Array 26 (2025) 100397

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

# Deep learning-based detection and classification of acute lymphoblastic leukemia with explainable AI techniques☆

Debendra Muduli [a], Sourav Parija [a], Suhani Kumari [a], Asmaul Hassan [b], Harendra S. Jangwan [c], Abu Taha Zamani [d,*], Sk. Mohammed Gouse [e], Banshidhar Majhi [a], Nikhat Parveen [f]

[a] Department of Computer Science and Engineering, C.V. Raman Global University, Bidya Nagar, Odisha, 752054, India
[b] Department of Information Technology Western Governors University 4001 S 700 E 300, Millcreek, UT 84107, USA
[c] Computer Science & Engineering Quantum University, Uttarakhand, India
[d] Department of Computer Science, Faculty of Science, Northern Border University, Arar, 73213, Kingdom of Saudi Arabia
[e] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India
[f] Department of Artificial Intelligence, College of Computing and Information Technology, University of Bisha, Saudi Arabia

ARTICLE INFO

*Keywords:*
Deep learning
Image data augmentation
ResNet34
ResNet50
ResNetV2
VGG16
Xception
EfficientNetB0
EfficientNetV2
DenseNet121
ResNet152

ABSTRACT

Leukemia is identified by an excess of immature white blood cells (WBC) being formed in the bone marrow, leading to cancer. It is divided into two main types: acute, which stems from early cell growth ab- normalities and involves rapid immature cell proliferation, and chronic, which progresses more slowly due to a blockage in the later stages of the cell life cycle. Detecting acute lymphoblastic leukemia (ALL) at an early stage is critical to reducing its associated mortality rate. This study presents an empirical analysis of various pre-trained deep learning models, including VGG16, VGG19, ResNet50, Xception, ResNet152, EfficientNet- B0, NASNetMobile, DenseNet169, DenseNet121, and EfficientNetV2B0, for the detection and classification of ALL. A comprehensive evaluation highlights the effectiveness of deep learning in distinguishing different types of ALL, demonstrating its potential as a reliable diagnostic tool in medical imaging. Additionally, we evaluated the performance of these models using different optimization techniques, including Adadelta, SGD, RMSprop, and Adam, to determine the most effective optimization strategy for improving classifica-tion accuracy. Our results demonstrate that EfficientNet-B0 achieved a classification accuracy of 72 %, while NASNetMobile attained 81 %. Notably, DenseNet121 outperformed these models with an accuracy of 99 %. Furthermore, the remaining seven models VGG16, VGG19, ResNet50, Xception, ResNet152, DenseNet169, and EfficientNetV2B achieved a perfect classification accuracy of 100 %, highlighting their robustness and effectiveness in our experimental setup. To improve the interpretability of the leukemia detection process, explainable AI techniques, including Grad-CAM, Score-CAM, and Grad-CAM++, were integrated to vi-sualize critical regions influencing model predictions. These techniques enhance transparency by providing visual explanations of classification decisions. A detailed comparative analysis was conducted, examining key parameters such as learning rate, optimization algorithms, and the number of training epochs to determine the most effective approach. The study leveraged a publicly available acute lymphoblastic leukemia dataset to ensure comprehensive model evaluation. By offering insights into model performance and interpretability.

## 1. Introduction

Acute lymphoblastic leukemia (ALL) is a disease that affects blood cells and has the potential to rapidly spread throughout the body, with fatal consequences if not addressed immediately. Detecting the disease early is crucial to increase treatment success rates, with a major diagnostic process the analysis of white blood cells in peripheral blood samples. These abnormal cells also reduce the immune system's strength, limiting the bone marrow's ability to produce red blood cells and platelets. Furthermore, cancerous white blood cells have the

☆ this research serves as a valuable reference for future studies, aiding in the development of more accurate and transparent deep-learning models for leukemia detection.

* Corresponding author.
*E-mail address:* abutaha.zamani@nbu.edu.sa (A.T. Zamani).

https://doi.org/10.1016/j.array.2025.100397
Received 16 January 2025; Received in revised form 18 March 2025; Accepted 3 April 2025
Available online 7 April 2025

potential to travel through the bloodstream and harm essential organs such as the liver, kidneys, spleen, and ALL, resulting in more severe health problems. The condition is defined by the overproduction of early-stage white blood cells in the bone marrow. Every year in the USA, over 6,500 instances of ALL are identified in adults and kids combined, constituting about one-fourth of childhood cancers, with a growing trend. The early detection of the disease is being improved by the latest progress in artificial intelligence and big data analytics, assisting healthcare professionals in making clinical decisions. Stem cells give rise to blood cells, and they have the potential to differentiate into myeloid or lymphoid stem cells.The growth of leukemia is linked to the production of young white blood cells by these stem cells. In 2022, approximately 6,660 new cases of acute lymphoblastic leukemia (ALL) were reported in the U.S., leading to 1,660 deaths among children and adults Saeed et al. [1]. Early detection is crucial for improving treatment options and survival rates. Research by Shafique et al. [2] involved customizing a pretrained AlexNet model, achieving high sensitivity and accuracy. A study by Ghorpade et al. [3] evaluated automated CNN models, including Xception and MobileNetV2, achieving 100 % accuracy in ALL classification. Additionally, Genovese et al.

[4] improved blood sample image analysis, while Jawahar et al. [5] introduced the ALNet model, achieving 91.13 % accuracy for ALL detection. Finally, Saeed et al. [1] proposed Multi-Attention EfficientNet models with accuracies of 99.73 % and 99.25 %.

With the continuous advancement of technology, machine learning is being used increasingly in different industries, such as healthcare. This technology can be used to detect a variety of health concerns and problems, as well as has been used in various other areas [6–10]. This study presents a comprehensive approach to preprocessing peripheral blood smear image datasets for classifying ALL subtypes using deep learning models. We analyze ten different pre-trained CNN architectures, leveraging advancements in AI for disease detection. Images are resized to 224 × 224 pixels and organized by class labels, while data augmen-tation techniques enhance the training dataset, enhancing model generalization, and minimizing overfitting. After dividing the data into train and test sets, we evaluated the performance of each model, achieving impressive precision, with several models reaching 100 %, underscoring the capability of neural networks

in medical diagnostics. To enhance healthcare automation, we integrated explainable artificial intelligence (XAI) techniques into our deep learning framework. Illustrated in Fig. 3. The XAI heatmap emphasizes critical regions with warmer colors, reflecting the model's confidence and input complexity. This system aids healthcare professionals in detecting Acute Lymphoblastic Leukemia (ALL) disease and understanding machine-generated decisions. By revealing the reasoning behind classifications, it reduces decision errors, fosters trust in Artificial Intelligence diagnostics, and ultimately facilitates earlier medical interventions, contributing to a more efficient diagnostic process.

Our proposed model has been evaluated against other top-performing models in the field of acute lym-phoblastic leukemia (ALL) disease. Our recommended model shows sensitivity to different classes and is straightforward to incorporate into a complete architecture. No extensive feature engineering is needed, and it can help hematologists and oncologists make precise and consistent diagnoses of ALL diseases.

The proposed study's key contributions can be outlined as follows.

- The efficacy of 10 highly efficient previously trained deep CNN models—VGG16, VGG19, ResNet50, Xception, ResNet152, EfficientNet-B0, NASNetMobile, DenseNet169, DenseNet121, and Efficient- NetV2B0—has been thoroughly assessed. This study examines the impact of various hyperparameters. In the end, the top-performing model is discovered, giving researchers a strong basis for creating a more effective CNN-based solution for detecting acute lymphoblastic leukemia (ALL) disease early
- We incorporated explainable AI to enhance transparency and trust in our model's predictions for diagnosing acute lymphoblastic leukemia (ALL) disease. Techniques like Grad-CAM, Grad-CAM++ & Score-CAM provide visual insights into critical regions influencing decisions, helping healthcare professionals understand the rationale behind predictions and fostering confidence in accurate diagnoses and treatment planning.
- The available datasets for public use are limited and show disparities. In order to tackle this problem, we introduced multi-operation data augmentation to maintain balanced samples among all four categories of acute lymphoblastic leukemia (ALL) illness.

To address the above contributions, this study will address the following research questions.

RQ1: How can transfer learning with pre-trained CNN models improve the accuracy and efficiency of Acute Lymphoblastic Leukemia (ALL) detection compared to training models from scratch?

RQ2: What are the optimal criteria for selecting and evaluating CNN architectures for classifying ALL subtypes in blood smear images?

RQ3: How does Explainable AI (XAI) improve the interpretability and trustworthiness of CNN-based ALL diagnostic models?

RQ4: What are the key challenges in using CNNs for ALL classification, and how can strategies like data augmentation and hyperparameter tuning address them?

The paper is organized as follows: Section 2 provides an analysis of related works, while Section 3 outlines the research methodology, including a description of the data, the preprocessing steps, and detailed information about the models utilized in this study. Section 4 outlines the results, while Section 6 concludes the paper.

## 2. Related works

Recent studies on leukemia detection have shown notable progress in the application of machine learning (ML) and deep learning (DL) techniques for classifying acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML).

Several studies have utilized traditional machine learning models to detect leukemia. Madhukar et al.

[11] improved image contrast and detected important features in AML detection, achieving 93.5 % accuracy using SVM as a classifier. Setiawan et al. [12] suggested a categorization system for AML subtypes M4, M5, and M7, reaching 92.9 % precision through a color k-means algorithm and multi-class SVM. Faivdullah and colleagues presented a system consisting of three layers that utilized a dense scale-invariant feature transform for extracting features, resulting in a classification accuracy of 79.37 % on a dataset of 400 samples. Hosseni et al. [13] present a mobile application that uses a lightweight convolutional neural network (CNN) model to accurately classify acute lymphoblastic leukemia (B-ALL) of B cells from noncancerous cells. Similarly, Ghaderzadeh et al. [14] proposed a fast and efficient CNN model for the accurate detection of B-ALL diagnosis and the classification of its subtypes using peripheral blood smear images. Laosai et al. [15] created an AML categorization method using k-means and contour signature approaches, the SVM classifier achieved a maximum accuracy of 92 % on 100 pictures. Patel et al. [16] created an automated detection method for leukemia based on microscopic images, incorporating noise and blur removal, followed by segmentation of white blood cells. For the detection of leukemia, Abbasi et al. [17] explored multi-omics data by employing machine learning (ML) and deep learning (DL) algorithms. Various ML methods were compared with DL approaches. In a separate study, Rangini M et al. [18] proposed an innovative method that utilizes First Order Histogram (FOH) and the Gray-Level Co-occurrence Matrix (GLCM) for extracting texture features. The results of their simulations show that combining GLCM feature extraction with the ANN model achieves the highest accuracy.

On the other hand, deep learning has proven to be more effective in improving accuracy and precision in leukemia detection. Shafique et al. [2] customized a pre-trained AlexNet model with data augmentation,

achieving high sensitivity, specificity, and accuracy for ALL disease detection. Ghorpade et al. [3] reported perfect accuracy (100 %) with automated CNN models, including CNN, Xception, and MobileNetV2, for ALL disease classification. Genovese et al. [4] enhanced blood sample images through image analysis and deep learning, improving classification precision for ALL. ALNet, a deep neural network utilizing depth-wise convolution introduced by Jawahar et al. [5], achieved 91.13 % accuracy for ALL diseases. Saeed et al. [1] proposed Multi-Attention EfficientNetV2S and EfficientNetB3 models, which attained accuracies of.

99.73 % and 99.25 %, respectively, in differentiating normal and blast cells in blood smear images. Liu et al.[19] developed a weakly supervised data augmentation classification network (WT-DFN) for the diagnosis of ALL diseases, demonstrating superior performance in various datasets. Additionally, Ansari et al. [20] created a modified DL model for acute lymphoblastic leukemia disease detection, achieving 99 % accuracy using Tversky loss and GANs for dataset enhancement. Kumar et al. [21] employed eight CNN models, including ResNet152 v2 and DenseNet201, achieving over 95 % accuracy in diagnosing COVID-19 pneumonia from X-rays and CT scans. In another domain, El-Rashidy et al. [22] developed a monitoring framework for gestational diabetes mellitus (GDM), utilizing a deep neural network to achieve 95.7 % accuracy from a dataset of 16,354 women. Ramaneswaran et al. [23] introduced a hybrid model combining Inception v3 and XGBoost for ALL disease classification, achieving a weighted F1-score of 0.986. Pal-czy'nski et al.

[24] presented a hybrid AI system for classifying white blood cells, achieving over 97.4 % accuracy using MobileNet v2 and various machine learning algorithms. Talaat et al. [25] conducted a systematic review on early leukemia detection, developing an optimized CNN model that achieved 99.99 % accuracy with the C-NMC Leukemia dataset through hyperparameter optimization. Rahmani et al. [26] used various transfer learning techniques as feature extractors, along with different feature selectors, and among all the classifiers, MLP performed the best in detecting ALL and HEM diseases. Alim et al. [27] classified different B-ALL subtypes from peripheral blood smear images using a novel ResNet-50 model, which included additional customized fully connected layers, such as dense and dropout layers. Genovese et al. [28] introduced an innovative decision support system for detecting ALL, combining deep learning (DL) and explainable artificial intelligence (XAI). Pawar et al. [29] introduced a deep, optimized convolutional neural network (CNN) consisting of five convolutional blocks, which included 11 convolutional layers and 4 max-pooling layers. The optimization of this deep CNN model involved tuning hyperparameters such as epochs, batch size, and the selection of specific optimizers, namely Adam and Adamax, with the Adam optimizer yielding the best performance. Lalithkumar K et al. [30] presented a novel model named CapsENet, which integrates Capsule Networks (CapsNet) with EfficientNet (ENet) for classifying ALL. Kumar et al. [31] introduced advanced CNN models, as well as two hybrid models, InceptionResNetV2 and XceptionInceptionResNetV2, for the detection and classification of ALL disease using microscopic images. Tusar et al. [32] utilized Deep Neural Networks for detecting ALL blast cells in microscopic blood smear images, achieving the highest accuracy with the MobileNetV2 model. Awad et al. [33] applied advanced DL models such as YOLOv8 and YOLOv11 for the detection of ALL disease. Senthil K et al. [34] proposed a novel CNN model for acute lymphoblastic leukemia (ALL) disease classification. For leukemia detection, Abbasi et al. [17] analyzed multi-omics data using both machine learning (ML) and deep learning (DL) techniques. Rangini M et al.

[18] introduced a novel approach to extract texture features using First Order Histogram (FOH) and the Gray-Level Co-occurrence Matrix (GLCM), employing both ML and DL methods for classification. Dangore et al. [35] proposed a new deep-optimized CNN model architecture consisting of five convolution blocks with 13 convolution layers and 5 max-pooling layers for multi-class classification of Acute Lymphoblastic Leukemia

using Blood Smear Images. Park et al. [36] employed an ensemble method of deep learning models, including EfficientNet-V1 and EfficientNet-V2, for classifying 12 different cell types. Cheng et al. [37] introduced a model for the detection and classification of acute myeloid leukemia (AML) and B-lymphoblastic leukemia (B-ALL) using flow cytometry, ResNet-50, and a novel EverFlow model. Abougarair et al. [38] used the MobileNetV2 model to extract deep features from a dataset of peripheral blood specimen images. Alsaykhan and Maashi [39] developed a hybrid detection model by combining support vector machine (SVM) and particle swarm optimization (PSO) approaches to automatically identify Acute Lymphoblastic Leukemia (ALL). The study by Jawahar et al. [40] presents a Deep Dilated Residual Convolutional Neural Network (DDRNet) for classifying blood cell images, incorporating Deep Residual Dilated Blocks (DRDB), Global and Local Feature Enhancement Blocks (GLFEB), Channel and Spatial Attention Blocks (CSAB), and sigmoid activation for enhanced feature extraction and classification. Huang & Huang [41] utilized an ensemble-ALL model with a diverse set of convolutional neural networks, including InceptionV3, EfficientNetB4, ResNet50, CONV POOL-CNN, ALL-CNN, Network in Network, and AlexNet.

To detect acute lymphoblastic leukemia (ALL), we used various deep-learning models which were pre-viously trained, including VGG16, VGG19, ResNet50, Xception, ResNet152, DenseNet121, DenseNet169, EfficientNet-B0, NASNetMobile, and EfficientNetV2B0. Our goal was to reduce the mortality rate associ-ated with ALL diseases through a comprehensive model. We conducted an empirical survey highlighting the role of deep learning in detecting and classifying ALL disease types. By incorporating explainable AI techniques like Grad-CAM, Grad-CAM++, and Score-CAM, we enhanced model interpretability. Our findings showed 100 % accuracy for several models, providing valuable insights for future researchers developing effective detection methods for all diseases.

