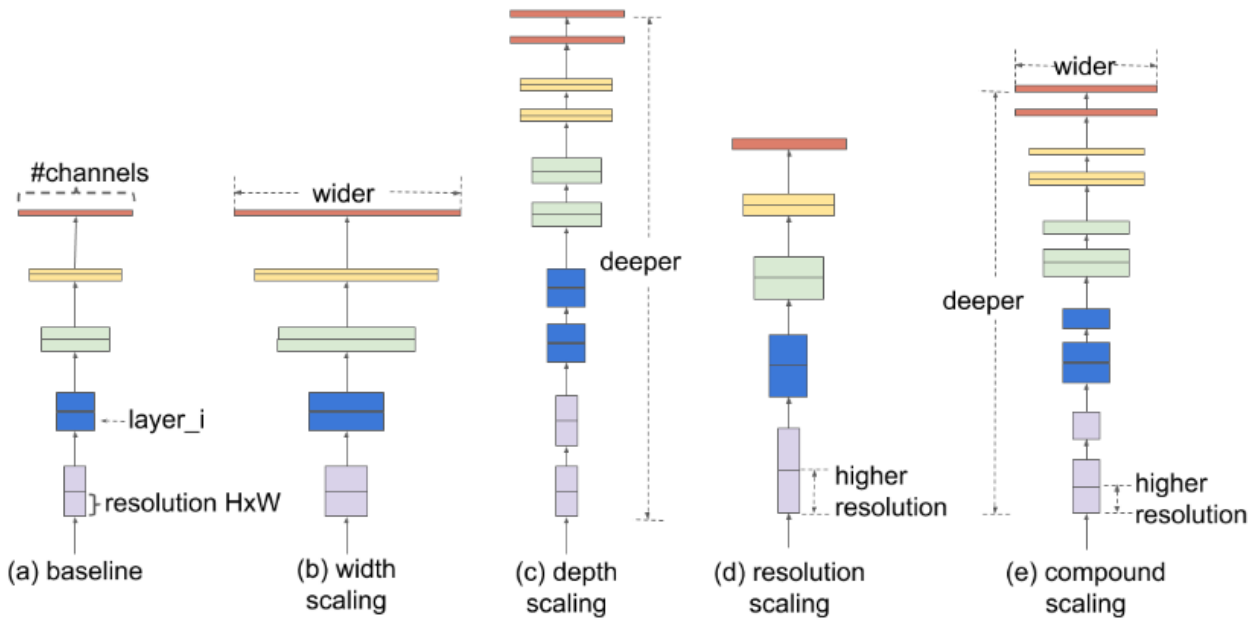


Fig. 4. Model scaling comparison which differs into baseline (a), width scaling (b), depth scaling (c), resolution scaling (d), and compound scaling which used in EfficientNet family (e).

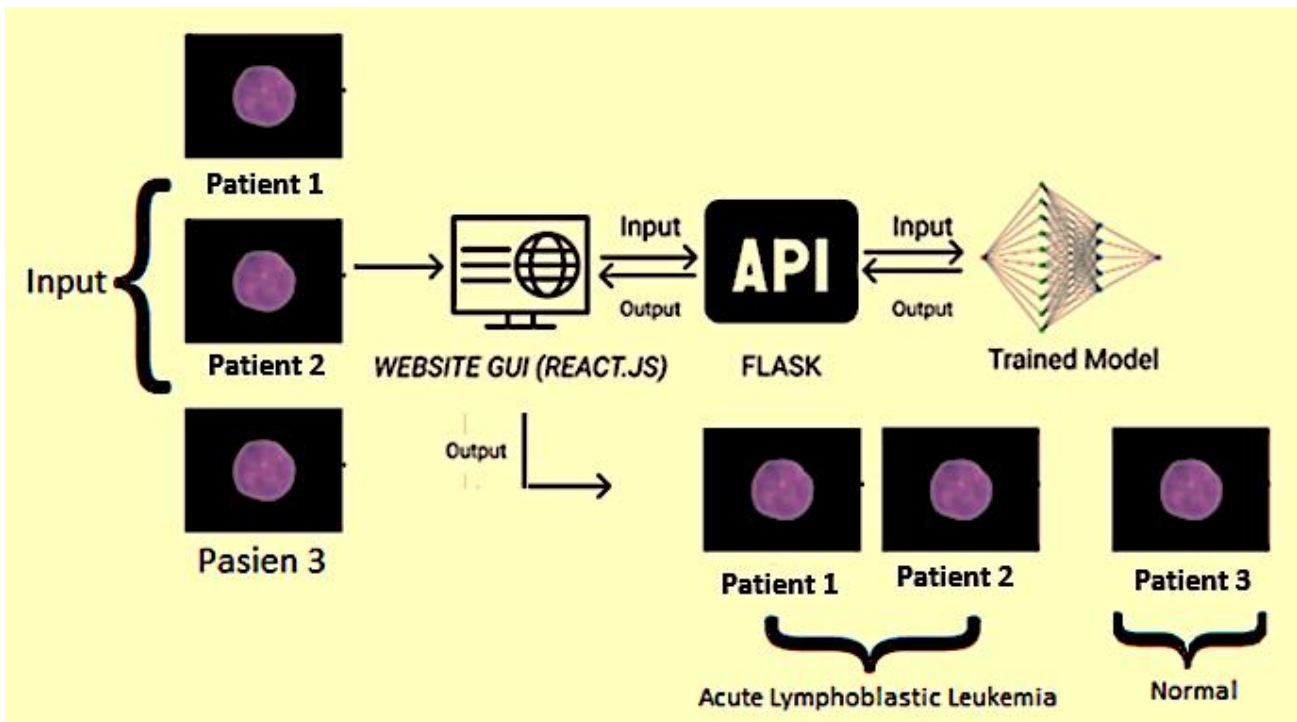


Fig. 5. The integrated workflow of the proposed ALL detection systems which demonstrates a seamless flow from image upload to model and classification result delivery.

