

An Intelligent Telediagnosis of Acute Lymphoblastic Leukemia using Histopathological Deep Learning

Md. Taufiqul Haque Khan Tusar ^{1,*}, Md. Touhidul Islam ², Abul Hasnat Sakil ¹, M N Huda Nahid Khandaker ¹, and Md. Monir Hossain ¹

¹ Department of Computer Science and Engineering, City University, Dhaka 1216, Bangladesh; e-mail : taufiqhantusar@gmail.com; abulhasnatsakil.cu@gmail.com; nahid.cse52@gmail.com; monir52cse.cu@gmail.com

² Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh; e-mail : touhid000cse@gmail.com

* Corresponding Author : Md. Taufiqul Haque Khan Tusar

Abstract: Leukemia, a global health challenge characterized by malignant blood cell proliferation, demands innovative diagnostic techniques due to its increasing incidence. Among leukemia types, Acute Lymphoblastic Leukemia (ALL) emerges as a particularly aggressive form affecting diverse age groups. This study proposes an advanced mechanized system utilizing Deep Neural Networks for detecting ALL blast cells in microscopic blood smear images. Achieving a remarkable accuracy of 97% using MobileNetV2, our system demonstrates high sensitivity and specificity in identifying multiple ALL subtypes. Furthermore, we introduce cutting-edge telediagnosis software facilitating real-time support for clinicians in promptly and accurately diagnosing various ALL subtypes from microscopic blood smear images. This research aims to enhance leukemia diagnosis efficiency, which is crucial for the timely intervention and managing this life-threatening condition.

Keywords: Acute Lymphoblastic Leukemia; Deep Learning; Image Processing; Healthcare; Telediagnosis.

1. Introduction

Blood is a complex biofluid containing an assortment of cellular and acellular elements crucial for sustaining human health. The primary blood constituents include plasma, leukocytes, erythrocytes, and platelets. Plasma is the predominant blood fraction, accounting for approximately 55% of the total blood volume. It is characterized by a heterogeneous mixture consisting primarily of water along with various solutes such as ions, proteins, lipids, hormones, and metabolic end products. The remaining 45% of the blood volume comprises erythrocytes (red blood cells), leukocytes (white blood cells), and platelets (thrombocytes). Erythrocytes represent over 40% of the entire blood volume; their concentration ranges between 4-6 million cells per microliter in healthy adults. Their primary function is to transport oxygen from the lungs to body tissues while simultaneously carrying carbon dioxide back to be exhaled[1]. Leukocytes are a diverse group of cells that play a crucial role in immune defense against foreign antigens and infections. Their concentration typically ranges from 4,500 to 11,000 cells per microliter of blood[2]. Disruptions or deficiencies in any component within the blood system can lead to various health complications. Acute Lymphoblastic Leukemia (ALL), a highly hazardous category of malignancy affecting either the bone marrow or circulating blood cells, is one such severe condition. It occurs across all age groups and affects both pediatric and adult populations alike. ALL is characterized by an excessive accumulation of immature lymphocytes (or blast cells) originating from hematopoietic stem cells residing in the bone marrow[3]. These abnormal cell proliferations interfere with normal blood production processes, leading to immunosuppression, increased susceptibility to infections, bleeding disorders due to impaired platelet function as well as anemia caused by decreased erythrocyte production. Lymphocytes are another type of white blood cell that plays a fundamental role in the immune system. They can be categorized into three primary types -

Received: March, 11th 2024

Revised: May, 8th 2024

Accepted: May, 9th 2024

Published: May, 13th 2024



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) licenses (<https://creativecommons.org/licenses/by/4.0/>)

normal, atypical, and reactive lymphocytes based on their morphological features observed under microscopic examination conducted by skilled pathologists[4]. In ALL cases specifically:

1. L1 subtype: The smallest cell type with homogeneous size & chromatin texture
2. L2 subtype: Larger than L1 & displays nuclear heterogeneity
3. L3 subtype: Features larger vacuoles distributed among the cells compared to those present in L1[5], [6].