## 3. Materials and methods

This section provides a comprehensive overview of the proposed methodology for detecting acute lym-phoblastic leukemia (ALL) disease, along with a description of the data utilized to validate the model.

### 3.1. Dataset description

The publicly available leukemia dataset [13,14,42] consists of a total of 3,242 peripheral blood smear (PBS) images, classified into two primary classes: benign and malignant. The benign class includes 512 files, representing hematogenous samples that do not exhibit signs of leukemia. The malignant class contains images from three subtypes of acute lymphoblastic leukemia (ALL): 955 files are classified as malignant Pre-B cells, 796 files as malignant Pro-B cells, and 979 files as malignant Early Pre-B cells. The sample dataset has been depicted in Fig. 1 and the dataset split size is given in Table 1.

### 3.2. Proposed methodology

The illustration of the automated diagnostic model for acute lymphoblastic leukemia (ALL) using pre-trained CNN models can be seen in Fig. 2, and its architectural details are provided in Table 2. This design is meant to categorize PBS images into different ALL groups through two key steps: preprocessing (which involves normalizing and augmenting the data) and using pre-trained CNN structures for classification.

Additionally, we have introduced an explainable AI component to enhance detection accuracy and ensure a more valid and transparent decision-making process for the model, as shown in 3. Each stage is explained in detail in the following sections.

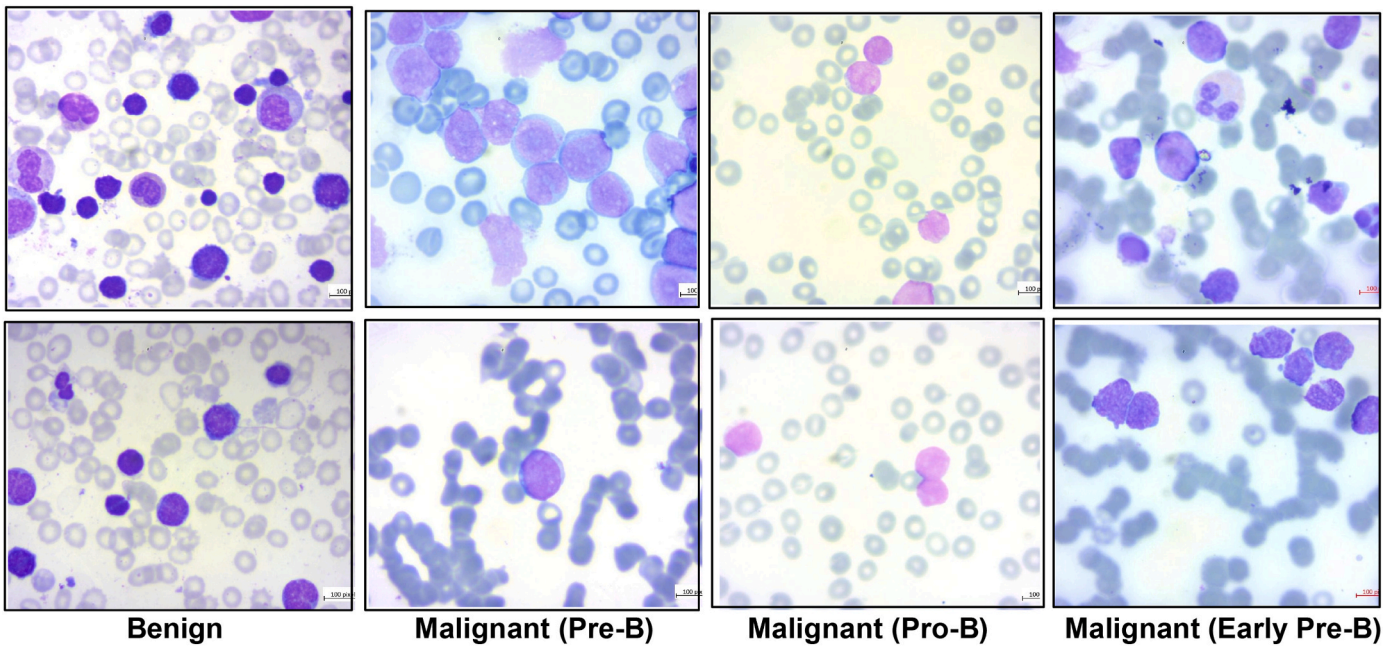



**Fig. 1.** Sample image from the acute lymphoblastic leukemia dataset.

**Table 1**
Details of splitting of the data set.

| Class | Given Dataset | | Augmented Dataset | | |
|---|---|---|---|---|---|
| | Train | Test | Train | Validation | Test |
| Benign | 409 | 103 | 736 | 82 | 103 |
| [Malignant] Pre-B | 764 | 191 | 1375 | 153 | 191 |
| [Malignant] Pro-B | 637 | 159 | 1147 | 127 | 159 |
| [Malignant] early Pre-B | 783 | 196 | 1409 | 157 | 196 |
| **Total** | **2593** | **649** | **4667** | **519** | **649** |

*3.2.1. Data preparation and analysis*

This part presents an extensive breakdown of the methods employed during the processing stages.

Data Augmentation Convolutional Neural Networks (CNNs) typically require extensive datasets to ef-fectively perform and generalize well. However, the availability of Peripheral Blood Smear (PBS) images was limited. To address this issue, data augmentation techniques were implemented to artificially expand the dataset and enhance model training. Data augmentation significantly improves the model's ability to generalize, addressing data scarcity issues and reducing overfitting, which is a critical concern for deep learning-based medical image analysis.

To improve the diversity and robustness of the training dataset, multiple data augmentation techniques were strategically implemented. These transformations help simulate real-world variations, ensuring the model's adaptability to different imaging conditions. Rotation has been applied by randomly adjusting images up to 40°, enhancing the model's ability to handle different orientations encountered in practice. Horizontal flipping has been incorporated to allow recognition of anatomical structures regardless of their

left-right alignment. To introduce realistic deformations, shearing transformations up to 40 % were utilized, replicating distortions that may occur during image acquisition. Furthermore, zooming within a 40 % range helped the model adapt to variations in object size and focus. To address positional inconsistencies, width shifting and height shifting were performed, allowing horizontal and vertical displacements of up to 40 % of the image dimensions. These transformations collectively applied changes within the range of −40 % to.

+40 %, further enhancing the diversity of the dataset. Collectively, these enhancements enriched the data set, enhancing the generalization capabilities of the model while reducing the susceptibility to overfitting. After applying augmentation techniques, the dataset has been expanded to a total of 5,186 samples, significantly enhancing its diversity. This increased sample size helps improve the generalization capability of deep learning models by reducing overfitting and ensuring robustness across varied data distributions. A selection of augmented image samples is presented in Fig. 4, illustrating the diverse transformations applied to enhance dataset variability.

Normalization Data normalization is a crucial process for maintaining numerical stability in CNN archi-tectures. Therefore, we have implemented it in our process. This begins by loading images from a specified directory, ensuring that all images are resized to a uniform dimension of 224 × 224 pixels. The program systematically navigates through the directory structure to collect images along with their corresponding class labels. Once loaded, these images are converted into numpy arrays for further manipulation. We

subsequently split the dataset into two parts: 80 % for training and 20 % for testing, which maintains class distribution through stratification. Subsequently, the pixel values are normalized to a range of 0–1. At last, the labels are converted into a categorical format using one-hot encoding to support the multi-class classification task.

*3.2.2. Acute lymphoblastic leukemia prediction using pretend CNN models*

CNN models have shown excellent performance in various medical image-processing tasks. Neverthe-less, it can be difficult to train these models for predicting Acute Lymphoblastic Leukemia cases because of the scarcity of PBS images. Using previously trained models using transfer learning (TL) can offer sig-nificant benefits in these circumstances. TL utilizes insights from a deep learning (DL) model trained on a vast dataset to tackle a similar task with a small dataset. This method eliminates the need for a bigger dataset and lessens the training time usually needed when starting from the beginning. In this research, we trained ten distinct pre-trained models, which include VGG16 [43], VGG19 [43], ResNet50 [44], Xception [45], ResNet152 [44], EfficientNetB0 [46], NASNetMobile [47], DenseNet169 [48], DenseNet121 [48], and Ef-ficientNetV2B0 [49]. Chosen for disease detection and classification were these networks renowned for their remarkable achievements in computer vision and medical image analysis. These

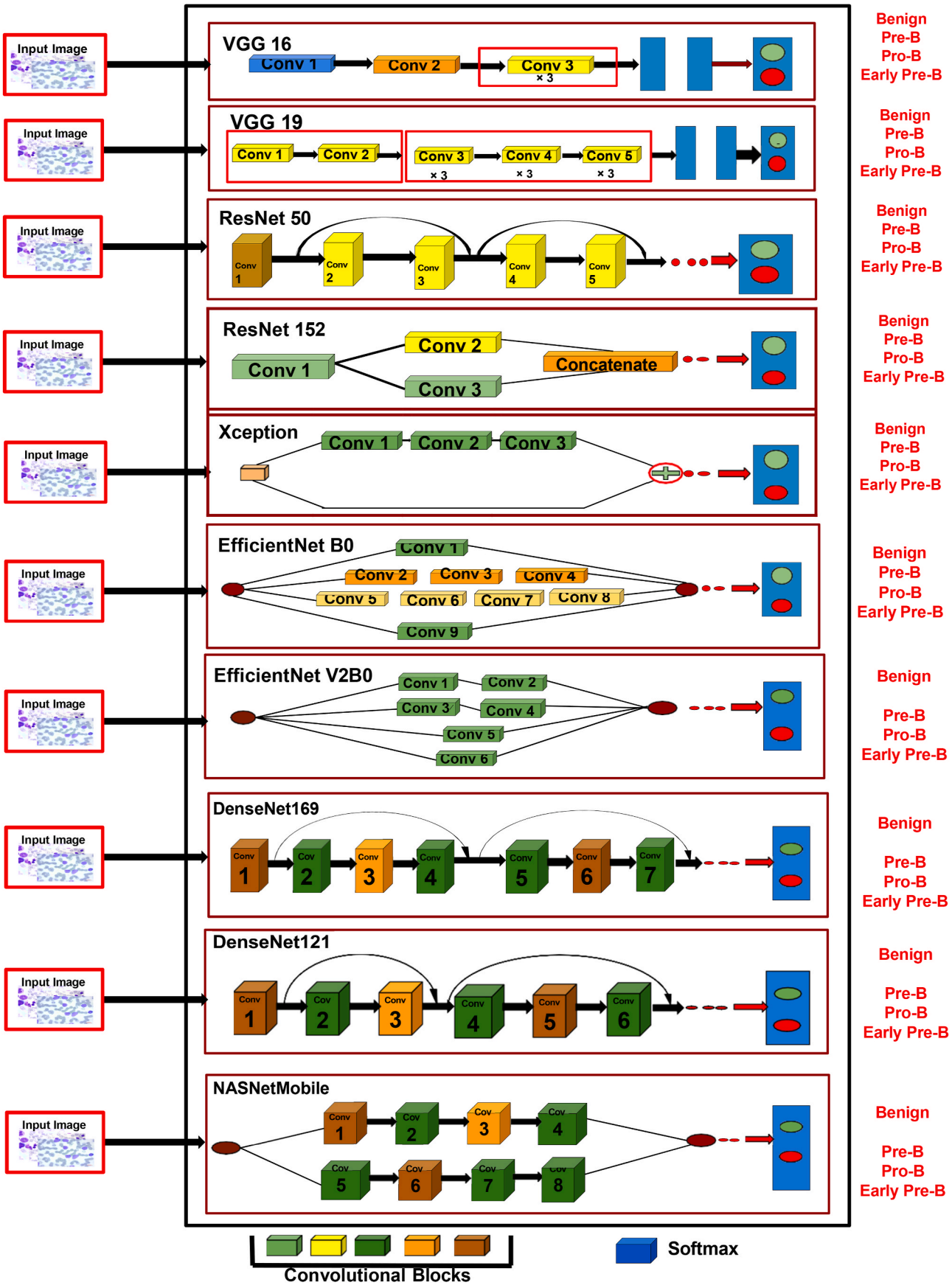



**Fig. 2.** Proposed model.

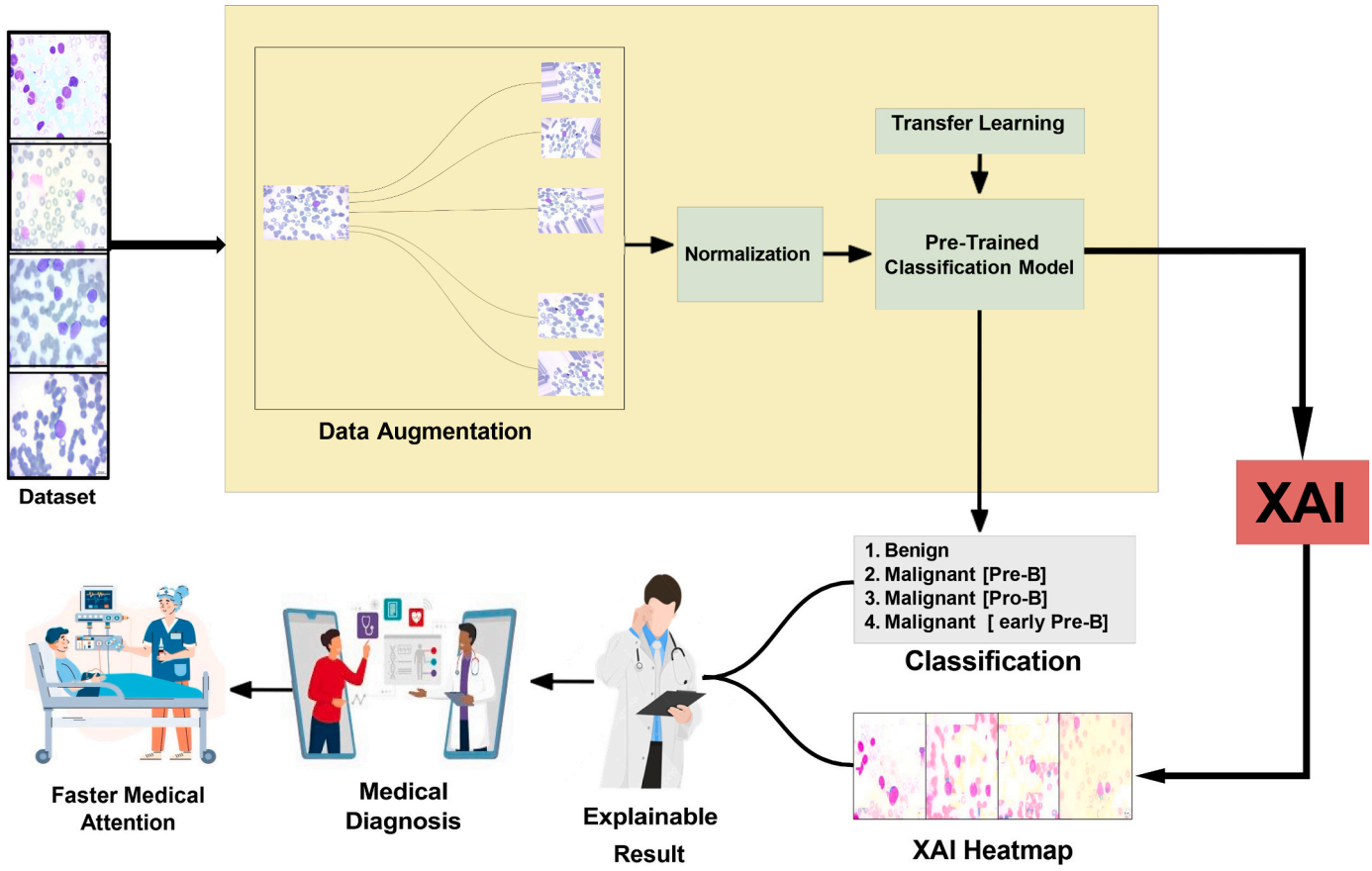



**Fig. 3.** Proposed XAI model.

**Table 2**
Architectural summarization of different deep learning models.

| Deep Learning Model | No of Lay-ers | Hyperparameters (in million) | Input Size | Size of classifica-tion layer |
|---|---|---|---|---|
| VGG-16 | 16 | 138 | (224, 224, 3) | (None, 4) |
| VGG-19 | 19 | 143.7 | (224, 224, 3) | (None, 4) |
| ResNet-50 | 50 | 25 | (224, 224, 3) | (None, 4) |
| ResNet-152 | 152 | 60 | (224, 224, 3) | (None, 1) |
| DenseNet-169 | 169 | 14 | (224, 224, 3) | (None, 4) |
| DenseNet-121 | 121 | 8 | (224, 224, 3) | (None, 4) |
| EfficientNetB0 | 50 | 5.3 | (224, 224, 3) | (None, 4) |
| EfficientNetV2B0 | 53 | 7.1 | (224, 224, 3) | (None, 4) |
| Xception | 71 | 22.9 | (299, 299, 3) | (None, 4) |
| NASNet-Mobile | 88 | 23 | (224, 224, 3) | (None, 4) |

models were first trained on the 'ImageNet' dataset before being fine-tuned on the PBS images. The architecture of each model includes custom layers added to pre-trained frameworks, ensuring optimal configurations for classification tasks. Hyperparameters are crucial for adjusting these deep learning models, and keeping them constant is essential for ensuring an equitable comparison. More information about the parameter configurations can be located in section 4.1. All models are trained with the Adam optimizer and categorical cross entropy as the loss function, and they undergo training for up to 30 epochs, incorporating early stopping and learning rate reduction callbacks. The performance of the proposed model is validated on validation and tasting dataset, with critical metrics such as accuracy, precision, ROC curves, recall, and F1-score computed for each model. The objective of this study is to identify the most effective architecture for classifying ALL disease subtypes, with several models achieving remarkable accuracy, including instances of 100 %.