The EfficientNetB3 was trained on the dataset with the Adam optimizer and binary cross-entropy as the loss function. Binary cross-entropy is defined as Eq. (2), where N is the number of class, p_i and q_i is the label and the sigmoid results. If the class is two, the cross-entropy known as binary cross entropy and formulated as Eq. (3).

$$CE = - \sum_{i=1}^N p_i \log q_i \quad (2)$$

$$CE = -p_i \log q_i - (1 - p_i) \log(1 - q_i) \quad (3)$$

The EfficientNetB3 was trained on the dataset with the Adam optimizer and binary cross-entropy as the loss function. The model development was conducted over 10 epochs with a batch size of 4. Additionally, an early stopping callback was applied during the training process, aiming to prevent the overfitting from occurring. The detail of the parameter set for model development is exhibited in Table 2.

Table 2. Detailed parameter set that was configured for ALL model development

Parameter Set	
Batch Size	4
Learning Rate	3×10^{-5}
Epoch	10
Height x Width	224 x 224
Channels	3
Weight Decay	10^{-4}
Optimizer	Adam
Label Smoothing	0.1

D. Web-based System Development

The development of the web-based ALL detection system consists of several integrated stages, namely user interface design, development of backend API, server setup, and integration of the pre-trained EfficientNetB3 architecture model. Fig. 5. displays the end-to-end of the proposed ALL system flow, which begins with the image uploading process of patient blood smear images as input through the website interface, which was developed with ReactJS. The Flask-based API was utilized as the middleware API to receive the uploaded images and also host the trained EfficientNetB3 model. Subsequently, the model processes the inputs, returns the predictions to the front end in a JSON response, and displays them back on the website.

Fig. 6 shows the initially constructed wireframe of the ALL detection system, including a drop zone area for uploading white blood cell images (Fig. 6 (a)), a form input field for the patient's name of each image (Fig. 6 (b)), two control buttons for adding more inputs and for submitting an image for classification process. ReactJS, as the open-source JavaScript library, was employed to build the system frontend. ReactJs was considered for it well-suited for a dynamic and responsive user interface construction [40].

The backend API of the proposed ALL system was established Flask API, as it is a lightweight Python microframework serving the detection model [41], [42]. The Flask API received the uploaded images associated with the name of the patient from the front end, classified the images, and returned the results in the form of JSON responses back to the front end. The pre-trained EfficientNetB3 model performed the feature extraction and classification using the TensorFlow framework. The classification was conducted based on the morphological features in the white blood cell images. The Flask API as the middleware allows easy communication between the web front end and the model inference engine. Code. 1.

shows the sample of JSON output, which ensures that both classification results and model confidence are transparently communicated in the real-time display on the web.

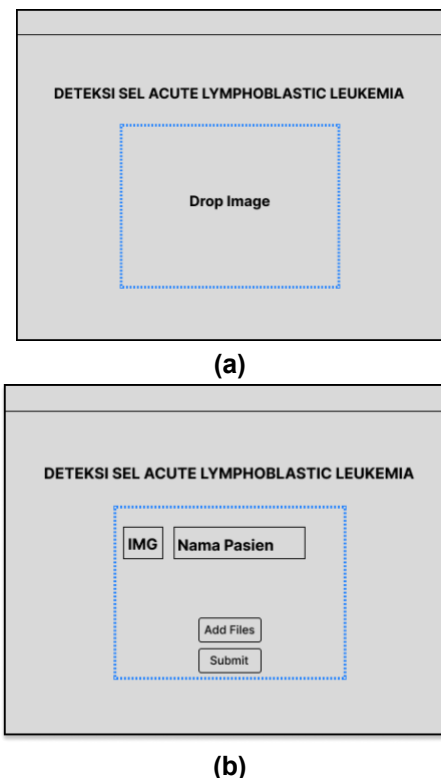


Fig. 6. The integrated workflow of the proposed ALL detection systems which demonstrates a seamless flow from image upload to model and classification result delivery (a) Image Upload Form, (b) Patient Name Input Field.

For development purposes, a localhost was configured for system hosting. Next, it allows the front end to send requests to the Flask API. The server processes the requests, performs predictions, and sends the results back to the web client to be displayed to the users. The rapid testing and debugging are also facilitated within this setup without requiring cloud infrastructure (Code 1).

```
{
  "data": {
    "Nama Pasien 1": {
      "output": "ALL",
      "pred": 98
    }
  }
}
```

Code. 1. Example of JSON output generated from API which shows the predicted lable and confidence score.

E. Model Verification

Model evaluation was conducted to assess the performance of the trained model in detecting ALL cells. The evaluation process utilized the test dataset, and the results are presented in the form of a confusion matrix as depicts in Fig. 7. The matrix values include True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) [43], [44].

		Predicted	
		0	1
Actual	0	TN	FP
	1	FN	TP

Fig. 7. Example of JSON output generated from API which shows the predicted label and confidence score.

Once the value of TP, FP, TN, and FN are provided, a comprehensive evaluation was carried out with three primary metrics computation, namely Recall, Precision, and Accuracy. Recall (Eq. (4) [44]) measures the model's ability to correctly identify ALL positive cases among all actual positive cases. Precision (Eq. (5) [44]) evaluates the proportion of true positive predictions out of all predicted positive results, indicating how accurately the model identifies patients with ALL. Accuracy (Eq. (6) [44]) reflects the overall correctness of the model in classifying both ALL and normal cells.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

In the realm of medical image classification, relying solely on accuracy is deemed inadequate, as false positives and false negatives can lead to significantly different clinical outcomes. Therefore, it is essential to employ a wider range of evaluation metrics for a more balanced and reliable assessment. To enhance the evaluation process, the model's performance was also analyzed by looking at specificity and the F1-Score. Specificity focuses on ensuring that the model does not incorrectly classify healthy cells as acute lymphoblastic leukemia (ALL). This metric quantifies the proportion of actual negative cases, normal cells, that the model accurately identifies as negative. The calculation for specificity is provided in Eq. (7). Additionally, the F1-Score was included as a performance evaluation metric because it effectively combines the mean of precision and recall into a single value, making it particularly useful for addressing class imbalance. The formula for calculating the F1-Score is presented in Eq. (8).