According to data published by the International Agency for Research on Cancer (IARC), an affiliate of the World Health Organization (WHO) [7], Leukemia was responsible for approximately 4,37,033 cases worldwide in 2018, with nearly 3,03,006 fatalities. The global incidence rate of this malignancy stood at 5.2 per 100,000 individuals, and its mortality rate at 3.5 per 100,000 individuals. Figure 1 shows the prevalence and death due to ALL globally over the last 30 years (1990 - 2019). Where only in 2019, around 975K were affected, and around 47.5K died. In response to this pressing health concern, a comprehensive research investigation was conducted within the Pediatric Hematology and Oncology Department at Dhaka Shishu Hospital in Bangladesh between January 2014 and December 2016[8]. The accurate diagnosis of acute lymphoblastic Leukemia (ALL) necessitates a comprehensive evaluation of both bone marrow and blood smear morphology by experienced hematopathologists. However, challenges persist regarding standardization and inter-observer variability, leading to potential inaccuracies when pathologists employ manual detection methods [9]. Inter- and intra-class variations among pathologists further exacerbate these issues, resulting in a low agreement rate of only 76.6% during leukemia diagnosis. Furthermore, intelligent detection of leukemic blast cells presents additional challenges due to the inherent complexity of white blood cells (WBCs), which encompass asymmetrical borders as well as textural resemblances with other constituents present within blood samples [10]. Effective data preparation is essential to enhance the learnability of deep learning (DL) models for ALL detection. The precision and temporal intricacy associated with accurately identifying ALL subtypes rely heavily on the quality and relevance of extracted attributes utilized for pixel-wise categorization model training.

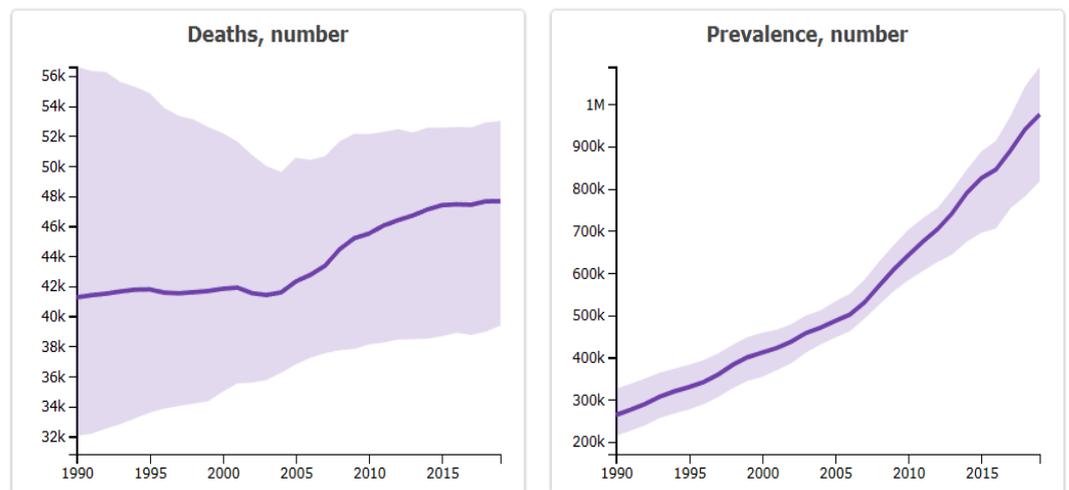


Figure 1. Death & Prevalence of ALL from 1990 – 2019

This study implements and experiments with several transfer learning modes, namely MobileNetV2, ResNet50, VGG19, and a base ConvNet, to develop a DNN model that can address the challenges of ALL subtypes detection with high performance. Their established efficacy in DNN development underpins the selection of these transfer learning models in this research. Transfer learning leverages pre-trained models on large datasets, allowing for knowledge transfer from source tasks to target tasks, thus significantly reducing computational costs and training time. Specifically, the CNN-based MobileNetV2, ResNet50, and VGG19 were chosen due to their demonstrated versatility and effectiveness across various computer vision tasks, including image classification, object detection, and feature extraction. Finally, achieves an impressive accuracy rate of 97% using MobileNetV2 from an imbalanced

nature multiclass dataset. Notably, this lightweight model exhibits robust performance even when dealing with imbalanced datasets without compromising accuracy. Moreover, this approach offers real-time tediagnosis support through a web-based application platform—an effective solution addressing limitations encountered by traditional diagnostic methods.