Table 2 presents a summary of the overview of the architecture of the pre-trained CNN models, while Fig. 2 illustrates the main components of each network. A deep convolutional network renowned for its use, VGG-16 is known for its simplicity and depth, consisting of 16 layers that contain a large number of parameters, allowing it to effectively capture intricate features in images. VGG-19 builds on this by adding three additional layers, thereby enhancing feature extraction capabilities while maintaining a straightforward design characterized by small receptive fields, which enables it to recognize complex patterns in visual data. ResNet-50 introduces residual learning in order to address the vanishing gradient problem, featuring 50 layers that facilitate high accuracy on challenging tasks while ensuring computational efficiency. ResNet-152 further extends this concept with even deeper layers, significantly improving its ability to model complex functions and excelling in robust feature extraction, thus attaining cutting-edge performance in a variety of image classification tasks. DenseNet-169 is distinguished by its densely connected architecture, where

every of layer receives input from all preceding layers, which enhances gradient flow and feature reuse, resulting in fewer parameters and high accuracy in classification tasks. Similarly, DenseNet-121 maintains this densely connected design with 121 layers but is more compact, demonstrating efficiency in training and outstanding

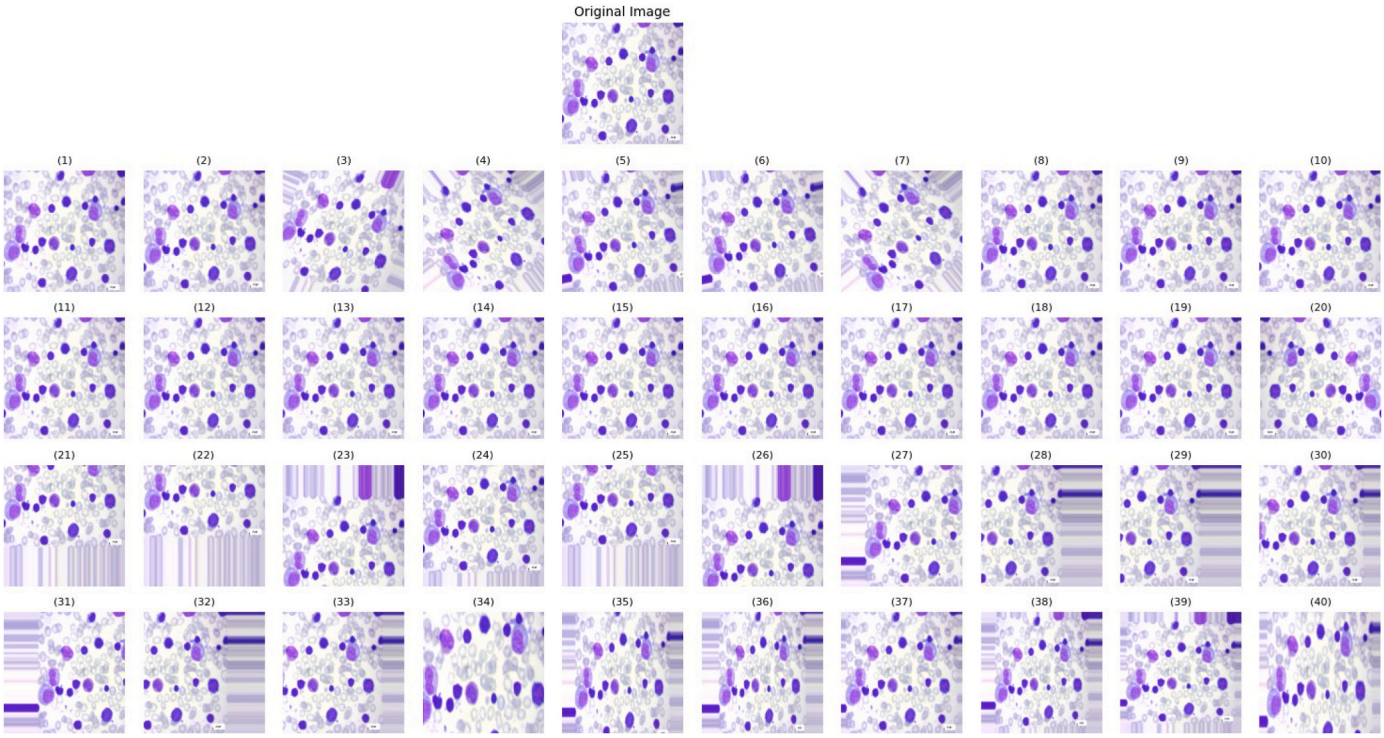



**Fig. 4.** Sample images of Data Augmentation: Images 1–7 undergo rotation augmentation, Images 8–14 experience shearing augmentation, Images 15–20 receive horizontal flip augmentation, Images 21–26 are subjected to width shifting augmentation, Images 27–33 undergo height shifting augmentation, and Images 34–40 are enhanced with zooming augmentation.

performance across various computer vision applications. EfficientNetB0 optimizes model scaling by balancing depth, width, and resolution, resulting in fewer parameters while improving accuracy, making it suitable for resource-constrained environments. EfficientNetV2B0 advances this further with new training techniques and architectural modifications, enhancing efficiency and performance, particularly for real-time image classification tasks, balancing speed and accuracy. Xception employs depthwise separable convolutions to achieve efficient computation while maintaining strong feature representation, excelling at capturing spatial hierarchies in images with its 71 layers, making it robust for complex classification tasks. Finally, NASNet-Mobile is a lightweight architecture tailored for mobile and edge devices, utilizing neural architecture search to optimize performance, with its 88 layers designed for efficiency, making it well-suited for applications that require rapid inference with limited resources. The primary goal of this empirical research is to identify the top-performing deep learning model for ALL disease screening. This will assist researchers in developing more efficient AI-driven solutions and guide future efforts in finding improved models for ALL disease detection.

### 3.2.3. *Acute lymphoblastic leukemia prediction using explainable AI*

In order to enhance the transparency and interpretability of healthcare automation systems, we have inte-grated several Explainable Artificial Intelligence (XAI) methods into our proposed deep learning framework. The architecture of the developed model is illustrated in Fig. 3, while Fig. 5 presents sample outputs gener-ated by each employed XAI technique, including Grad-CAM (Gradient-weighted Class Activation Mapping), Grad-CAM++, and Score-CAM. Grad-CAM generates a coarse localization map by utilizing the gradients of the targeted class flowing into the final convolutional layer. This method highlights image regions crucial to the model's classification decisions, visualizing the areas of significant contribution toward predictions. Specifically, Grad-CAM provides insights by capturing important features that heavily influence the model's outputs. Grad-CAM++ enhances the foundational Grad-CAM method by improving both resolution and interpretability. Grad-CAM++ addresses limitations inherent in Grad-CAM through an advanced weighting scheme, resulting in sharper, more accurate localization maps. This refinement significantly benefits medical applications where precise identification of influential features is critical, thereby offering clearer insights into the model's decision-making processes. Score-CAM, on the other hand, approaches visualization differently by eliminating reliance on gradients, which can often introduce noise or mislead interpretability. Instead, Score-CAM leverages activation maps from the final convolutional layer, computing scores for different im-age regions based on their contribution to the predicted class. This results in robust localization maps that effectively represent class-discriminative regions, mitigating common issues associated with gradient-based visualizations. The XAI-generated heatmaps visually emphasize critical regions by assigning warmer colors to areas that significantly impact classification decisions. The intensity of the highlighted regions

varies according to the model's confidence and the intrinsic complexity of the analyzed images. Comparative visualizations across the four classes within our dataset demonstrate variations in model confidence and inter-pretability among these XAI techniques. This visualization mechanism is particularly valuable for healthcare practitioners, assisting them in both the automated detection of Acute Lymphoblastic Leukemia (ALL) and in comprehending the rationale behind AI-derived diagnostic decisions. By analyzing XAI-generated expla-nations, medical professionals can verify model predictions, substantially decreasing diagnostic inaccuracies. The clear localization of leukemic cells or suspicious regions provided by these visualizations aids practi-tioners in accurately identifying ALL, thereby improving diagnostic reliability. Consequently, this increased interpretability fosters greater confidence and trust in AI-assisted medical diagnostics. Ultimately, our XAI- enhanced system for ALL disease prediction not only streamlines diagnostic processes but also elevates trust in AI-supported medical decisions. The integration of XAI leads to a cost-effective, efficient, and rapid diagnostic approach, facilitating earlier

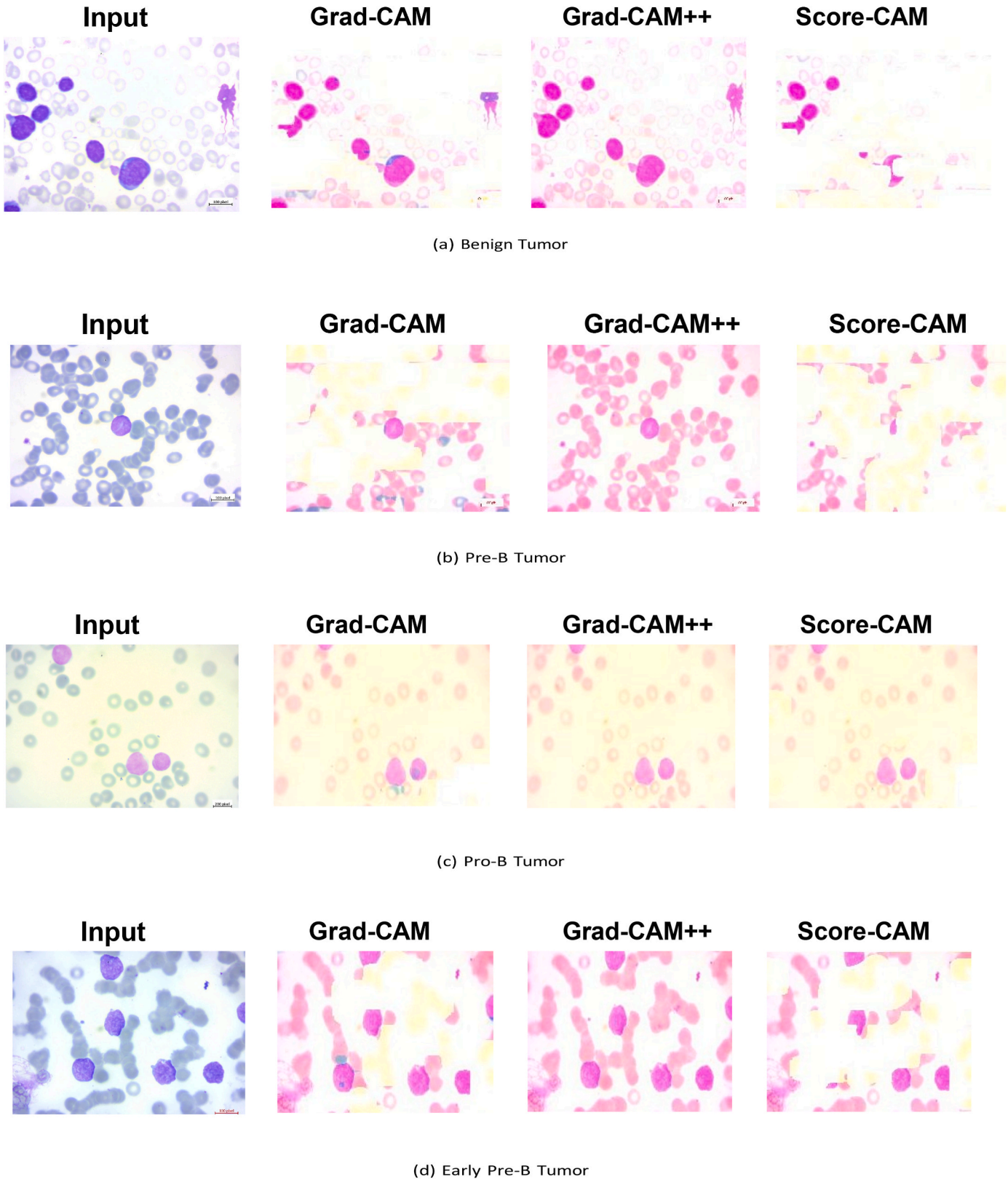



**Fig. 5.** XAI visualization of the model's decision over input images.

interventions and improved patient outcomes, particularly in critical healthcare scenarios.

## 4. Experiment and results

The part shows the outcome of multiple experiments. An empirical study was conducted to predict acute lymphoblastic leukemia (ALL) disease from peripheral blood smear images using 10 previously-trained

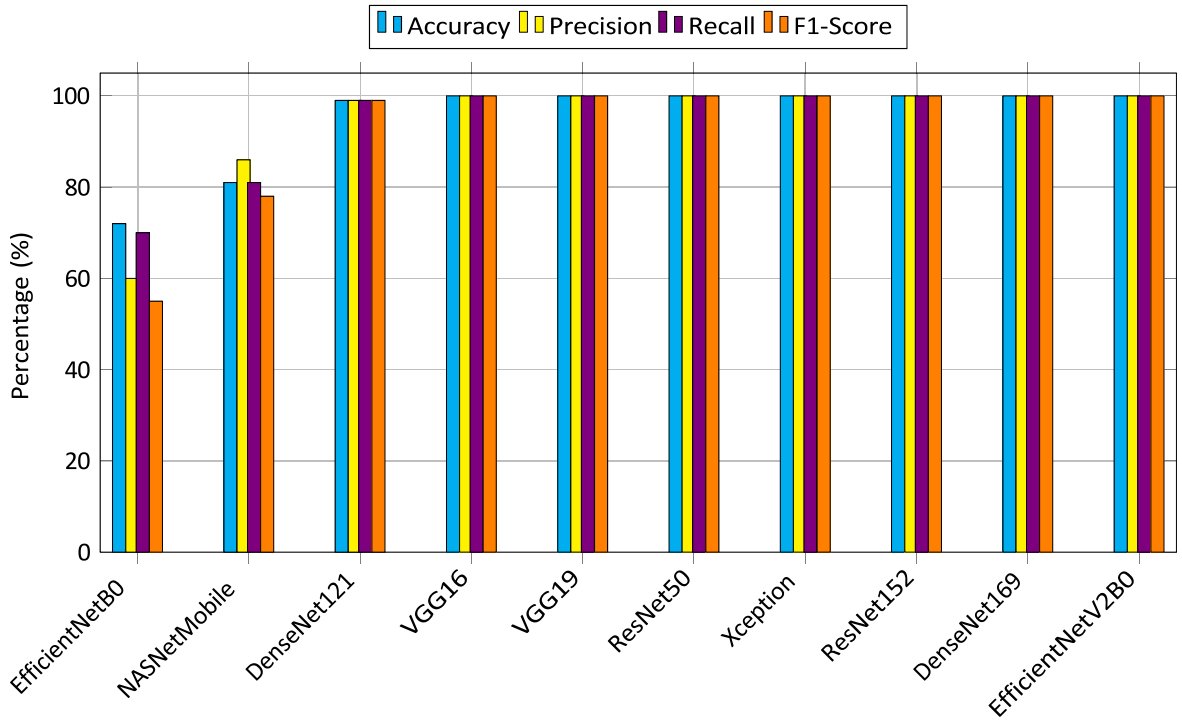



**Fig. 6.** Performance comparison of different pre-trained Classification models during testing.

deep CNN models: VGG16, VGG19, ResNet50, Xception, ResNet152, EfficientNetB0, NASNetMobile, DenseNet169, DenseNet121, and EfficientNetV2B0. We have examined how various hyperparameters af-fect this and compared the 10 CNN models. At last, the model that performs the best is achieved. We also evaluated the results against the latest cutting-edge methods.