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

III. Results

The training phase was conducted over 10 epochs using the EfficientNetB3 architecture. The output layer was configured with a sigmoid activation function for binary classification, which is 0 for the normal cell and 1 for ALL cells (Fig. 8). The Adam optimizer was employed to optimize the gradient during training. The early stopping mechanism was also implemented to prevent overfitting from occurring. The finding shows that the model demonstrated robust learning behavior, achieving a training accuracy of 93% and a validation accuracy of 91%, with minimal difference between training and validation loss. No indication of overfitting was observed, as both the training and validation loss remained stable throughout the training epochs. This stability indicates that the model generalizes well to unseen data and is not merely memorizing the training samples.

However, when the training was extended beyond 10 epochs, the validation loss began to increase, as illustrated in Fig. 9. The increasing error of validation is an indicator of overfitting, where the model did not learn to generalize features but started memorizing the training data instead of learning generalizable features. Therefore, the number of epochs was limited to 10 to maintain optimal performance and ensure that the model could accurately classify new and unseen data. The model was further evaluated using the test dataset, and its performance was summarized in the confusion matrix shown in Fig. 10. with the following values: True Positive (TP) = 1165, True Negative (TN) = 557, False Positive (FP) = 91, and False Negative (FN) = 54. Based on these values, the results shows that the accuracy of 92.23%, indicating that 92.23% of all predictions made by the model were correct across both ALL and normal cell classes. The precision was calculated, yielding the value of 93.75%. This means that the model accurately predicted 93.75% of all instances labeled as ALL, showing strong reliability in identifying true leukemia cases. The recall results shows that the model was able to correctly identify 95.75% of all actual ALL cases, demonstrating a high sensitivity to leukemia-positive samples. Moreover, the specificity and F1-score obtained with value of 85.95% and 94.14%, respectively. The results show that the model accurately identified 85.95% of the actual normal cells.

This suggests that the model was quite effective in reducing unnecessary alerts for healthy patients. The value of 94.14% in F1-score demonstrates a strong balance between precision and recall. This indicates that the model effectively maintains both sensitivity and precision in identifying ALL cells. The results indicate that the deep learning model is more particular when compared to traditional machine learning approaches. For instance, the K-Nearest Neighbor (KNN) algorithm achieved an accuracy of only 88.25% [2]. Thus, the proposed EfficientNetB3-based model not only surpasses traditional methods in terms of performance but also exhibits strong generalization capabilities across varied sample distributions. EfficientNet's compound scaling allows it to extract features from high-resolution medical

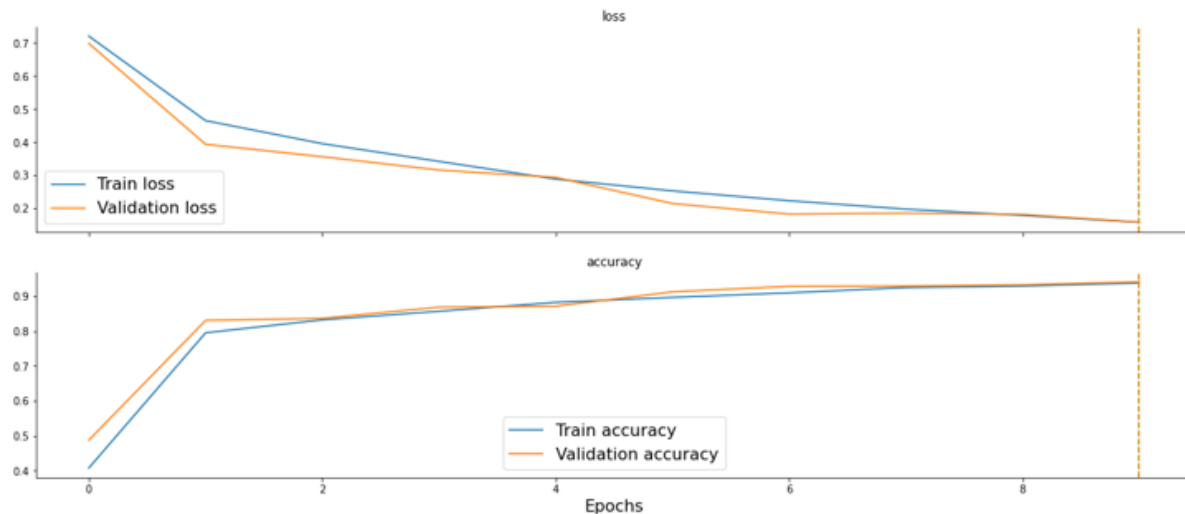


Fig. 8. Graph showing training loss, validation loss, training accuracy and validation accuracy over 10 epoch using EfficientNetB3 model.