2. Literature Review

Researchers have explored various methodologies to identify leukemia cells from peripheral blood and bone marrow smears, including Machine Learning (ML), medical image processing techniques, and Deep Neural Networks (DNNs).

The study [3] proposed a fully automated system that employed the YOLO v4 algorithm for detecting ALL blast cells in microscopic blood smear images. Their approach achieved impressive results with Mean Average Precision (MAP) rates of 96.06% for the ALL-IDB1 dataset and 98.7% for the CNMC 2019 dataset.

In [6], the authors proposed a three-part system designed for leukemia cell detection in blood smears using datasets such as ALL-IDB1 and ALL-IDB2. The first mechanism involved a fusion of features extracted from Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and Fuzzy Color Histogram (FCH) techniques. These features were then utilized with both Artificial Neural Network (ANN) and Feed-Forward Neural Network (FFNN) models, achieving 100% and 98.11% precision rates, respectively. Next, transfer learning was employed using CNN models like AlexNet, GoogleNet, and ResNet-18. All these models achieved an accuracy rate of 100%. The third part combined CNNs with Support Vector Machines (SVMs) for feature map classification; specifically, AlexNet + SVM and ResNet-18 + SVM attained 100% accuracy, while Goog-LeNet + SVM reached 98.1% accuracy. However, there are concerns regarding the robustness of the data preparation pipeline due to limited denoising techniques used during pre-processing.

The study [11] proposed a multi-step deep learning (DL) framework for automated leukemia cell segmentation in bone marrow images. The pipeline involved several stages: initial segmentation, fine-tuned region-based convolutional neural network (FRCNN) training, manual feature extraction, and subsequent training of multiple DL models for binary predictions. Notably, a pre-trained ResNet50 model was employed to predict NPM1 status. The FRCNN model achieved an impressive precision rate of 0.97 when segmenting cells from bone marrow smear images. Furthermore, the binary classification model generated area under the curve (AUC) values of 0.97 for both receiver operating characteristic (ROC) and precision-recall curves; it also yielded a micro-average accuracy of 0.91 in distinguishing between healthy bone marrow donor samples and acute myeloid Leukemia (AML). Additionally, their DL model demonstrated an elevated precision level of 0.86 in predicting NPM1 mutation status. However, it is essential to acknowledge that employing manual feature extraction may introduce inefficiencies and errors when handling large datasets; this limitation should be addressed in future research endeavors.

In [12], the authors introduced a feature extraction technique that combines DL and image processing methodologies. The proposed approach consists of two key phases: first, identifying regions of interest (ROIs) using a CMYK-moment based localization technique; second, adopting a DL-based feature fusion strategy to extract features from the identified regions. Notably, their method achieved an impressive overall classification accuracy of 97.57%. However, it is important to consider that the pipeline employed for image data preparation may not be universally applicable, potentially leading to reduced accuracy when dealing with new or diverse image datasets.

In a study [13], the author proposed an advanced multiclass WBC differentiation approach utilizing three DL models: ResNet50, ResNext50-32, and ResNext101-32. The proposed method achieved impressive performance metrics, including accuracy of 0.8149, average precision (AP) of 0.7982, area under the curve (AUC) of 0.8293, and F1 score of 0.8073. Moreover, in this investigation, researchers compiled a retrospective dataset consisting of 1,732 bone marrow images containing a total of 27,184 cellular entities; this extensive dataset encompassed both individual cells (24,165 instances) and cell debris samples (2,983 instances), obtained from a cohort comprising 89 pediatric leukemia patients who were treated at the Shanghai Children's Medical Center. To train the model effectively, 70% of the cellular entities were allocated to the training set for learning purposes, while the remaining cells were carefully partitioned among test sets and validation sets following an organized scheme.

In the study [14], the authors constructed an object detection model based on Faster RCNN, incorporating Region Proposal Network (RPN) and Fast R-CNN for candidate target box selection and accurate target classification and regression. Compared to standard approaches, their technique showed a marginal decrease in recall by 4% but achieved significant improvements in precision by 26.4%, F1-score by 12.1%, and Average Precision at IoU=0.5 (AP@50) by 3%. The model reached a recall of 0.710, precision of 0.496, AP@50 of 0.533, and F1-score of 0.575; this pioneering methodology demonstrates its effectiveness in providing a reliable framework for morphological evaluation of cells derived from bone marrow samples.