### 4.1. Experimental setup and performance metrics

In this study, we assessed the performance of various CNN models on peripheral blood smear (PBS) samples obtained from acute lymphoblastic leukemia (ALL) disease datasets [**?** ]. The splitting strategy for both the augmented and non-augmented samples is shown in Table 1, highlighting the proportion of data used for training and testing. Specifically, the augmented samples (created through data augmentation techniques) were employed to enrich the training set, thereby improving the robustness of the trained models. The images designated for training were further partitioned into 80 % for training and 20 % for validation, ensuring a balanced approach that mitigates overfitting risks while providing sufficient data for unbiased performance evaluation. The image dimensions were set to 224× 224 pixels for the majority of models (*i.e.*, VGG16, VGG19, ResNet50, ResNet152, EfficientNetB0, NASNetMobile, DenseNet169, DenseNet121, and EfficientNetV2B0), whereas for the Xception model the images were resized to 229× 299. In our experiments, we standardized common hyperparameters such as the batch size (32) and the number of training epochs.

(30) based on preliminary empirical studies. The **Adam optimizer** guided the training process, and its initial learning rate was also established through trial and error to balance quick convergence with model stability.

**Performance metrics.** We evaluated the classification performance using multiple metrics: *accuracy*, *precision*, *sensitivity* (or recall), *specificity*, *F1 score*, and *AUC* (area under the ROC curve). These met-rics illuminate various facets of the classification task. For instance, *precision* and *sensitivity* gauge how frequently the classifier is correct when it predicts a positive class, and how often it captures the actual positives in the data, respectively. *Specificity*, conversely, characterizes how effectively the model identifies negative (non-ALL) cases. When all of these indicators are high, it implies a strong correlation between the model's predictive power and accurate classification outcomes.

Each of these metrics is derived from elements of the confusion matrix ($TP_i$, $FP_i$, $TN_i$, $FN_i$), where $TP_i$ represents the count of true positives, $FP_i$ the count of false positives, $TN_i$ the count of true negatives, and $FN_i$ the count of false negatives for class *i*. By analyzing the confusion matrix, we observe *how* and *why* certain mistakes occur, thereby uncovering correlations between correct classifications and misclassifications, as well as guiding possible improvements for future models or training pipelines.

The mathematical definitions are presented below, where $N$ denotes the total number of classes.

Accuracy.

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FP_i + FN_i)} \tag{1}$$

Accuracy reflects the overall proportion of correctly classified instances. A higher accuracy corresponds to a strong correlation between the trained model's predictions and ground-truth labels across *all* classes.

Precision for class i.

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{2}$$

Precision gauges the proportion of true positives among all predicted positives for class *i*. When precision is high, it indicates that the model makes fewer false alarms and correlates predicted positives with actual positives effectively.

Recall (Sensitivity) for class i.

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{3}$$

Recall (or sensitivity) measures the proportion of true positives successfully identified among all actual positives for class *i*. High recall

signifies the model rarely misses positive cases, underscoring a strong correlation between ground-truth positives and predicted positives.

F1 Score for class i.

$$F1_i = 2\,\frac{\text{Precision}_i\ \text{Recall}i}{\text{Precision}_i + \text{Recall}i} \tag{4}$$

The F1 score harmonically balances both precision and recall for class *i*. A high F1 score signifies that the model achieves a good trade-off between not missing positive instances (*FN*) and avoiding false positives (*FP*).

Overall F1 Score (Macro-average).

$$F1_{\text{macro}} = \frac{1}{N}\sum_{i=1}^{N} F1_i \tag{5}$$

The macro-average F1 score aggregates the individual F1 scores equally across all *N* classes. This reveals how well the model performs overall, regardless of potential class imbalances.

Specificity for class i.

$$\text{Specificity}_i = \frac{TN_i}{TN_i + FP_i} \tag{6}$$

Specificity evaluates how effectively the model identifies negative instances for class *i*. A high specificity denotes strong correlation between actual negatives and predicted negatives, indicating fewer false positives.

AUC for multi-class (One-vs-Rest approach).

$$\text{AUC}_i = \int_0^1 \text{TPR}_i d(\text{FPR}_i) \tag{7}$$

AUC (Area Under the Curve) further elaborates on the true positive rate (TPR) and false positive rate (FPR) across varying classification thresholds. It offers a threshold-independent measure of a model's discriminative capacity, illuminating the correlation between the model's sensitivity (TPR) and its tendency to generate false positives (FPR). By analyzing these metrics, researchers and practitioners can develop a nuanced understanding of each model's strengths, limitations, and classification tendencies in detecting ALL. The results can highlight whether a model is particularly strong at detecting positive cases (high recall), avoiding false alarms (high precision), or maintaining robust overall performance (high accuracy and F1 scores). Consequently, drawing correlations between different metrics and confusion matrix elements aids in refining CNN architectures and training processes, ensuring that any identified weaknesses are addressed in future research or clinical implementations (see Fig. 6).

### 4.2. Results

Tables 3 and 4 displays the training loss, validation loss, validation accuracy, training accuracy, and learning rate at different epochs for the training performance. Through analysis, it is evident that our proposed approach accurately classified all cases in 7 different models namely VGG16, VGG19, ResNet50, Xception, ResNet152, DenseNet169, and EfficientNetV2B0. For evaluating the modest performance we have also plotted the accuracy plots in Figs. 7–10), and ROC curves (Figs. 13 and 14), providing clear insights into their classification capabilities. Models such as VGG16, VGG19, ResNet50, EfficientNetV2B0, Xception, and DenseNet121 achieved outstanding validation accuracies of 100 % across all metrics, including Precision, Sensitivity, Specificity, F1-Score, Accuracy, and AUC as detailed in Table 5 and it's pictorial representation is shown in 6. Notably, the classification results indicate that these models consistently recorded 100 % for precision, sensitivity, specificity, F1-score, and accuracy.Additionally, we examine the performance trends of the proposed models over different epochs, which enhances our understanding of their training dynamics, as illustrated in Tables 3 and 4 With the transfer learning techniques, the confusion metrics of the 10 CNN models have been shown in Figs. 11 and 12.

#### 4.2.1. Result comparison with different optimization methods

All the networks were trained using the Adam optimizer to test their performance. Its performance is compared to other effective optimization techniques, like SGD [50], Adadelta [51], Adam [52], and RMSProp.

[53]. The classification results for these two optimizers across seven top-performing CNN models are presented in Table 6, with evaluations conducted on the test set. The results show that the Adam optimizer outperforms the other optimization techniques.

**Table 3**
Performance of the Proposed Models during training (30 Epochs) - 1.

| Model | Epoch | Train Loss | Valid Loss | Valid (%) | Accuracy | Train (%) | Accuracy | Learning Rate |
|---|---|---|---|---|---|---|---|---|
| **EfficientNet-B0** | 1 | 1.2687 | 1.4340 | 15.80 | | 55.62 | | 0.0001 |
| | … | … | … | … | | … | | |
| | 15 | 0.9856 | 1.7901 | 20.25 | | 95.25 | | |
| | … | … | … | … | | … | | |
| | 30 | 0.0286 | 2.4605 | 28.90 | | 98.79 | | |
| **VGG-19** | 1 | 1.0755 | 12.8821 | 29.48 | | 54.58 | | 0.0001 |
| | … | … | … | … | | … | | |
| | 15 | 0.0548 | 0.3163 | 90.56 | | 98.22 | | |
| | … | … | … | … | | … | | |
| | 30 | 0.0147 | 0.0243 | 99.23 | | 99.62 | | |
| **NASNetMobile** | 1 | 1.1268 | 1.2346 | 42.39 | | 60.55 | | 0.0001 |
| | … | … | … | … | | … | | |
| | 15 | 0.0123 | 0.8436 | 66.86 | | 99.63 | | |
| | … | … | … | … | | … | | |
| | 30 | 0.0286 | 0.6137 | 79.58 | | 99.25 | | |
| **Xception** | 1 | 0.7938 | 1.0457 | 60.12 | | 72.10 | | 0.0001 |
| | … | … | … | … | | … | | |
| | 15 | 0.0102 | 0.0105 | 99.61 | | 99.66 | | |
| | … | … | … | … | | … | | |
| | 30 | 0.0034 | 0.0061 | 99.61 | | 99.88 | | |
| **DenseNet-121** | 1 | 0.9232 | 1.5411 | 52.60 | | 69.93 | | 0.0001 |
| | … | … | … | … | | … | | |
| | 15 | 0.0123 | 0.0300 | 99.04 | | 99.50 | | |
| | … | … | … | … | | … | | |
| | 30 | 0.0019 | 0.0097 | 99.81 | | 99.97 | | |

**Table 4**
Performance of the Proposed Models during training (30 Epochs) - 2.

| Model | Epoch | Train Loss | Valid Loss | Valid (%) | Accuracy | Train (%) | Accuracy | Learning Rate |
|---|---|---|---|---|---|---|---|---|
| DenseNet-169 | 1 | 0.7543 | 0.7112 | 66.28 | 74.36 | | 0.0001 | |
| | … | … | … | … | … | | | |
| | 15 | 0.0063 | 0.0320 | 99.23 | 99.75 | | | |
| | … | … | … | … | … | | | |
| | 30 | 0.0011 | 0.0114 | 99.61 | 99.97 | | | |
| **ResNet-50** | 1 | 0.7844 | 7.1554 | 29.48 | 72.74 | | 0.0001 | |
| | … | … | … | … | … | | | |
| | 15 | 0.0308 | 0.6235 | 78.42 | 99.16 | | | |
| | … | … | … | … | … | | | |
| | 30 | 0.0056 | 0.0056 | 99.61 | 100.00 | | | |
| **VGG-16** | 1 | 0.7867 | 2.5722 | 30.25 | 68.90 | | 0.0001 | |
| | … | … | … | … | … | | | |
| | 15 | 0.0171 | 0.0142 | 99.61 | 99.60 | | | |
| | … | … | … | … | … | | | |
| | 30 | 0.1711 | 0.0141 | 99.72 | 99.72 | | | |
| **ResNet-152** | 1 | 0.7459 | 2.7392 | 30.25 | 75.93 | | 0.0001 | |
| | … | … | … | … | … | | | |
| | 15 | 0.0032 | 0.6309 | 73.03 | 99.95 | | | |
| | … | … | … | … | … | | | |
| | 30 | 0.0087 | 0.0018 | 99.81 | 99.74 | | | |
| **EfficientNet-V2B0** | 1 | 1.2087 | 1.4278 | 31.79 | 57.70 | | 0.0001 | |
| | … | … | … | … | … | | | |
| | 15 | 0.0283 | 0.0041 | 100.00 | 98.99 | | | |
| | … | … | … | … | … | | | |
| | 30 | 0.0158 | 0.0030 | 99.81 | 99.47 | | | |
| **ResNet-50** | 1 | 0.7844 | 7.1554 | 29.48 | 72.74 | | 0.0001 | |
| | … | … | … | … | … | | | |
| | 15 | 0.0308 | 0.6235 | 78.42 | 99.16 | | | |
| | … | … | … | … | … | | | |
| | 30 | 0.0056 | 0.0056 | 99.61 | 100.00 | | | |

### *4.3. Comparison with existing models for ALL disease classification*

The comparison of the existing models, as shown in Table 7, reveals varying levels of performance on both the C-NMC-2019 and ISBI-2019 datasets. Jawahar et al. [5] proposed the ALNet model, which achieved a remarkable accuracy of 99.73 %, significantly outperforming all other methods. Among the previous ap-proaches, Shi et al. [54] achieved the highest accuracy on the C-NMC-2019 dataset with 87.9 % using a combination of three PNASNet-5 models and a voting mechanism. Following closely, Shah et al. [55] uti-lized a hybrid model involving AlexNet and LSTM components, reaching an accuracy of 86.6 %. Mondal et al. [56] reported a slightly lower performance with an ensemble of CNNs, attaining an accuracy of 86.2 %. Jiang et al. [57] reached 98.90 % accuracy on the ISBI-2019 dataset, using a Vision Transformer, while their ensemble model combining EfficientNetB0 and Vision Transformer achieved 99.03 %. Overall, these results highlight the effectiveness of ALNet attaining an accuracy of 86.2 %and other advanced architectures in various classification tasks.

## 5. Discussion

In this paper, we conduct an empirical analysis using various deep learning techniques for the automatic detection of acute lymphoblastic leukemia (ALL) and the classification of its subtypes into four categories. Specifically, we evaluate the performance of ten distinct DL models, including VGG16, VGG19, ResNet50, Xception, ResNet152, NASNet-Mobile, DenseNet169, and EfficientNetV2B0, to identify the model that per-forms best in detecting ALL.

Deep neural networks generally need a large volume of labeled data for effective model training. How-ever, in this study, we worked with a relatively limited dataset. Despite this challenge, we achieved 100 % accuracy in subtype classification using seven models like VGG16, VGG19, ResNet50, Xception, ResNet152, DenseNet169, and EfficientNetV2B0 by employing data augmentation techniques. Our findings suggest that deep neural networks are indeed powerful tools for detecting leukemia, even with limited data.

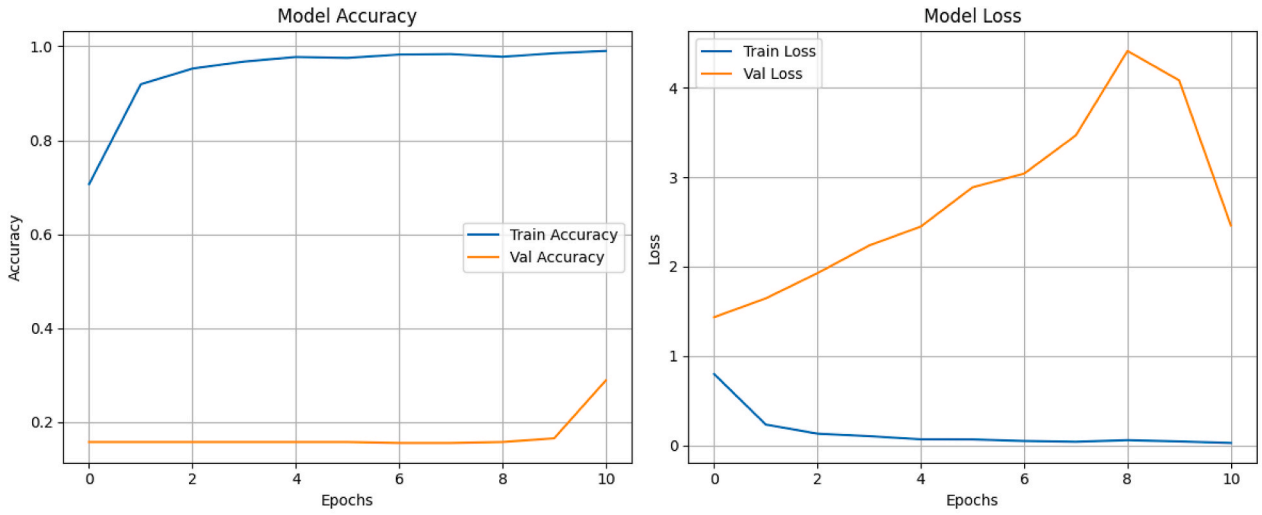
Many prior studies have focused on leukemia detection but have often overlooked subtype classification due to the inter-class similarities and variability. However, classifying the subtypes of ALL is essential for accurate diagnosis and appropriate treatment planning. One limitation of our study was the lower perfor-mance of the EfficientNetV2B0 model, which did not yield satisfactory results. Additionally, NASNetMobile and DenseNet121 showed lower accuracy, but we believe that fine-tuning these models could improve their performance. Moreover, the use of explainable AI techniques, such as Grad-CAM and Score-CAM, enhanced

model interpretability, allowing healthcare professionals to better understand the decision-making process behind predictions.