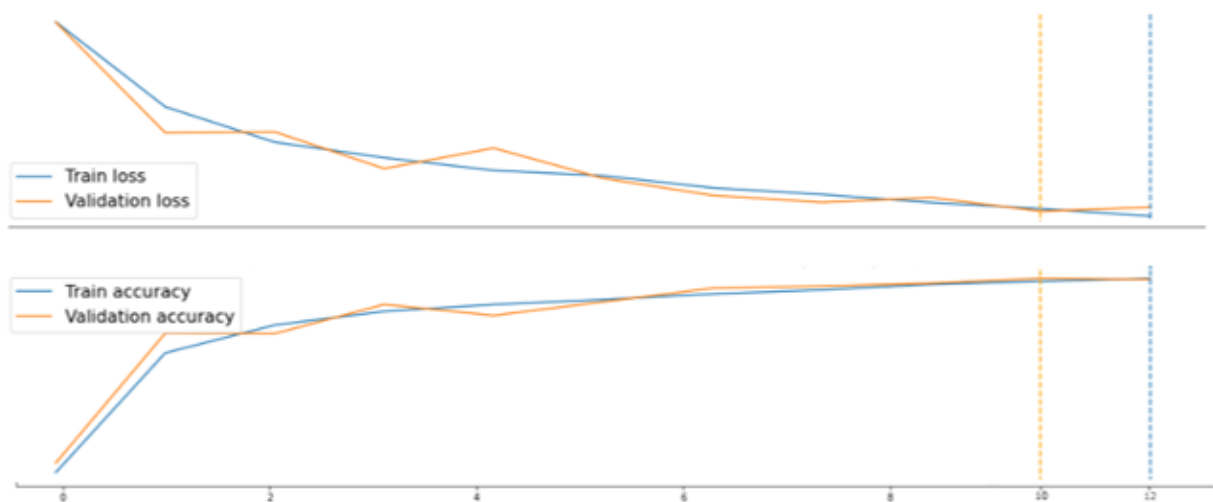


Fig. 9. Graph displaying training loss, validation loss, training accuracy and validation accuracy trend over 12 epochs, indicating overfitting in the model.

images efficiently. The high recall indicates its suitability for medical screening, minimizing false negatives

The user interface of the proposed web-based ALL detection system is designed to provide a simple yet functional workflow for clinical or laboratory users. Fig. 11 (a) shows the interface feature for uploading images (dropzone); users can drag and drop microscopic blood smear images for classification. Moreover, the interface also allows multiple file uploads at once, streamlining the process for cases with several patient samples. Next, the system automatically generates a preview of each image to ensure the correct files are selected. There is also a form to input the patient name after the previous phase, as shown in Fig. 11 (b). so that the uploaded images and patient name can be associated. This input ensures that classification results are linked to the correct patient record.

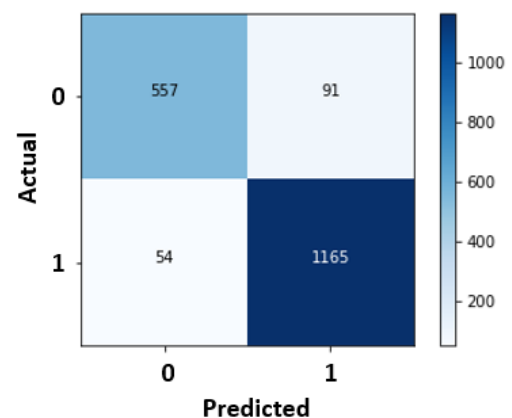


Fig. 10. Confusion matrix illustrating model performance in classifying ALL cells and normal white blood cells.

After processing, the results are displayed as shown in Fig. 11 (c), where the system displays the classification output for each image, indicating whether the sample is predicted as ALL or Normal, along with the confidence score. The layout is built to be responsive and interactive, providing a user-friendly experience and enabling immediate access to prediction outcomes in a clinical setting.

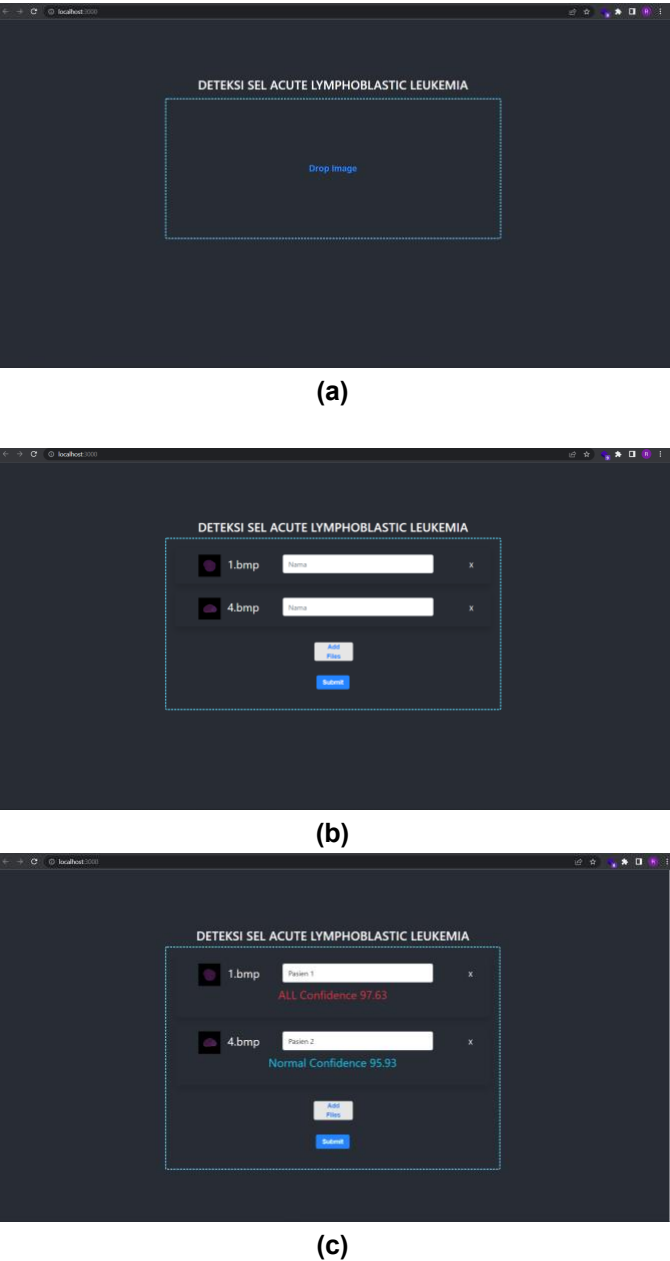


Fig. 11. User interface of the proposed web-based ALL detection system with image upload section (a), patient name input form (b), and displayed classification output results (c).

IV. Discussion

The experimental results demonstrate the effectiveness of the proposed ALL detection system using the EfficientNetB3 architecture. The model achieved a classification accuracy of 92%, significantly outperforming

traditional machine-learning approaches and earlier feature-based methods. As seen in Table 3, classical distance-based classifiers such as Canberra, Euclidean, Manhattan, and Chebyshev yielded accuracies ranging from 76.92% to 82.31%, while ensemble and K-Nearest Neighbor (KNN) methods improved performance up to 86.2% and 88.25%, respectively. However, these conventional techniques rely heavily on handcrafted features and often struggle to generalize well on complex image data.