Furthermore, a study [15] proposes a revolutionary modality for Acute Lymphoblastic Leukemia (ALL) recognition using the C-NMC-2019 dataset with an ensemble strategy that leverages state-of-the-art deep learning techniques combined with transfer learning applications from five pre-trained networks: Xception, VGG-16, InceptionResNet-V2, MobileNetV2, and DenseNet-121. The visual input undergoes convolutional layers with small receptive fields (3×3), followed by integration into a Kappa-based ensemble model to achieve outstanding performance in ALL recognition – yielding impressive results such as a Weighted F-Score (WFS) of 89.72% and Area Under the Curve (AUC) value of 94.8%. These findings highlight the immense potentiality of this approach as both a robust and accurate framework for ALL recognition; it could potentially revolutionize the medical imaging diagnosis domain while enabling timely identification & treatment options, leading to improved patient care quality. In the study [16], the authors addressed the critical challenge of brain tumor classification from Magnetic Resonance Imaging (MRI) images. They proposed a custom, lightweight Convolutional Neural Network (CNN) model based on the modified VGG-19 architecture, which effectively reduced computational complexity while achieving an impressive 96.42% classification accuracy.

Similarly, in the study [17], researchers introduced a novel and innovative feature selection algorithm that synergistically combined the strengths of wrapper, filter, and ensemble methods intending to maximize prediction accuracy in a chronic kidney disease dataset. The proposed approach achieved an astounding 100% accuracy rate – highlighting its immense potential for facilitating medical diagnosis and treatment processes.

Research in acute lymphoblastic leukemia detection and classification has seen significant advancements, particularly with the utilization of DNNs and ML techniques. However, several research gaps persist despite these advancements. Firstly, there is a need for a comprehensive approach and a generalized model that can effectively handle ALL blast cells' varied patterns, shapes, and textures. While existing studies have made strides in this direction, there is still room for improvement in developing a methodology that can be integrated with operational software and accurately identify blast cells in real-time. Additionally, the lack of a universally applicable methodology for image analysis tasks in leukemia diagnosis remains a challenge. Addressing these gaps will contribute to the development of more robust and widely applicable systems for leukemia detection and classification, ultimately enhancing healthcare accessibility and affordability.

3. Proposed Method

We have chosen to utilize publicly available ALL datasets to develop our deep neural network (DNN) models. The schematic workflow of our proposed methodology is presented in Figure 2. Initially, we accessed the dataset via Kaggle API and partitioned it into training, validation, and testing sets with the aim of mitigating potential data leakage and overfitting issues. Subsequently, our pre-processing pipeline was applied solely to the training set while normalizing images from both validation and testing sets. This critical pre-processing phase ensures that disparities between global and local minima-maxima are preserved, thereby reducing risks associated with data leakage and overfitting. Next, we inputted training and validation images into DNN architectures; multiple DNN algorithms were trained using these datasets, followed by their evaluation during the validation stages. Based on achieved validation accuracy scores among various algorithms under consideration, optimal models were selected for further use. In subsequent phases of experimentation, normalized test images underwent rigorous testing using multiple optimized DNN models, which subsequently classified them into four distinct subtypes: Benign cells, Early Pre-B cells, Pre-B cells, and Pro-B

cells - thus facilitating identification across different subtypes of Acute Lymphoblastic Leukemia (ALL).