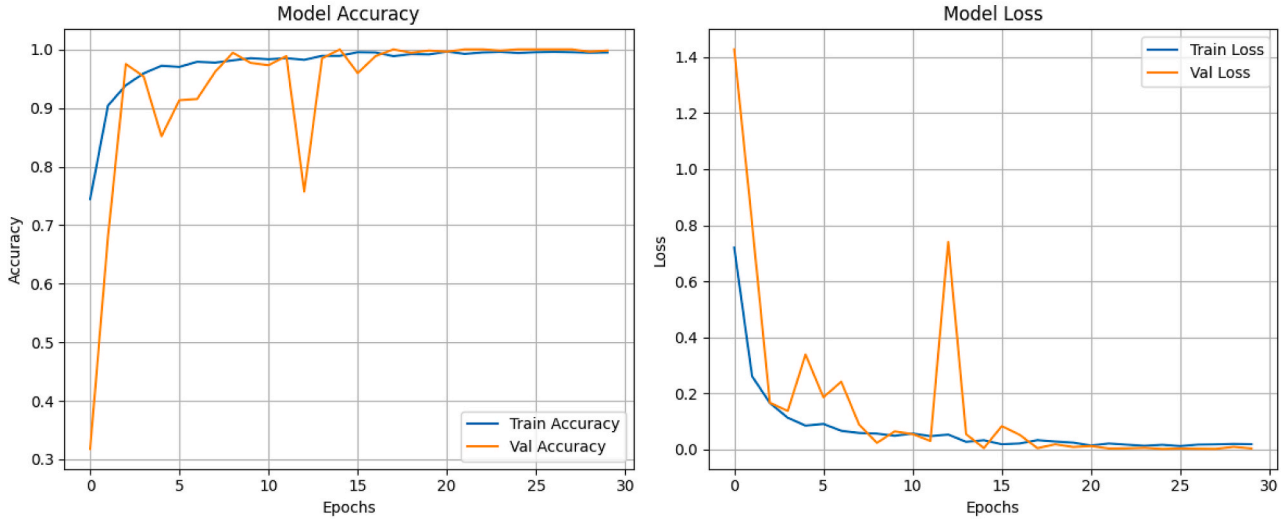
A central focus of this study is to explore various deep learning techniques, particularly convolutional neural networks (CNNs), and to evaluate their effectiveness in detecting and classifying acute lymphoblastic leukemia (ALL) based on histopathological images. We compared ten CNN architectures: VGG16, VGG19, ResNet50, Xception, ResNet152, NASNetMobile, DenseNet169, and EfficientNetV2B0. Our goal was to identify which model performs best in both detection and classification tasks, given the limited dataset available.

### *5.1. Needs of transfer learning in medical imaging*

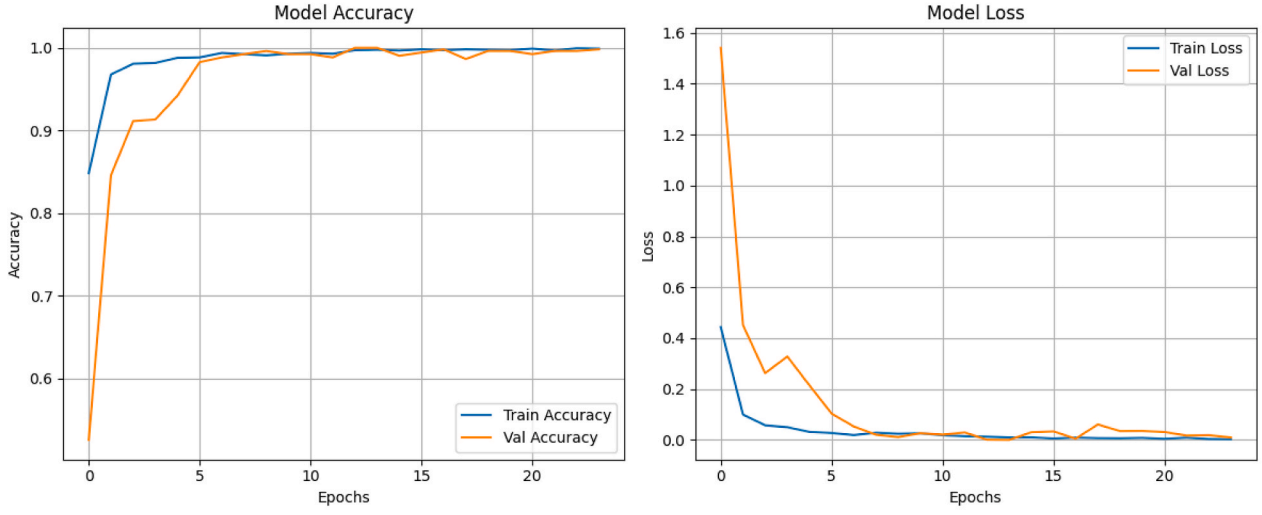
The use of transfer learning in training deep learning models has shown considerable promise in harnessing pre-trained knowledge to improve performance, particularly in specialized fields like blood cell classification. Convolutional neural networks (CNNs) typically require extensive datasets for training, which may be scarce in medical tasks such as leukemia detection. To overcome this challenge, we applied transfer learning, a technique that has become essential in deep learning, particularly for medical image analysis. Transfer learning involves utilizing pre-trained models, which have already been trained on large-scale datasets like ImageNet, to extract learned features (e.g., edges, textures, patterns) and adapt them to new tasks. This approach




(a) EfficientNet B0

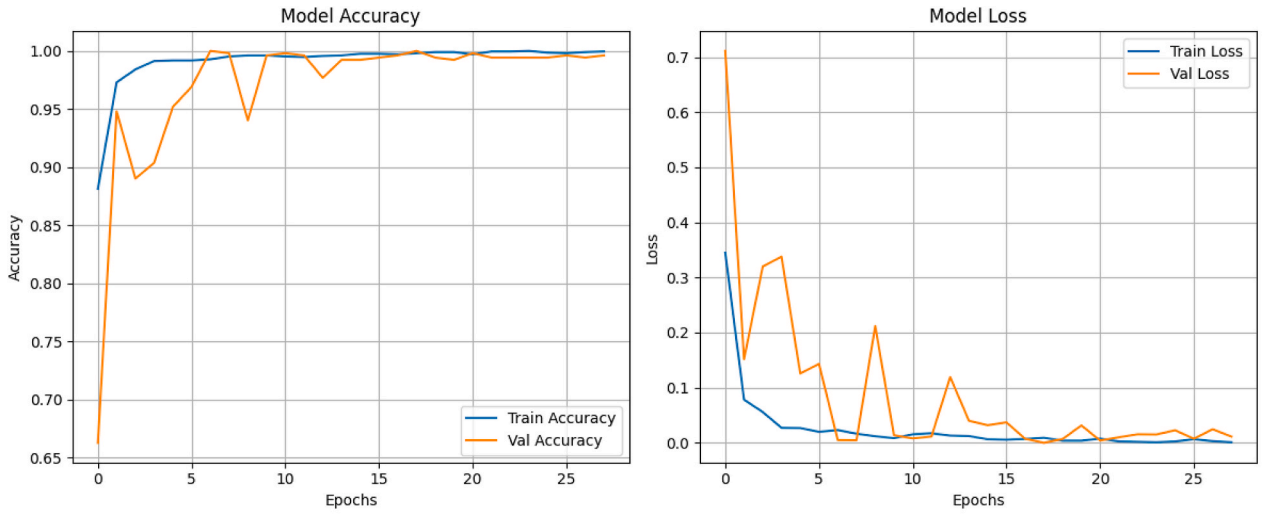



(b) EfficientNetV2 B0

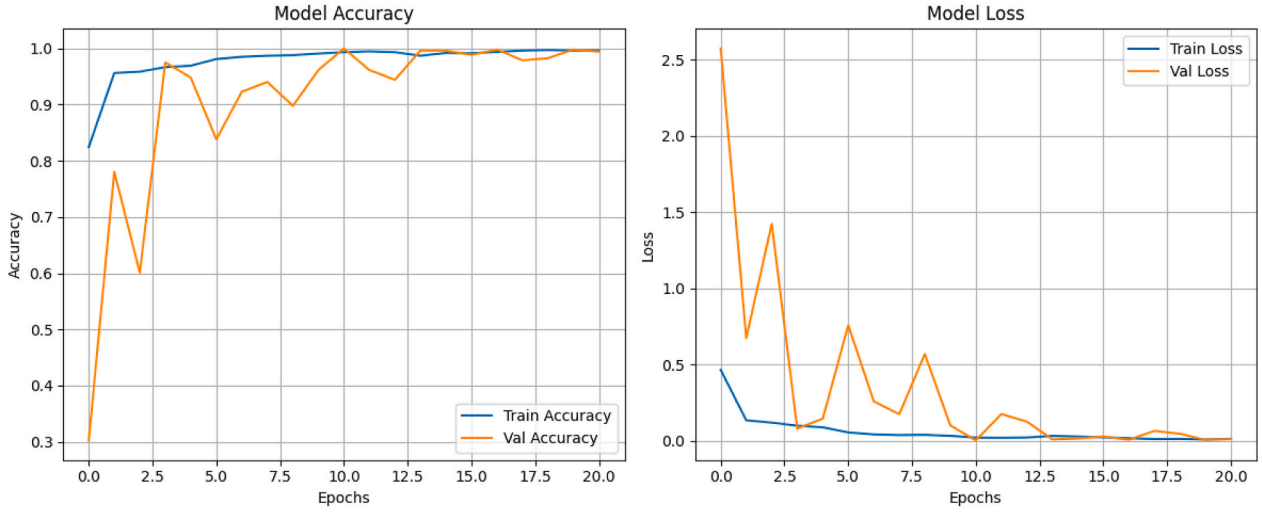



(c) DenseNet 121

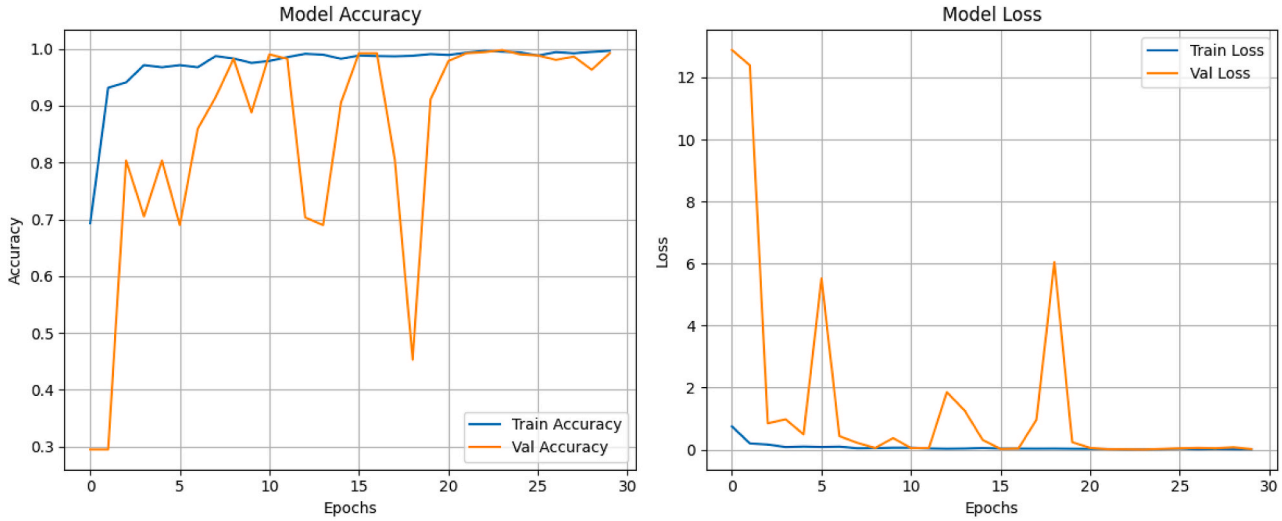
**Fig. 7.** Accuracy plots 1.




(a) DenseNet 169

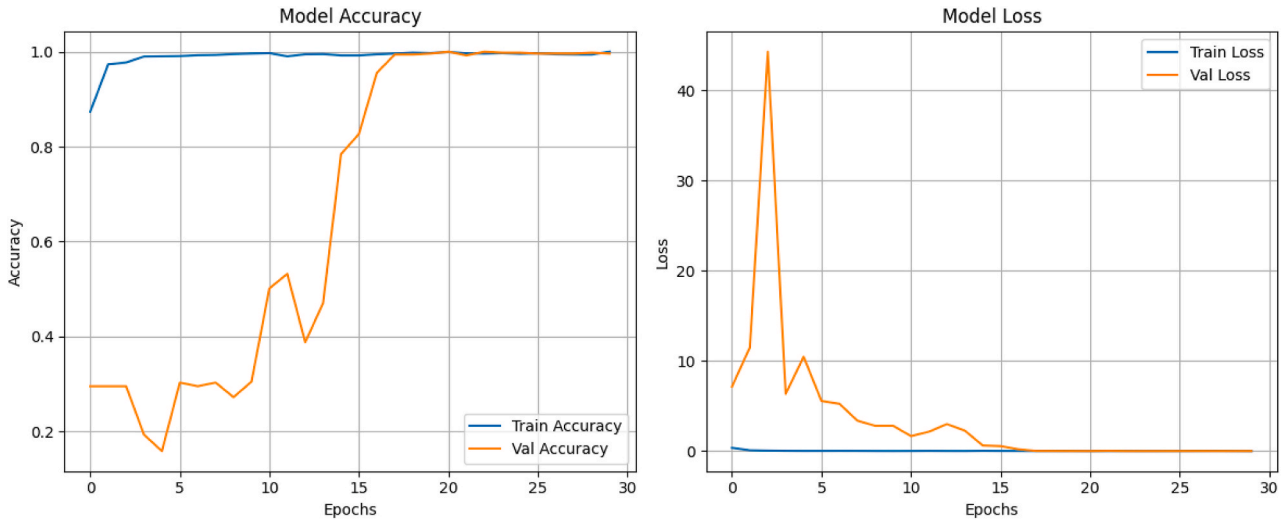



(b) VGG16

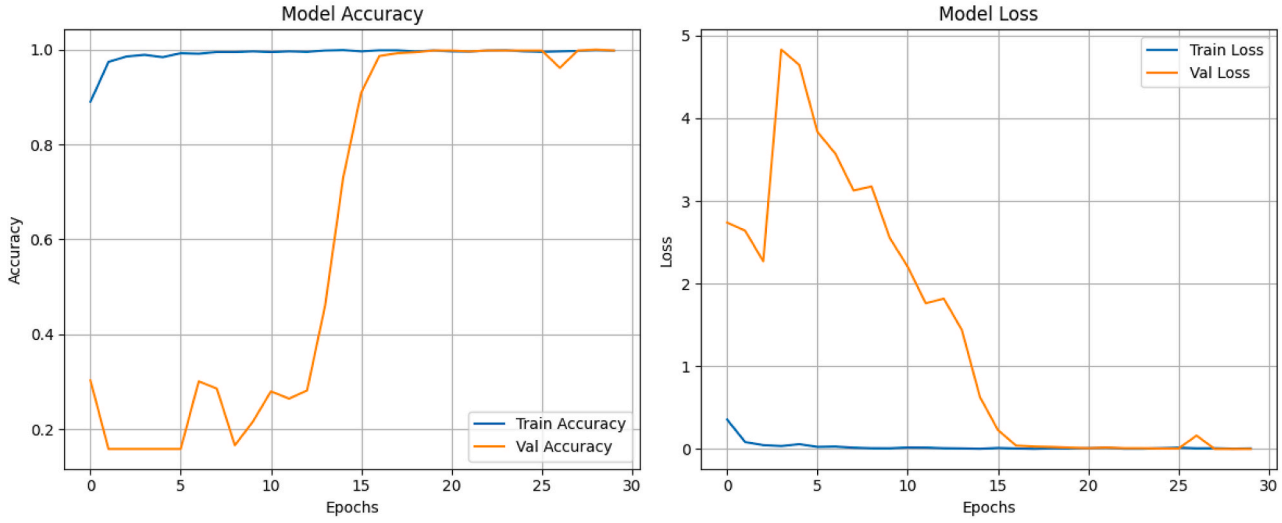



(c) VGG19

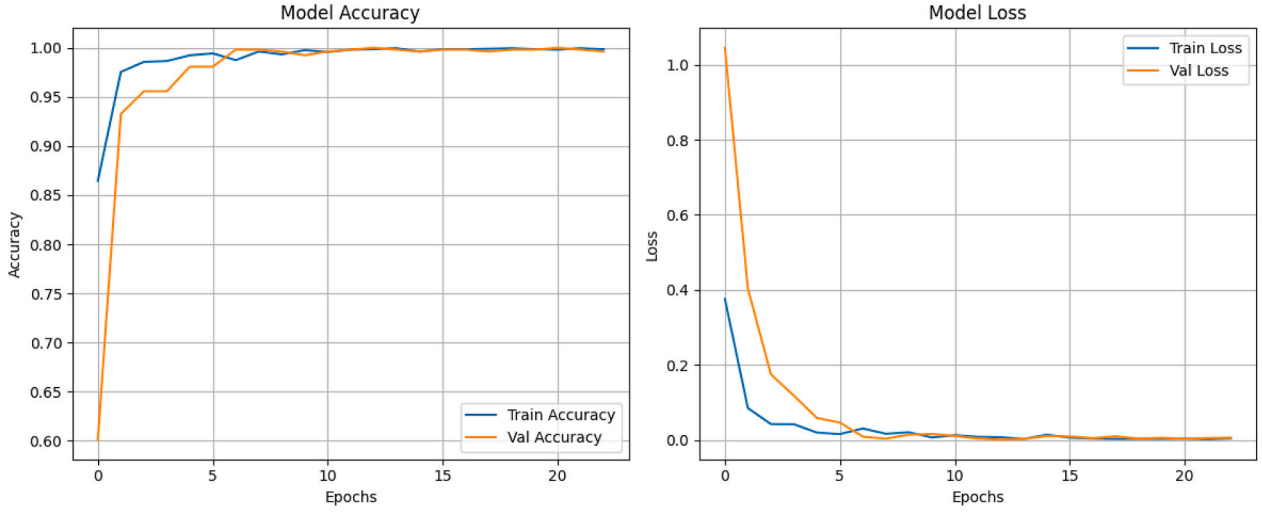
**Fig. 8.** Accuracy plots 2.