By contrast, the EfficientNetB3 model, trained with transfer learning and optimized using the Adam optimizer and EarlyStopping, was able to learn deep hierarchical features from the input images. The training process reached 93% training accuracy and 91% validation accuracy with minimal overfitting, as confirmed by the stable training and validation loss curves. Further experimentation showed that training the model for more than 10 epochs led to increased validation loss, indicating overfitting and reduced generalization, which justifies limiting the training to 10 epochs.

Table 3. Comparison chart of classification accuracy across multiple methods, highlighting EfficientNetB3 as the most accurate among traditional and deep learning techniques.

Methods	Accuracy(%)
K-Nearest Neighbor [2]	88.25
DenseNet201 [45]	98.45
ResNet50 [46]	98.45
VGG-16 [47]	99.66
DenseNet-201 [47]	99.90
Pre-trained EfficientNetB3	92.23

To better understand the performance of the proposed EfficientNetB3-based model, we compared it with several recent studies that have applied deep learning techniques to the classification of Acute Lymphoblastic Leukemia (ALL). For instance, Chand et al. [45] utilized a ResNet-50 architecture on a small dataset of just 108 blood smear images, which included 59 healthy blood images and 49 ALL images, reporting an impressive accuracy of 100%. However, the limited size of this dataset raises concerns about overfitting and restricts the model's generalizability in clinical settings. Similarly, Rodrigues et al. [46] employed a ResNet-50V2-based model trained on 360 images (180 ALL and 180 normal) and achieved an accuracy of 98.46%. While their dataset is more balanced, it is still relatively small, and the absence of external validation raises questions about its robustness in real-world applications. In contrast, El-Seddek et al. [47] used VGG16 and DenseNet-201 architectures on a heavily augmented dataset of 20,000 synthetic images, achieving high accuracies of 99.66% and 99.9%, respectively. However, an overreliance on data augmentation may not accurately capture the true variability present in actual medical samples, potentially introducing bias and diminishing model reliability in clinical settings.

In comparison to these studies, our EfficientNetB3-based model achieved an accuracy of 92.23% on a

curated dataset with moderate augmentation, demonstrating strong generalization and stable performance without excessive data synthesis. Furthermore, EfficientNetB3 provides superior computational efficiency compared to VGG16 and DenseNet201, making it more suitable for real-time deployment in low-resource or point-of-care environments. These comparisons underscore that while other models may report higher accuracies, they often do so under constrained or artificially controlled conditions. In contrast, our model strikes a well-balanced trade-off between performance, reliability, and practical deployment in realistic clinical workflows.

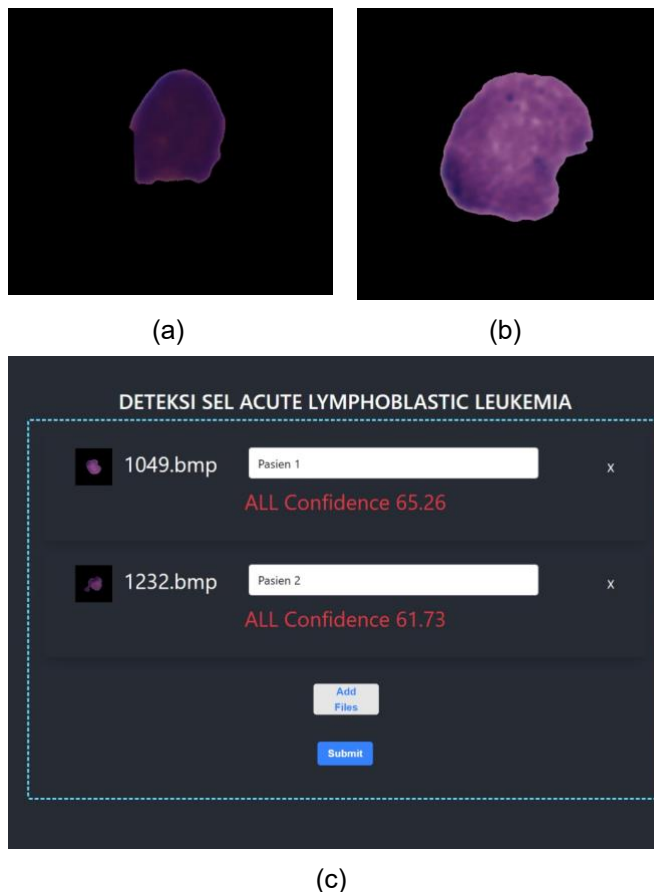


Fig. 12. Examples of misclassified ALL cell images, showing low-quality regions that contributed to erroneous predictions, (a) (b) Low Quality Cell Image, (c) Misclassified Screen Result.

The final performance was validated using a test dataset, and the confusion matrix showed strong classification capability with True Positives = 1165, True Negatives = 557, False Positives = 91, and False Negatives = 54. These results reflect the model's high precision, recall, and accuracy in detecting Acute Lymphoblastic Leukemia (ALL) cells from blood smear images. However, the misclassified samples frequently featured blurred boundaries or overlapping cells. Fig. 12 illustrates examples of misclassified ALL images, indicating that the integration of segmentation or attention mechanisms could enhance performance. From an

analytical point of view, the significant improvement in accuracy using EfficientNetB3 can be attributed to its compound scaling approach, which balances network depth, width, and input resolution more effectively than traditional CNN architectures. Unlike earlier models that either increase depth or width independently, EfficientNet applies a principled compound scaling strategy, allowing the model to capture more relevant morphological features in medical images with fewer parameters and reduced computational overhead. This result is especially beneficial for medical imaging tasks like ALL detection, where datasets are often limited and computational resources in clinical settings may be constrained.

To validate the statistical significance of the results of our proposed pre-trained EfficientNetB3 model, the McNemar's test was performed to compare the baseline methods (KNN and ensemble classifiers). The test obtained $p < 0.01$ for all comparisons, indicating statistically significant improvements. Confidence intervals, which is 95%, were calculated as: accuracy in the range of 91.2% - 93.1%, precision between 92.3% - 94.9%, and recall between 94.2% - 96.7%.

Overall, the use of EfficientNetB3 in our experiment is one of the solutions for key limitations of earlier methods by offering high accuracy while maintaining computational efficiency. This model is a compelling solution for practical use, particularly in healthcare settings where resources are limited. It stands out due to its reliable performance on unseen data and its modest resource requirements, which are notably less demanding than those of deeper CNNs like ResNet or Inception.

Despite the encouraging performance of the proposed EfficientNetB3-based model in detecting ALL, several important limitations should be acknowledged. First, the model was trained and tested using only one publicly available dataset, which may restrict its generalizability to other clinical environments, including variations in imaging tools, staining protocols, or differing patient conditions. Second, while data augmentation was employed to enhance data variety and model robustness, the real-world dataset remains limited, particularly regarding variations in cell morphology, such as rare cell types, low-quality samples, and blurry images. This limitation could impact the model's ability to generalize to typical low-quality samples encountered in actual clinical settings.

Additionally, the evaluation was conducted based on a single split of the data into training, validation, and testing sets. A more robust approach would involve using cross-validation or testing the model on external datasets to better assess its generalizability [44]. The model also lacks interpretability mechanisms, such as Grad-CAM or saliency maps, which are increasingly important in clinical AI for understanding and validating predictions. Finally, while the model demonstrated high performance, its clinical relevance and impact on actual diagnostic workflows were not tested or compared with expert hematologists, an aspect that should be addressed in future validation studies.

While the proposed ALL detection system is not intended to replace clinical expertise, its performance suggests that it could potentially assist hematologists by providing fast, consistent, and preliminary assessments, thereby helping to reduce clinician workload. This highlights a valuable opportunity for integration into clinical workflows, particularly in resource-limited settings where hematologists may not always be readily available.

V. Conclusion

This study presents a deep learning-based system for the detection of ALL utilizing the EfficientNetB3 architecture. The proposed model achieved a training accuracy of 93% and a testing accuracy of 92.23%, indicating high performance and strong generalization capability. Our experiment results not only in building a model but also in a fully functioning system that works with a web-based interface, enabling practical application for early ALL detection through microscopic blood cell images. Experimental results demonstrate that the proposed model outperforms traditional machine learning approaches. Specifically, while the K-Nearest Neighbor (KNN) algorithm reached an accuracy of only 88.25%, the EfficientNetB3-based model surpassed this benchmark, confirming its effectiveness in real-world scenarios. However, our system also has some drawbacks, such as adequate memory resources during operation, including the need for server hardware for running the web server. Future work may involve model optimization and deployment in lightweight environments to enhance scalability and accessibility.

References

- [1] C. K. Tebbi, "Etiology of acute leukemia: A review," *Cancers*, vol. 13, no. 9, p. 2256, 2021, doi: 10.3390/cancers13092256.
- [2] S. O. Heriawati, R. A. Revanda, C. Fatichah, and N. Suciati, "Blood cells classification for identification of acute lymphoblastic leukemia on microscopic images using image processing," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 6, pp. 1–10, 2021, doi: 10.1109/IES53407.2021.9593939.
- [3] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceeding of the 36th International Conference on Machine Learning (ICML)*, pp. 6105–6114, 2019.
- [4] C. Mondal, M. K. Hasan, M. T. Jawad, A. Dutta, M. R. Islam, and M. A. Awal, "Acute lymphoblastic leukemia detection from microscopic images using weighted ensemble of convolutional neural networks," *Engineering Research Express*, vol. 3, no. 1, p. 015037, 2021, doi: 10.48550/arXiv.2105.03995.
- [5] P. K. Das and S. Meher, "An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia," *Expert Systems with Applications*, vol. 183, p. 115311, 2021, doi: 10.1016/j.eswa.2021.115311.
- [6] R. P. Gale, "Radiation and leukaemia: Which leukaemias and what doses?," *Blood Reviews*, vol. 58, p. 101017, 2023, doi: 10.1016/j.blre.2022.101017.
- [7] H. J. Yook, J. H. Son, Y. H. Kim, J. H. Han, J. H. Lee, Y. M. Park, N. G. Chung, H. J. Kim, and C. H. Bang, "Leukaemia cutis: Clinical features and outcomes of 56 patients," *Acta Dermato-Venereologica*, vol. 102, p. adv00647, 2022, doi: 10.2340/actadv.v102.1123.
- [8] C. Matek, S. Schwarz, K. Spiekermann, and C. Marr, "Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks," *Nature Machine Intelligence*, vol. 1, pp. 538–544, 2019, doi: 10.1038/s42256-019-0101-9.
- [9] S. W. Brady et al., "The genomic landscape of pediatric acute lymphoblastic leukemia," *Nature Genetics*, vol. 54, no. 9, pp. 1376–1389, 2022, doi: 10.1038/s41588-022-01159-z.
- [10] H. T. K. Nguyen, M. A. Terao, D. M. Green, C. H. Pui, and H. Inaba, "Testicular involvement of acute lymphoblastic leukemia in children and adolescents: Diagnosis, biology, and management," *Cancer*, vol. 127, no. 17, pp. 3067–3081, 2021, doi: 10.1002/cncr.33609.
- [11] A. S. Duffield, C. G. Mullighan, and M. J. Borowitz, "International consensus classification of acute lymphoblastic leukemia/lymphoma," *Virchows Archiv*, vol. 482, pp. 11–26, 2023, doi: 10.1007/s00428-022-03448-8.
- [12] P. K. Das, A. Pradhan, and S. Meher, "Detection of acute lymphoblastic leukemia using machine learning techniques," in *Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication*, vol. 749, E. S. Gopi, Ed. Singapore: Springer, pp. 373–382, 2021, doi: 10.1007/978-981-16-0289-4_32.
- [13] A. R. Revanda, C. Fatichah, and N. Suciati, "Classification of acute lymphoblastic leukemia on white blood cell microscopy images based on instance segmentation using Mask R-CNN," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 5, pp. 1–10, 2022.
- [14] S. Rezayi, N. Mohammadzadeh, H. Bouraghi, S. Saeedi, and A. Mohammadpour, "Timely diagnosis of acute lymphoblastic leukemia using artificial intelligence-oriented deep learning methods," *Computational Intelligence and Neuroscience*, vol. 2021, p. 5478157, 2021, doi: 10.1155/2021/5478157.
- [15] R. Baig, A. Rehman, A. Almuhaimeed, A. Alzahrani, and H. T. Rauf, "Detecting malignant leukemia cells using microscopic blood smear images: A deep learning approach," *Applied Sciences*, vol. 12, no. 13, p. 6317, 2022, doi: 10.3390/app12136317.
- [16] I. Abunadi and E. M. Senan, "Multi-method diagnosis of blood microscopic sample for early detection of acute lymphoblastic leukemia based on deep learning and hybrid techniques," *Sensors*, vol. 22, no. 4, p. 1629, 2022, doi: 10.3390/s22041629.
- [17] A. Mittal, S. Dhalla, S. Gupta, and A. Gupta,