(a) ResNet50

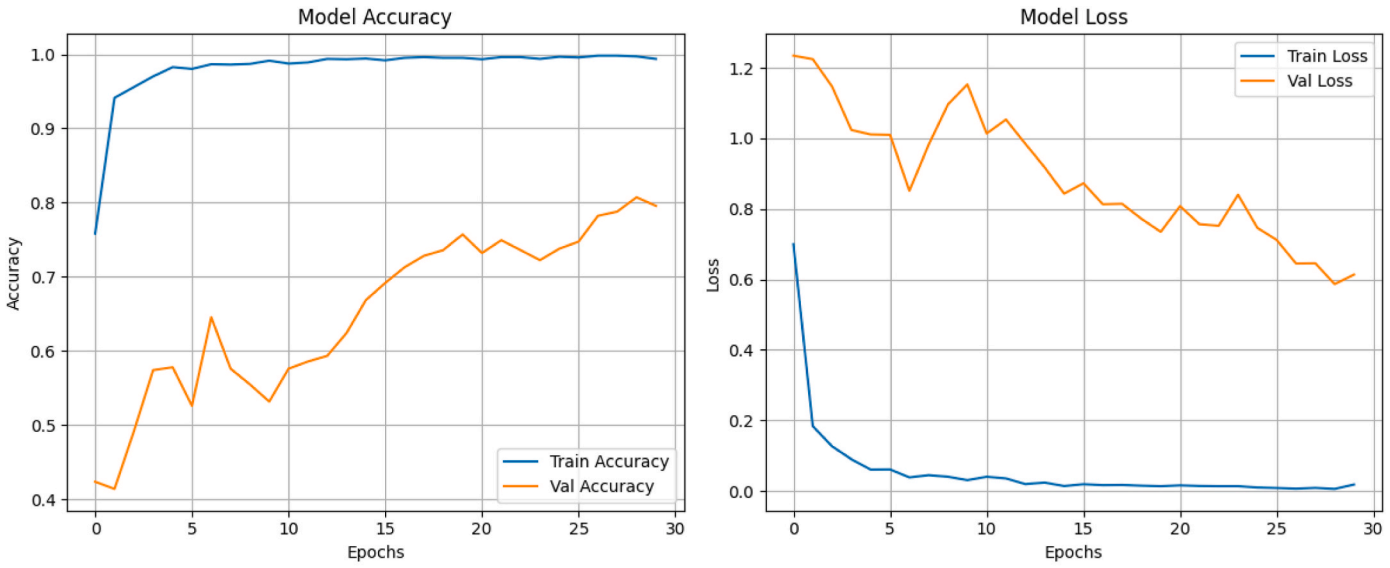



(b) ResNet152




(c) Xception

**Fig. 9.** Accuracy plots 3.




(a) NASNetMobile

**Fig. 10.** Accuracy plots 4.

**Table 5**
Performance comparison of different pre-trained model during testing.

| Model name | Precision (%) | Sens (%) | Spec (%) | F1-s (%) | Accu (%) | AUC |
|---|---|---|---|---|---|---|
| EfficientNetB0 | 60.00 | 70.00 | 80.00 | 55.00 | 72.00 | 0.68 |
| NASNetMobile | 86.00 | 81.00 | 93.00 | 78.00 | 81.00 | **0.97** |
| DenseNet121 | 99.00 | 99.00 | **100.00** | 99.00 | 99.00 | **0.98** |
| **VGG16** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **0.96** |
| **VGG19** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **0.97** |
| **ResNet50** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **0.97** |
| **Xception** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **0.96** |
| **ResNet152** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **0.96** |
| **DenseNet169** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **0.97** |
| **EfficientNetV2B0** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **0.95** |

Precision: Precision, **Sens**: Sensitivity, **Spec**: Specificity, **F1-s**: F1 Score, **Accu**: Accuracy, **AUC**: **Area Under the Curve**.

enables us to significantly reduce the need for labeled data during training while still achieving high performance in tasks such as Acute Lymphoblastic Leukemia (ALL) detection.

### 5.1.1. *Advantages of transfer learning*

- **Reduced Training Time:** Transfer learning allows for fine-tuning a pre-trained model, saving both time and computational resources compared to training a model from scratch.
- **Improved Performance:** The use of features learned from large datasets enhances the generalization ability of the model, even when the target dataset is small.
- **Better Handling of Limited Data:** Medical datasets often suffer from limited labeled data, making transfer learning an essential tool for overcoming this issue.
- **High-Quality Feature Extraction:** Pre-trained models like ResNet50, VGG16, and others are adept at extracting meaningful features from images, which are crucial for tasks such as ALL detection.

**RQ1:** *How can transfer learning with pre-trained CNN models improve the accuracy and efficiency of* Acute *Lymphoblastic Leukemia (ALL) detection compared to training models from scratch?*

**Answer:** Training a CNN from scratch requires vast amounts of labeled data, computational resources, and time. Medical datasets, particularly those for rare diseases like ALL, often lack the scale required for effective training. Transfer learning mitigates these challenges by enabling us to fine-tune models previously trained on large datasets, utilizing learned patterns that can be adapted to our specific task.

In this study, we explore a variety of CNN architectures, including simpler models like VGG16 and VGG19, deeper models such as ResNet50, and more efficient models like EfficientNetV2B0 and NAS-NetMo- bile. Each of these models brings its own strengths.

- **VGG16 and VGG19:** These models are simple but effective at feature extraction. Their simplicity allows for faster training, making them a good starting point for medical image classification tasks.
- **ResNet50 and ResNet152:** ResNet models are known for their residual learning architecture, which helps mitigate the problem of vanishing gradient during training. This is crucial for medical images, where fine distinctions are often necessary for accurate classification.
- **Xception:** Xception employs depthwise separable convolutions, making it computationally efficient while maintaining high performance. This makes it particularly useful in real-time diagnostic scenarios.
- **NASNetMobile:** Optimized for mobile devices, NASNetMobile is ideal for low-resource settings while attaining cutting-edge performance in real-time medical image analysis.
- **EfficientNetV2B0:** Known for its efficiency in balancing accuracy with computational resources, EfficientNetV2B0 is ideal when both accuracy and speed are required.

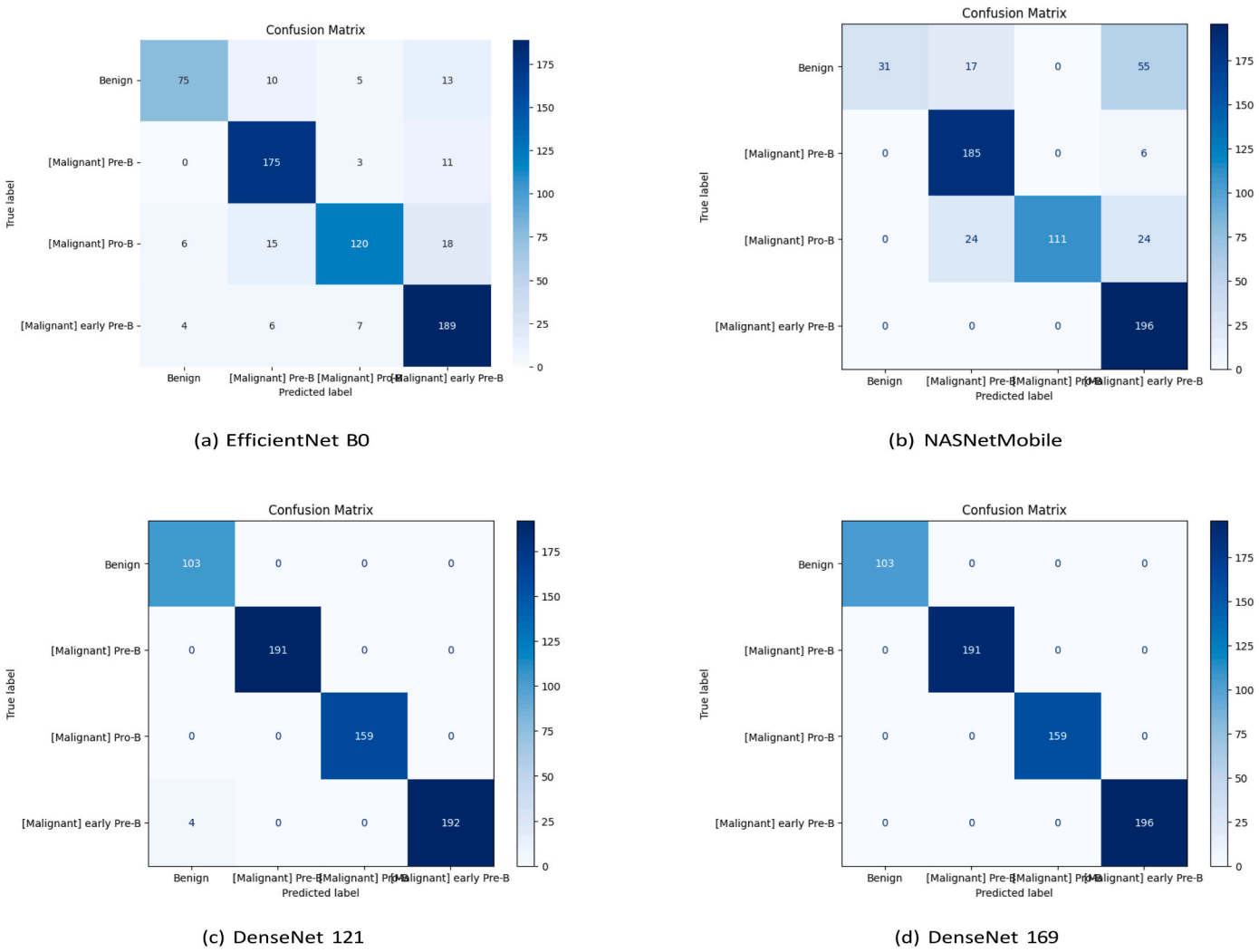



**Fig. 11.** Model architecture confusion Matrices 1.

**RQ2:** *What are the optimal criteria for selecting and evaluating CNN architectures for classifying ALL subtypes in blood smear images?*

**Answer:** By evaluating different CNN models, we aimed to identify the most suitable architecture for ALL detection and subtype classification. Each model has its own strengths, and this broad approach helps ensure we can achieve optimal performance even with a limited dataset.

### *5.1.2. Advantages of using CNNs for medical image classification*

CNNs are highly effective for image classification tasks because they can automatically extract hierarchical features from raw image data. In medical imaging, CNNs offer several advantages.

- **Identify Subtle Patterns:** CNNs can capture low-level characteristics like edges and textures and combine them into higher-level representations, which is crucial for distinguishing between different disease states.
- **Handle Complex Image Data:** Medical images, like histopathological slides, often contain complex information that traditional methods struggle to analyze. CNNs excel in capturing these complexities without requiring manually designed features.
- **Provide Robustness Against Noise:** Medical images often suffer from noise due to artifacts or variations in preparation. CNNs are adept at focusing on important patterns while ignoring irrelevant noise.

## *5.2. Roles of explainable AI (XAI) in medical imaging*

Despite their high accuracy, deep learning models, including CNNs, are often regarded as ”black-box” models, which makes it difficult for healthcare professionals to trust their predictions. To overcome this challenge, we integrated Explainable AI (XAI) techniques, such as Grad-CAM and Score-CAM, which offer insight into the model’s decision-making process. These methods highlight the image regions that had the greatest influence on the model’s prediction, thereby enhancing trust and interpretability.

**RQ3:** *How does Explainable AI (XAI) improve the interpretability and trustworthiness of CNN-based ALL* diagnostic *models?*

**Answer:** XAI ensures that healthcare professionals can trust AI predictions by providing insights into the reasoning behind each decision. This is critical in the medical domain, where decisions must be transparent and verifiable. By facilitating collaboration between clinicians and AI systems, XAI makes these technologies more acceptable and useful in clinical practice.

### *5.2.1. Advantages of explainable AI*

- **Model Transparency:** XAI techniques help clinicians interpret the model’s predictions, fostering trust in AI-based medical systems.
- **Improved Clinical Decision-Making:** By showing which areas of an image the model focuses on, XAI enables clinicians to make more

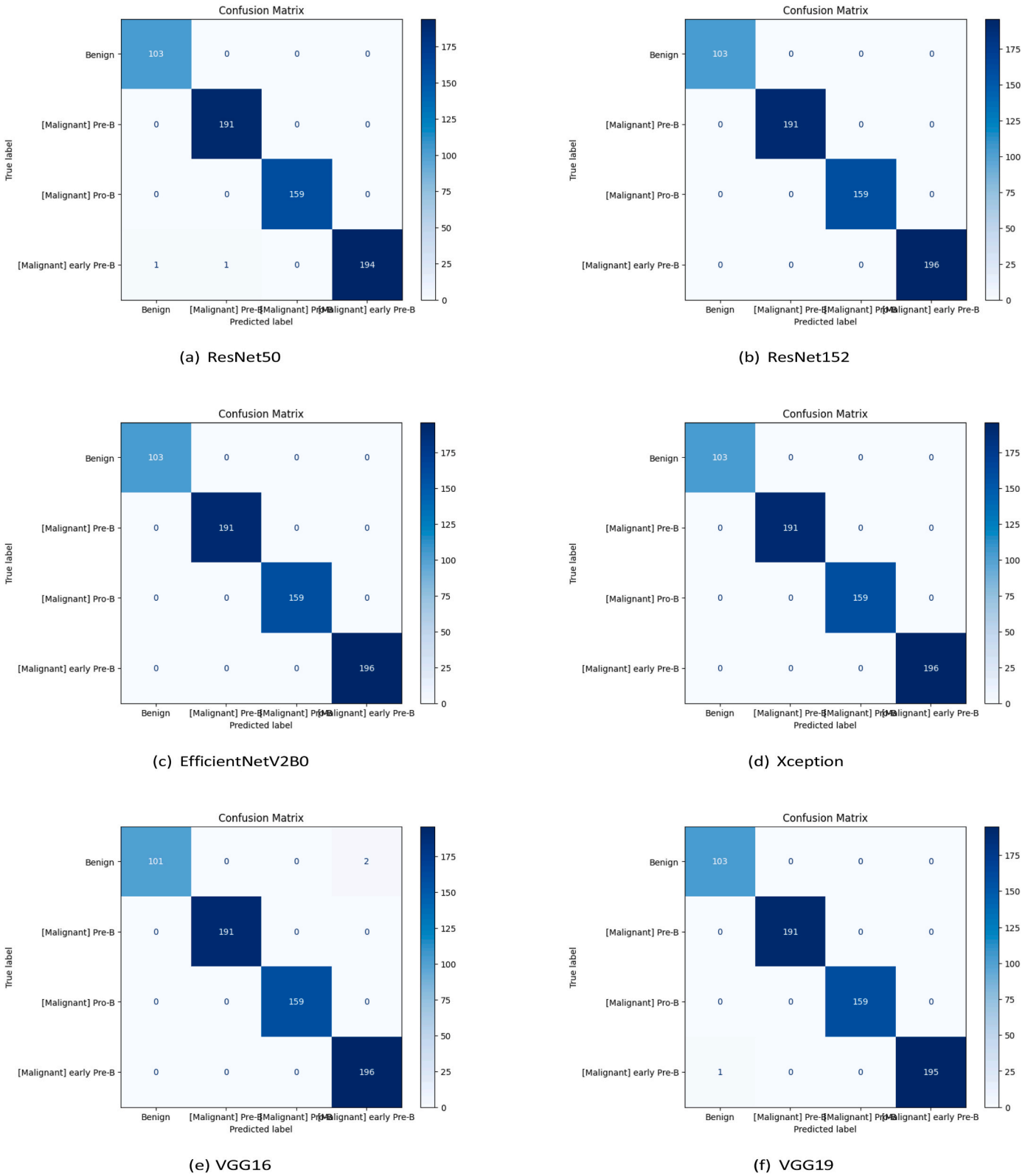



**Fig. 12.** Model architecture confusion Matrices 2.

informed decisions and identify features that may have been overlooked.