- [18] G. Zini, O. Barbagallo, F. Scavone, and M. C. Béné, "Digital morphology in hematology diagnosis and education: The experience of the European LeukemiaNet WP10," *International Journal of Laboratory Hematology*, vol. 44, pp. 37–44, 2022, doi: 10.1111/ijlh.13908.
- [19] Z. F. Mohammed and A. A. Abdulla, "An efficient CAD system for ALL cell identification from microscopic blood images," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 6355–6368, 2021.
- [20] A. Muntasa, R. T. Wahyuningrum, Z. Tuzzahra, D. R. Anamisa, and M. K. Sophan, "AlexNet architecture modification to classify Acute Lymphoblastic Leukemia images," in *AIP Conference Proceedings*, vol. 2952, no. 1, AIP Publishing, July 2024, doi: 10.1063/5.0211911.
- [21] S. Chand and V. P. Vishwakarma, "A novel deep learning framework (DLF) for classification of acute lymphoblastic leukemia," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 37243–37262, 2022, doi: 10.1007/s11042-022-13543-2.
- [22] M. Narendra and S. Negi, "Enhancing Leukemia Diagnosis through Machine Learning: A Customized ResNet50 Approach," in *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, pp. 728–733, 2024, doi: 10.1109/InCACCT61598.2024.10551130.
- [23] G. Sharma, V. Anand, and S. Gupta, "Utilizing the inception-ResNetV2 pre-trained model for binary classification of leukemia cells: An advanced approach to hematological diagnostics," in *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, pp. 1–6, 2023, doi: 10.1109/GCAT59970.2023.10353360.
- [24] A. Muntasa, A. Sugiarti, R. T. Wahyuningrum, M. A. Ghufro, R. M. Almohamedh, A. Motwakel, Y. P. Asmara, D. A. Dewi, and Z. Tuzzahra, "A Novel Hybrid Convolutional and Network Encapsulation Approach in EfficientNetV2-S Architecture for Acute Lymphoblastic Leukemia Classification," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 6, 2024, doi: 10.22266/ijies2024.1231.52.
- [25] M. Zolfaghari and H. Sajedi, "A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells," *Multimedia Tools and Applications*, vol. 81, no. 5, pp. 6723–6753, 2022, doi: 10.1007/s11042-022-12108-7.
- [26] J. Padhi, L. Korada, A. Dash, P. K. Sethy, S. K. Behera, and A. Nanthamornphong, "Paddy leaf disease classification using EfficientNet B4 With compound scaling and swish activation: A deep learning approach," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3451557.
- [27] A. Mohan, *Enhanced Multiple Dense Layer EfficientNet*, Purdue University, 2024.
- [28] L. Arora, S. K. Singh, S. Kumar, H. Gupta, W. Alhalabi, V. Arya, S. Bansal, K. T. Chui, and B. B. Gupta, "Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy," *Scientific Reports*, vol. 14, no. 1, p. 30554, 2024, doi: 10.1038/s41598-024-81132-4.
- [29] P. Kaushik and P. Sharma, "Optimizing Automated Eye Disease Detection: A Deep Learning Approach Using EfficientNetB3 for Accurate Multi-Class Classification," in *2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT)*, pp. 1–6, 2024, doi: 10.1109/INSPECT63485.2024.10896145.
- [30] A. Alshoraihy, "EfficientNetB3 in Leukemia Detection: Advancements in Medical Imaging Analysis," *Medinformatics*, 2025, doi: 10.47852/bonviewMEDIN52023293.
- [31] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabdullah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023, doi: 10.3390/su15075930.
- [32] W. Ladines-Castro, G. Barragán-Ibañez, M. A. Luna-Pérez, A. Santoyo-Sánchez, J. Collazo-Jaloma, E. Mendoza-García, and C. O. Ramos-Peñafiel, "Morphology of leukaemias," *Revista Médica del Hospital General de México*, vol. 79, no. 2, pp. 107–113, 2016, doi: 10.1016/j.hgmx.2015.06.007.
- [33] M. M. Amin, S. Kermani, A. Talebi, and M. G. Oghli, "Recognition of acute lymphoblastic leukemia cells in microscopic images using K-means clustering and support vector machine classifier," *Journal of Medical Signals and Sensors*, vol. 5, no. 1, pp. 49–58, Jan.–Mar. 2015.
- [34] A. Mvd, "Leukemia classification," *Kaggle*, 2020. [Online]. Available: <https://www.kaggle.com/datasets/andrewmvd/leukemia-classification>
- [35] G. Verma, "Eye Disease Classification Enhanced by EfficientNetB3: A Deep Learning Approach," in *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, pp. 1016–1021, 2024.
- [36] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019, doi: 10.1186/s40537-019-0197-0.
- [37] G. Kaur and N. Sharma, "EfficientNetB3-based lung cancer classification from CT scan images: A deep learning approach," in *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, pp. 206–212, 2024, doi: 10.1109/ICTACS62700.2024.10841049.
- [38] R. Ochoa-Ornelas, A. Gudiño-Ochoa, and J. A. García-Rodríguez, "A hybrid deep learning and machine learning approach with mobile-EfficientNet and grey wolf optimizer for lung and

- colon cancer histopathology classification," *Cancers*, vol. 16, no. 22, p. 3791, 2024, doi: 10.3390/cancers16223791.
- [39] P. Nasra, S. Gupta, and G. R. Kumar, "Leveraging EfficientNetB3 for accurate breast cancer classification: Insights and innovations," in *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1473–1478, 2024, doi: 10.1109/ICCES63552.2024.10859795.
- [40] R. Chauhan, H. P. Semwal, J. B. Fernandes, V. N. Alone, and R. Maranan, "Website design and development using advance technology React-JS," in *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)*, pp. 678–684, 2023, doi: 10.1109/AECE59614.2023.10428222.
- [41] M. Lathkar, *Building Web Apps with Python and Flask: Learn to Develop and Deploy Responsive RESTful Web Applications Using Flask Framework (English Edition)*. BPB Publications, 2021.
- [42] Z. Zheng, X. Cai, H. Wang, and X. Fu, "Lightweight intelligent fire monitoring system based on Flask and YOLOv8," in *2024 2nd International Conference on Artificial Intelligence and Automation Control (AIAC)*, pp. 365–369, 2024, doi: 10.1109/AIAC63745.2024.10899473.
- [43] M. K. Al Yassar, M. Fitria, M. Oktiana, M. A. Yufnanda, K. Saddami, K. Muchtar, and T. R. A. Isma, "The role of U-Net segmentation for enhancing deep learning-based dental caries classification," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, pp. 253–269, 2025, doi: 10.35882/ijeemi.v7i2.75.
- [44] A. Bahri, M. Oktiana, M. Fitria, and Z. Zulfikar, "Impact of image quality enhancement using homomorphic filtering on the performance of deep learning-based facial emotion recognition systems," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 11, no. 2, pp. 206–224, Apr. 2025, doi: 10.26555/jiteki.v11i2.30409.
- [45] S. Chand and V. P. Vishwakarma, "Acute Leukaemia Diagnosis Using Transfer Learning on Resnet-50," in *Proceeding of the 4th International Conference on Information Management & Machine Intelligence*, pp. 1–7, 2022.
- [46] L. F. Rodrigues, A. R. Backes, B. A. N. Travençolo, and G. M. B. de Oliveira, "Optimizing a deep residual neural network with genetic algorithm for acute lymphoblastic leukemia classification," *Journal of Digital Imaging*, vol. 35, no. 3, pp. 623–637, 2022.
- [47] M. El-Seddek, "A Deep Learning Technique for Multi-Classification of Acute Lymphocytic Leukemia," in *2024 International Telecommunications Conference (ITC-Egypt)*, pp. 404–409, 2024.

Author Biography



Sayed Muchallil graduated with a Bachelor of Engineering (B.Eng) degree from the Electrical Engineering Department at Diponegoro University, Semarang, Central Java, in 2004. He has served as a lecturer at Syiah Kuala University since 2005. In 2009, he was awarded the Fulbright Scholarship to continue his studies. He obtained his Master of Science in Computer Science degree from the Computer Science and Engineering (CSE) department at the University of Texas at Arlington (UTA). After completing his master's degree, in 2011, he was elected as the final project and fieldwork coordinator in the Electrical Engineering Department. In 2012, he was appointed as the head of the computer section in Electrical Engineering and has been the head of the Computer Network Laboratory since 2014. In 2018, he began his PhD studies at the Institutt for Informatikk (IFI) at the Universitetet i Oslo. His research interests include mobile technologies, web services, information security and privacy, databases, and data and computer security.



Maya Fitria began her career as a lecturer and member of the Department of Electrical Engineering and Computer Science at Syiah Kuala University in 2017 and has been there ever since. She earned her Bachelor's degree in Computer Science from the University of Indonesia (UI) in 2012. In 2013, she continued her studies at the Department of Computer Engineering, University of Duisburg-Essen, Germany, specializing in Interactive Systems and Visualization. During her studies, she received support from the DAAD-LPSDM Aceh Scholarship. She completed her master's degree in 2016, earning a Master of Science in Computer Engineering. Her research interests is in Human-Centered AI and IoT for Diagnostic and Interactive Systems.



Ridha Arrahman held a Bachelor's degree in Computer Engineering at the Faculty of Engineering, Syiah Kuala University. His research focuses on computer vision and artificial intelligence, including machine learning and deep learning. During his studies, he has developed expertise as a programmer in software development, mobile applications, and website engineering, and has served as an assistant laboratory in the programming. Additionally, he participated in the MBKM Bangkit program, undertaking independent study with a focus on cloud computing to build a technical foundation that supports further exploration in artificial intelligence, particularly in the integration and operationalization of AI models.



Khairun Saddami obtained Bachelor's degree in 2015 from Syiah Kuala University, Indonesia. He received his PhD degree in Electrical and Computer Engineering from Syiah Kuala University, Indonesia, in 2020, where he awarded a scholarship from the Ministry of Research, Technology, and Higher Education, the Republic of Indonesia, under the Scheme of Pendidikan Magister Menuju Doktor Untuk Sarjana Unggul (PMDSU). He has

been with Multimedia and Signal Processing Research Group (MUSIG), Syiah Kuala University since 2020. He also acts as reviewer in reputable journal such as IEEE Access and ACM Computer Survey. He is a member of IEEE and International Association for Pattern Recognition (IAPR). His research interests are in document image analysis, deep learning, biometric and biomedical image processing and pattern recognition.