- **Model Debugging and Improvement:** XAI can also highlight areas where the model may be underperforming, providing insights for further improvements.
- **Regulatory Compliance:** In medical applications, regulations often require models to be inter-pretable, and XAI can help meet these standards.

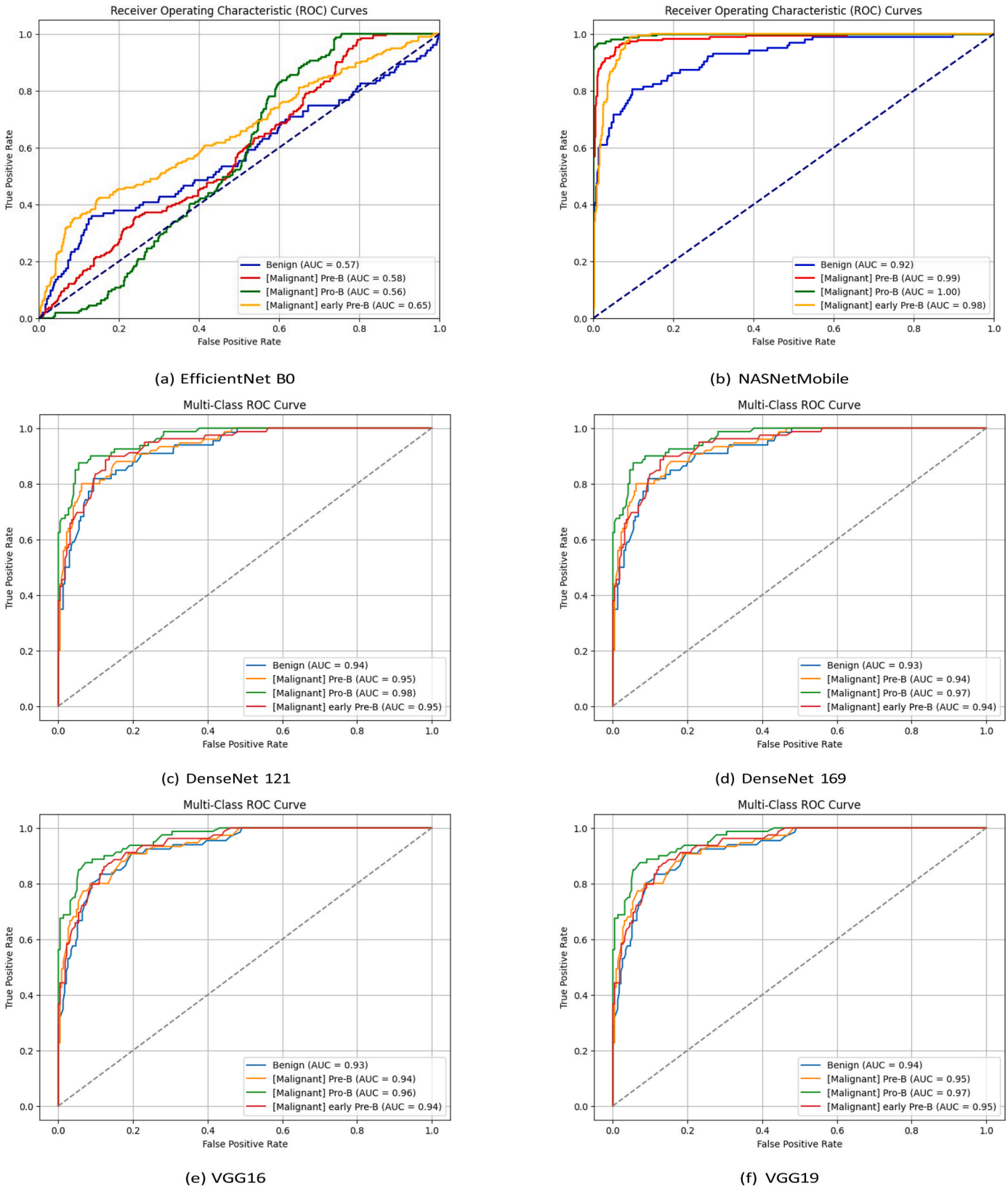



**Fig. 13.** Model architecture Graphs 1.

### 5.3. *Needs of data augmentation in medical imaging*

Given the limited size of the dataset, data augmentation played a crucial role in improving model per-formance. This technique entails artificially increasing the dataset size by applying transformations like rotations, flipping, and scaling. By exposing the model to a variety of image variations, data augmentation helps the model generalize better, mimicking real-world conditions. Without data augmentation, the

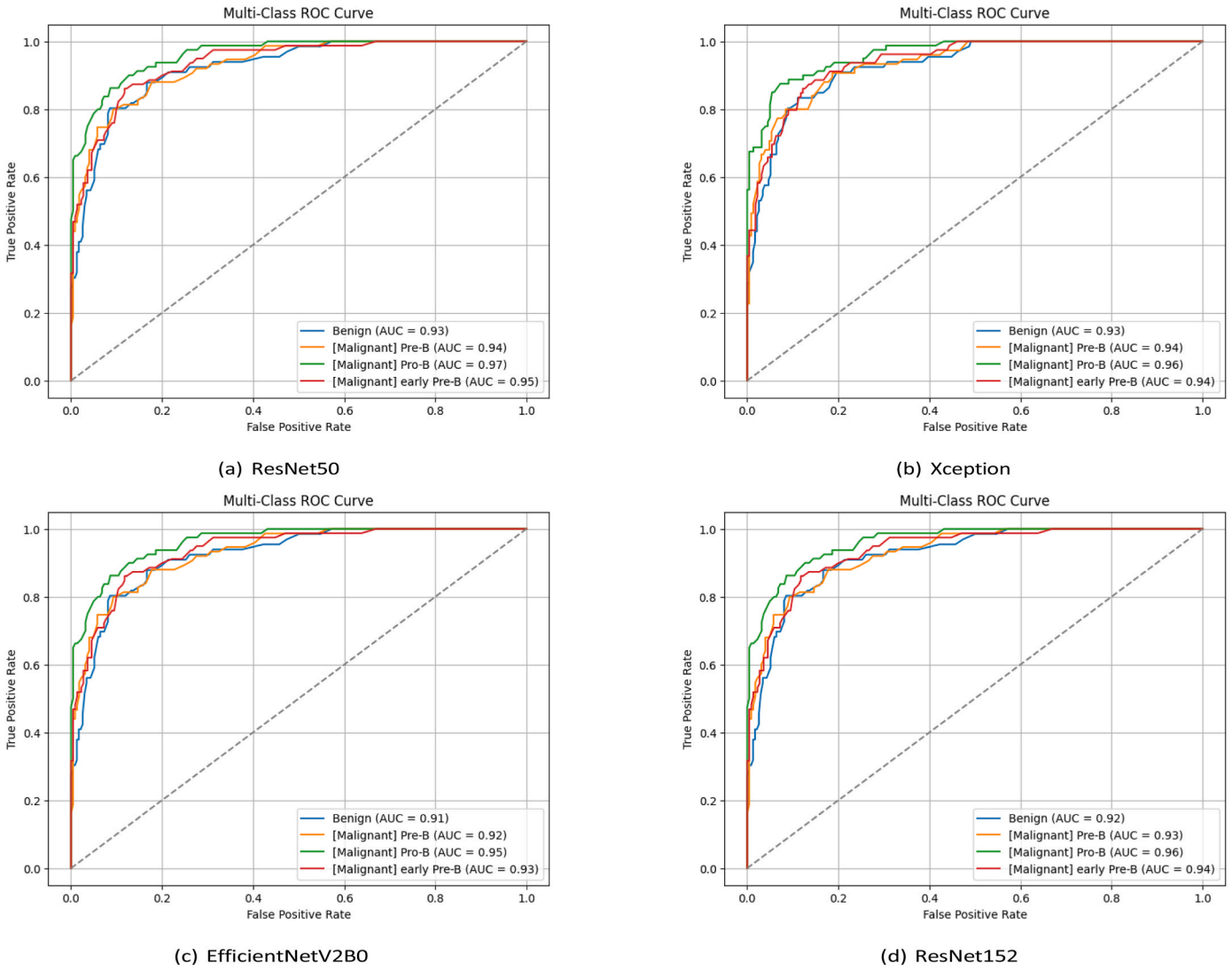



**Fig. 14.** Model architecture Graphs 2.

model is at risk of overfitting, as it would learn specific patterns from a small set of data that may not represent the diversity of cases in real-world scenarios. Data augmentation helps prevent overfitting by artificially expanding the training dataset, enabling the model to generalize better to unseen data and enhancing its accuracy. In addition to traditional methods like rotation and scaling, other augmentation techniques like SMOTE (Synthetic Minority Over-sampling Technique) can be used to balance the dataset by generating synthetic samples of the minority class. However, augmentation preserves the contextual integrity of the data, especially in complex domains like medical imaging. It reduces overfitting by exposing models to

diverse scenarios and can be tailored to domain-specific needs, such as simulating variations in texture or orientation. By enriching both majority and minority classes, augmentation provides a robust and realistic approach to addressing class imbalance effectively.

**RQ4:** *What are the key challenges in using CNNs for ALL classification, and how can strategies like data* augmentation *and hyperparameter tuning address them?*

**Answer:** The primary challenges in using Convolutional Neural Networks (CNNs) for Acute Lym-phoblastic Leukemia (ALL) classification include limited labeled medical data, variability in image quality, and the computational demands of explainable AI (XAI). These challenges can be mitigated through strate-gies such as data augmentation, which helps generate more diverse training samples, and transfer learning, which leverages pre-trained models to improve performance with limited data. Standardizing preprocess-ing methods can reduce the impact of image quality variations, ensuring more consistent model inputs. In addition, hyperparameter tuning plays a crucial role in optimizing CNN performance. The selection of appropriate hyperparameters such as learning rate, batch size, number of layers, and kernel size can signifi-cantly influence the accuracy and generalization ability of the model. Grid search, random search, and more advanced techniques like Bayesian optimization are commonly employed to systematically explore the hyper-parameter space and identify the best configurations. Fine-tuning hyperparameters enables models to adapt more effectively to the specific characteristics of medical imaging data, leading to better overall performance. Additionally, while XAI offers interpretability, its computational intensity requires further optimization to become more practical for clinical use. In conclusion, our study underscores the potential of deep learning, particularly CNNs and transfer learning, for the detection and classification of ALL subtypes, even with sparse datasets. The integration of XAI techniques further improves model transparency, fostering trust in clinical environments. However, challenges such as data variability, model generalization, computational efficiency, and hyperparameter optimization must be addressed to enhance the feasibility and effectiveness of AI applications in healthcare.

**Table 6**
Classification performance of various CNN models with different optimizers techniques during testing.

| Model | Optimizer | Precision (%) | Sensitivity (%) | Specificity (%) | F1-score | Accuracy (%) |
|---|---|---|---|---|---|---|
| VGG16 | Adadelta | 92.00 | 93.50 | 90.00 | 0.9200 | 91.00 |
| | SGD | 95.50 | 94.00 | 96.00 | 0.9440 | 95.00 |
| | RMSProp | 94.50 | 95.00 | 93.00 | 0.9400 | 92.50 |
| | **Adam** | **100.00** | **100.00** | **100.00** | **1.0000** | **100.00** |
| VGG19 | Adadelta | 90.50 | 92.00 | 89.00 | 0.9100 | 90.00 |
| | SGD | 96.00 | 95.50 | 96.50 | 0.9540 | 94.50 |
| | RMSProp | 95.00 | 94.00 | 95.00 | 0.9500 | 93.00 |
| | **Adam** | **100.00** | **100.00** | **100.00** | **1.0000** | **100.00** |
| ResNet50 | Adadelta | 91.00 | 90.00 | 92.00 | 0.9100 | 90.50 |
| | SGD | 94.00 | 92.50 | 95.00 | 0.9300 | 92.00 |
| | RMSProp | 96.00 | 94.00 | 97.00 | 0.9500 | 92.50 |
| | **Adam** | **100.00** | **100.00** | **100.00** | **1.0000** | **100.00** |
| Xception | Adadelta | 92.50 | 90.80 | 91.70 | 0.9120 | 91.00 |
| | SGD | 95.70 | 94.20 | 96.40 | 0.9490 | 93.50 |
| | RMSProp | 94.00 | 93.50 | 95.50 | 0.9200 | 92.00 |
| | **Adam** | **100.00** | **100.00** | **100.00** | **1.0000** | **100.00** |
| ResNet152 | Adadelta | 90.00 | 90.50 | 89.50 | 0.8950 | 90.00 |
| | SGD | 93.00 | 92.00 | 94.00 | 0.9200 | 91.50 |
| | RMSProp | 96.50 | 95.00 | 97.00 | 0.9500 | 93.50 |
| | **Adam** | **100.00** | **100.00** | **100.00** | **1.0000** | **100.00** |
| DenseNet169 | Adadelta | 91.50 | 90.00 | 92.50 | 0.9000 | 90.50 |
| | SGD | 94.50 | 93.00 | 95.00 | 0.9300 | 92.50 |
| | RMSProp | 95.50 | 94.00 | 96.50 | 0.9400 | 92.00 |
| | **Adam** | **100.00** | **100.00** | **100.00** | **1.0000** | **100.00** |
| EfficientNetV2B0 | Adadelta | 90.20 | 89.00 | 91.00 | 0.9000 | 90.00 |
| | SGD | 95.20 | 94.50 | 96.20 | 0.9400 | 92.80 |
| | RMSProp | 94.50 | 93.00 | 95.30 | 0.9200 | 92.50 |
| | **Adam** | **100.00** | **100.00** | **100.00** | **1.0000** | **100.00** |

**Table 7**
Comparison of Existing models.

| Source | Method | Accuracy (%) |
|---|---|---|
| Shi et al. [54] | 3 x PNASNet-5 + vote11 | 87.9 |
| Shah et al. [55] | AlexNet + LSTMDENSE + DCTLSTM | 86.6 |
| Mondal et al. [56] | Ensemble of CNN | 86.2 |
| Jawahar et al. [5] | ALNet | 99.73 |
| Saeed et al. [1] | VGG16 + ECA module | 91.0 |
| Jiang et al. [57] | EfficientNetB0 | 95.18 |
| Jiang et al. [57] | Vision Transformer | 98.90 |
| Kasani et al. [58] | NasNetLarge + VGG19 | 96.58 |
| Ghaderzadeh et al. [59] | Ensemble model using the ma-jority voting method | 98.50 |
| **Proposed Model** | **VGG16, VGG19, ResNet50, Xception, ResNet152, DenseNet169, Efficient- NetV2B0** | **100.00** |

### 5.4. Challenges and future directions

While our study yielded promising results, several challenges remain.

- **Data Quality and Variability:** Variations in image quality, including differences in resolution, lighting, and staining, can impact the performance of the model.
- **Interpretability in Complex Models:** XAI methods, while valuable, can be computationally ex-pensive and may not always provide easily interpretable explanations.
- **Limited Generalization:** Models trained on one dataset may not generalize well to other datasets or clinical settings, even with transfer learning.

Future work should focus on improving model robustness, enhancing interpretability techniques, and exploring semi-supervised or unsupervised learning approaches to reduce reliance on large labeled datasets.

### 5.5. Limitations

Despite the promising outcomes, there are several limitations in our study that need to be recognized. These limitations may influence the models' generalizability, performance, and interpretability, offering opportunities for future research to enhance the accuracy and robustness of the proposed approach.

- **Limited Dataset:** One of the primary challenges in our study was the limited size of the available dataset for training. While we utilized data augmentation techniques to mitigate this issue, the overall performance of the models may still be impacted by the lack of sufficient labeled data. A larger dataset could provide more diverse examples, helping the models generalize better and avoid overfitting.
- **Data Quality and Variability:** The dataset used in this study might suffer from inherent incon-sistencies in image quality, such as variations in resolution, lighting, and staining techniques. These factors could affect the performance of the models, as CNNs are sensitive to image quality. More-over, variations between different data sources or medical institutions could result in reduced model robustness when deployed in real-world scenarios.
- **Model Interpretability in Complex Models:** While we incorporated Explainable AI (XAI) tech-niques, such as Grad-CAM and Score-CAM, to enhance the transparency of our models, these tech-niques are still computationally expensive and may not always provide fully intuitive or easily inter-pretable results. Understanding the decision-making process behind complex deep-learning models remains a challenge, particularly when dealing with high-dimensional data like medical images.
- **Generalization Across Datasets:** Despite the use of transfer learning, which helps address the challenge of limited data, models trained on a specific dataset may not generalize well to different datasets or clinical settings. Variations in image characteristics, such as staining protocols or scanner types, can cause discrepancies in model performance. This limitation highlights the need for ongoing research to enhance the robustness of deep learning models when applied to diverse datasets.

- **Overfitting on Small Datasets:** Even with techniques like transfer learning and data augmentation, the models may still be prone to overfitting due to the small size of the training set. This can lead to high performance on the training data but reduced accuracy when evaluating the model on unseen data. More effective regularization methods or larger datasets could alleviate this issue.
- **Model Performance Variability:** While some models, such as VGG16 and EfficientNetV2B0, per-formed well in detecting ALL disease and classifying subtypes, other models, like NASNetMobile, showed lower accuracy. The performance variability between different CNN architectures indicates that the choice of model can significantly influence the results. Future work should explore methods for fine-tuning underperforming models to improve their performance.
- **Computational Resource Requirements:** Deep learning models, especially more sophisticated architectures like ResNet152 and EfficientNetV2B0, can be computationally demanding, necessitating substantial hardware resources for both training and inference. This poses a challenge for deployment in resource-limited environments, such as healthcare settings with constrained resources. Future research will need to focus on optimizing these models for efficiency while maintaining high performance.
- **Clinical Adoption and Trust:** Although our study integrated Explainable AI techniques, the adop-tion of AI-based systems in clinical practice still faces significant barriers, including trust in AI pre-dictions and integration with existing clinical workflows. Despite the potential benefits of improved diagnostic accuracy, healthcare professionals may be hesitant to rely on AI systems without a deeper understanding of their inner workings and limitations.

Future research should aim to address these limitations by exploring strategies for acquiring larger, more diverse datasets, improving model generalization across different medical imaging conditions, and enhancing the interpretability and efficiency of deep learning models for medical image classification. Additionally, the development of more robust, scalable, and efficient AI systems will be crucial for their integration into real-world clinical environments. In conclusion, our research highlights the potential of deep learning models, especially convolutional neural networks (CNNs) and transfer learning, in effectively detecting and classifying subtypes of acute lymphoblastic leukemia (ALL), even when data is limited. The incorporation of explainable AI (XAI) techniques further enhances the interpretability and trustworthiness of these models, offering promising avenues for their deployment in clinical practice. However, challenges such as data variability, model generalization, and computational efficiency remain, and addressing these issues will be essential for advancing the real-world implementation of AI in healthcare.

## 6. Conclusion and future work

A comprehensive analysis of various previously-trained DL models—VGG16, VGG19, ResNet50, Xcep-tion, ResNet152, DenseNet169, and EfficientNetV2B0—was conducted to evaluate their performance in clas-sifying acute lymphoblastic leukemia (ALL) disease subtypes from peripheral blood smear images. Utilizing a robust dataset of 3,242 images, meticulous data preparation was performed, including normalization and augmentation techniques to ensure optimal training conditions. The comparison of models across varying hyperparameters revealed that several architectures achieved a remarkable accuracy of 100 %. These results not only demonstrate the efficacy of deep learning approaches in medical image classification but also high-light the potential for significant contributions to diagnostic processes in hematology. Models like VGG16, VGG19, and ResNet50 proved particularly effective for this classification task. Our model can assist on-cologists and hematologists in enhancing screening processes while alleviating their workload. Furthermore, it lays the groundwork for developing more effective deep CNN models aimed at achieving more accurate diagnoses of all diseases. Future work may explore further optimizations, additional datasets, and real-world

applications of these models to enhance diagnostic accuracy and speed in clinical settings. Overall, this study reinforces the radical transformation brought by deep learning technologies in medical diagnostics, while in-corporating explainable AI techniques like Grad-CAM, Grad-CAM++ and ScoreCAM to foster greater trust and understanding in automated systems.

## CRediT authorship contribution statement

**Debendra Muduli:** Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Sourav Parija:** Investigation, Formal analysis, Data curation. **Suhani Kumari:** Visualization, Validation, Software, Resources. **Asmaul Hassan:** Methodology, Investigation, Data curation, Conceptualization. **Harendra S. Jangwan:** Validation, Supervision, Software, Resources. **Abu Taha Zamani:** Writing – review & editing, Funding acquisition, Formal analysis, Conceptualization. **Sk. Mohammed Gouse:** Writing – review & editing, Validation, Supervision, Software. **Banshidhar Majhi:** Software, Formal analysis, Data curation. **Nikhat Parveen:** Validation, Supervision, Software.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number ‘‘NBU-FFR-2025-1850-02’’.

## Data availability

The data that has been used is confidential.

## References

[1] Saeed A, Shoukat S, Shehzad K, Ahmad I, Eshmawi A, Amin AH, Tag-Eldin E. A deep learning- based approach for the diagnosis of acute lymphoblastic leukemia. Electronics 2022;11(19):3168.
[2] Shafique S, Tehsin S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technol Cancer Res Treat 2018;17:1533033818802789.
[3] Ghorpade N, Bale AS, Suman S, Divya V, Mandal S, Parashivamurthy C. Acute lymphoblastic leukemia detection employing deep learning and transfer learning techniques. In: 2024 international conference on advances in computing, communication and applied informatics (ACCAI). IEEE; 2024. p. 1–6.
[4] Genovese A, Hosseini MS, Piuri V, Plataniotis KN, Scotti F. Acute lymphoblastic leukemia detection based on adaptive unsharpening and deep learning. In: ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2021. p. 1205–9.
[5] Jawahar M, Sharen H, Gandomi AH, et al. Alnett: a cluster layer deep convolutional neural network for acute lymphoblastic leukemia classification. Comput Biol Med 2022;148:105894.
[6] Sharma SK, Muduli D, Rath A, Dash S, Panda G, Shankar A, Dobhal DC. Discrete ripplet-ii transform feature extraction and metaheuristic-optimized feature selection for enhanced glaucoma de- tection in fundus images using least square-support vector machine. Multimed Tool Appl 2024:1–33.
[7] Muduli D, Dash R, Majhi B. Automated breast cancer detection in digital mammograms: a moth flame optimization based elm approach. Biomed Signal Process Control 2020;59:101912.
[8] Muduli D, Dash R, Majhi B. Automated diagnosis of breast cancer using multi-modal datasets: a deep convolution neural network based approach. Biomed Signal Process Control 2022;71:102825.
[9] Muduli D, Dash R, Majhi B. Fast discrete curvelet transform and modified pso based improved evolutionary extreme learning machine for breast cancer detection. Biomed Signal Process Control 2021;70:102919.

[10] Sharma SK, Muduli D, Priyadarshini R, Kumar RR, Kumar A, Pradhan J. An evolutionary sup- ply chain management service model based on deep learning features for automated glaucoma detection using fundus images. Eng Appl Artif Intell 2024;128:107449.

[11] Madhukar M, Agaian S, Chronopoulos AT. Deterministic model for acute myelogenous leukemia classification. In: 2012 IEEE international conference on systems, man, and cybernetics (SMC). IEEE; 2012. p. 433–8.

[12] Setiawan A, Harjoko A, Ratnaningsih T, Suryani E, Palgunadi S, et al. Classification of cell types in acute myeloid leukemia (aml) of m4, m5 and m7 subtypes with support vector machine classifier. In: 2018 international conference on information and communications technology (ICOIACT). IEEE; 2018. p. 45–9.

[13] Hosseini A, Eshraghi MA, Taami T, Sadeghsalehi H, Hoseinzadeh Z, Ghaderzadeh M, Rafiee M. A mobile application based on efficient lightweight cnn model for classification of b-all cancer from non-cancerous cells: a design and implementation study. Inform Med Unlocked 2023;39:101244.

[14] Ghaderzadeh M, Aria M, Hosseini A, Asadi F, Bashash D, Abolghasemi H. A fast and efficient cnn model for b-all diagnosis and its subtypes classification using peripheral blood smear images. Int J Intell Syst 2022;37(8):5113–33.

[15] Laosai J, Chamnongthai K. Acute leukemia classification by using svm and k-means clustering. In: 2014 international electrical engineering congress (iEECON). IEEE; 2014. p. 1–4.

[16] Patel N, Mishra A. Automated leukaemia detection using microscopic images. Procedia Comput Sci 2015;58:635–42.

[17] Abbasi EY, Deng Z, Ali Q, Khan A, Shaikh A, Al Reshan MS, Sulaiman A, Alshahrani H. A machine learning and deep learning-based integrated multi-omics technique for leukemia prediction. Heliyon 2024;10(3):e25369.

[18] Rangini M, Pundir S, Soudagar MEM, Vekariya D, Patil H, Rajkumar S. Texture-based feature extraction and machine learning model for the detection of acute lymphoblastic leukemia. In: 2024 5th international conference on mobile computing and sustainable informatics (ICMCSI). IEEE; 2024. p. 505–11.

[19] Liu Y, Chen P, Zhang J, Liu N, Liu Y. Weakly supervised ternary stream data augmentation fine- grained classification network for identifying acute lymphoblastic leukemia. Diagnostics 2021;12(1):16.

[20] Ansari S, Navin AH, Sangar AB, Gharamaleki JV, Danishvar S. A customized efficient deep learning model for the diagnosis of acute leukemia cells based on lymphocyte and monocyte images. Electronics 2023;12(2):322.

[21] Kumar N, Hashmi A, Gupta M, Kundu A. Automatic diagnosis of covid-19 related pneumonia from cxr and ct-scan images. Eng Technol Appl Sci Res 2022;12(1):7993–7.

[22] El-Rashidy N, ElSayed NE, El-Ghamry A, Talaat FM. Retracted article: prediction of gestational diabetes based on explainable deep learning and fog computing. Soft Comput 2022;26(21):11435–50.

[23] Ramaneswaran S, Srinivasan K, Vincent PDR, Chang C-Y. Hybrid inception v3 xgboost model for acute lymphoblastic leukemia classification. Comput Math Methods Med 2021;2021(1):2577375.

[24] Pal-czyn′ski K, Smigiel S, Gackowska M, Ledzin′ski D, Bujnowski S, Lutowski Z. Iot application of transfer learning in hybrid artificial intelligence systems for acute lymphoblastic leukemia classification. Sensors 2021;21(23):8025.

[25] Talaat FM, Gamel SA. Machine learning in detection and classification of leukemia using c- nmc leukemia. Multimed Tool Appl 2024;83(3):8063–76.

[26] Rahmani AM, Khoshvaght P, Alinejad-Rokny H, Sadeghi S, Asghari P, Arabi Z, Hosseinzadeh M. A diagnostic model for acute lymphoblastic leukemia using metaheuristics and deep learning methods. arXiv preprint arXiv:2406.18568. 2024.

[27] Alim MS, Bappon SD, Sabuj SM, Islam MJ, Tarek MM, Azam MS, Islam MM. Integrating convolutional neural networks for microscopic image analysis in acute lymphoblastic leukemia classification: a deep learning approach for enhanced diagnostic precision. Systems and Soft Computing 2024;6:200121.

[28] Genovese A, Piuri V, Scotti F. A decision support system for acute lymphoblastic leukemia detection based on explainable artificial intelligence. Image Vis Comput 2024;151:105298.

[29] Pawar J, Bhosle AA, Gupta P, Shiyal HM, Borate VK, Mali YK. Analyzing acute lymphoblas- tic leukemia across multiple classes using an enhanced deep convolutional neural network on blood smear. In: 2024 IEEE international conference on information technology, electronics and intelligent communication systems (ICITEICS). IEEE; 2024. p. 1–6.

[30] Lalithkumar K, Priyanga M, Sandhya S, Karthiga M, et al. Capsenet: deep learning based acute lymphoblastic leukemia detection approach. In: 2024 8th international conference on I-SMAC (IoT in Social, mobile, analytics and Cloud)(I-SMAC). IEEE; 2024. p. 1577–84.

[31] Kumar A, Kumar N, Kuriakose J, Sisodia PS. A deep transfer learning based approaches for the detection and classification of acute lymphocytic leukemia using microscopic images. Multimed Tool Appl 2024:1–25.

[32] Khan Tusar MTH, Islam MT, Sakil AH, Khandaker M, Hossain MM. An intelligent teledi- agnosis of acute lymphoblastic leukemia using histopathological deep learning. Journal of Computing Theories and Applications 2024;2(1):1–12.

[33] Awad A, Hegazy M, Aly SA. Early diagnoses of acute lymphoblastic leukemia using yolov8 and yolov11 deep learning models. arXiv preprint arXiv:2410.10701. 2024.

[34] Senthil K, Monikaa R, Gopinath M, Vishwa S, Sivagami S, et al. Deeplymphodetect: leveraging deep learning techniques for acute lymphoblastic leukemia detection in blood cells. In: 2024 IEEE international conference on smart power control and renewable energy (ICSPCRE). IEEE; 2024. p. 1–6.

[35] Dangore M, Ashwini S, Chendke AG, Shirbhate R, Mali YK, Borate VK. Multi-class inves- tigation of acute lymphoblastic leukemia using optimized deep convolutional neural network on blood smear images. In: 2024 MIT art, design and technology school of computing international conference (MITADTSoCiCon). IEEE; 2024. p. 1–6.

[36] Park S, Park YH, Huh J, Baik SM, Park DJ. Deep learning model for differentiating acute myeloid and lymphoblastic leukemia in peripheral blood cell images via myeloblast and lymphoblast classification. Digital Health 2024;10:20552076241258079.

[37] Cheng F-M, Lo S-C, Lin C-C, Lo W-J, Chien S-Y, Sun T-H, Hsu K-C. Deep learning assists in acute leukemia detection and cell classification via flow cytometry using the acute leukemia orientation tube. Sci Rep 2024;14(1):8350.

[38] Abougarair AJ, Alshaibi M, Alarbish AK, Qasem MA, Abdo D, Qasem O, Thabit F, Can O. Blood cells cancer detection based on deep learning. Journal of Advances in Artificial Intelligence 2024;2(1):108–21.

[39] Alsaykhan LK, Maashi MS. A hybrid detection model for acute lymphocytic leukemia using support vector machine and particle swarm optimization (svm-pso). Sci Rep 2024;14(1):23483.

[40] Jawahar M, Anbarasi LJ, Narayanan S, Gandomi AH. An attention-based deep learning for acute lymphoblastic leukemia classification. Sci Rep 2024;14(1):17447.

[41] Huang M-L, Huang Z-B. An ensemble-acute lymphoblastic leukemia model for acute lymphoblastic leukemia image classification. Math Biosci Eng 2024;21(2):1959–78.

[42] Aria M, Ghaderzadeh M, Bashash D, Abolghasemi H, Asadi F, Hosseini A. Acute lymphoblastic leukemia (all) image dataset. Kaggle; 2021.

[43] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. arXiv preprint arXiv:1409.1556.

[44] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.

[45] Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–8.

[46] Tan M. Efficientnet: rethinking model scaling for convolutional neural networks. 2019. p. 6105–14. arXiv preprint arXiv:1905.11946.

[47] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 8697–710.

[48] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.

[49] Tan M, Le Q. Efficientnetv2: smaller models and faster training. In: International conference on machine learning. PMLR; 2021. p. 10096–106.

[50] Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: International conference on machine learning. PMLR; 2013. p. 1139–47.

[51] Zeiler MD. Adadelta: an adaptive learning rate method. 2012. arXiv preprint arXiv:1212.5701.

[52] Kingma DP. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.

[53] Hinton G, Srivastava N, Swersky K. Lecture 6a overview of mini–batch gradient descent. Coursera Lecture slides 2012. https://class.coursera.org/neuralnets-2012-001/lecture.

[54] Shi T, Wu L, Zhong C, Wang R, Zheng W. Ensemble convolutional neural networks for cell classifi- cation in microscopic images. In: ISBI 2019 C-NMC challenge: classification in cancer cell imaging: select proceedings. Springer; 2019. p. 43–51.

[55] Shah S, Nawaz W, Jalil B, Khan HA. Classification of normal and leukemic blast cells in b-all cancer using a combination of convolutional and recurrent neural networks. In: ISBI 2019 C-NMC challenge: classification in cancer cell imaging: select proceedings. Springer; 2019. p. 23–31.

[56] Mondal C, Hasan MK, Jawad MT, Dutta A, Islam MR, Awal MA, Ahmad M. Acute lymphoblastic leukemia detection from microscopic images using weighted ensemble of convolutional neural networks. arXiv preprint arXiv:2105.03995 2021.

[57] Jiang Z, Dong Z, Wang L, Jiang W. Method for diagnosis of acute lymphoblastic leukemia based on vit-cnn ensemble model. Comput Intell Neurosci 2021;2021(1):7529893.

[58] Kasani PH, Park S-W, Jang J-W. An aggregated-based deep learning method for leukemic b- lymphoblast classification. Diagnostics 2020;10(12):1064.

[59] Ghaderzadeh M, Hosseini A, Asadi F, Abolghasemi H, Bashash D, Roshanpoor A. Automated detection model in classification of b-lymphoblast cells from normal b-lymphoid precursors in blood smear microscopic images based on the majority voting technique. Sci Program 2022;2022(1):4801671.