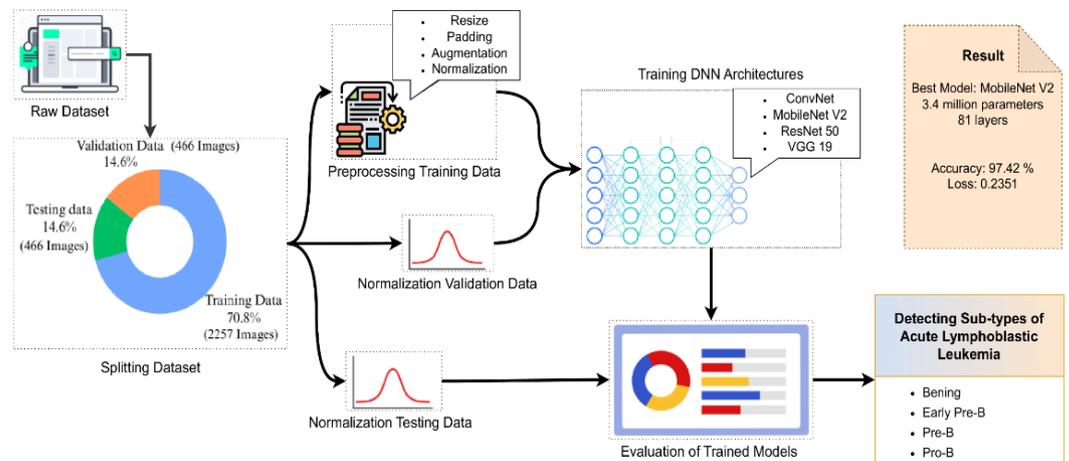


Figure 2. The Proposed Methodology's Workflow.

3.1. Dataset Description

This study made use of the meticulously curated ALL Image dataset, which comprises 3256 peripheral blood smear images collected from 89 patients suspected of Acute Lymphoblastic Leukemia (ALL) at Taleqani Hospital's Bone Marrow Laboratory in Tehran, Iran [18], [19]. The dataset encompasses images sourced from 25 healthy individuals diagnosed with benign conditions (hematogenous) and 64 patients who received definitive diagnoses for various subtypes of ALL: Early Pre-B, Pre-B, and Pro-B. The ALL dataset served as both training material and a key reference point for developing and validating multiple deep neural networks (DNN) models. Visual images were obtained using a Zeiss camera integrated with a microscope equipped with 100x magnification capabilities; specialists employed flow cytometry to accurately identify cell types and subtypes. Figure 3 illustrates examples of Benign (hematogone) Cells alongside early pre-B ALL Cells, pre-B ALL Cells, and pro-B ALL Cells, respectively.

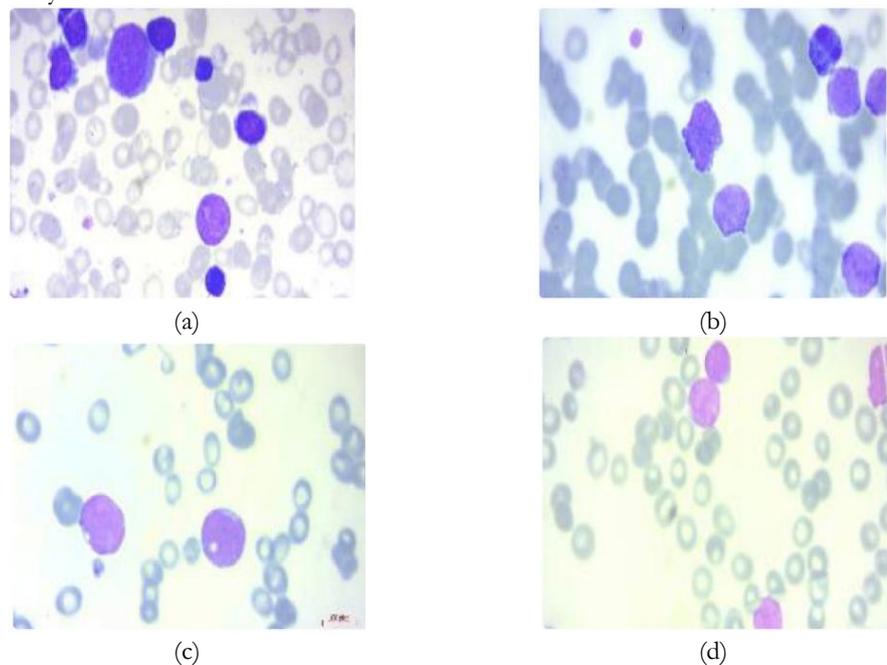


Figure 3. Representative Images of Different Subtypes of Acute Lymphoblastic Leukemia (a) Benign; (b) Early Pre-B; (c) Pre-B; (d) Pro-B.

3.2. Data Preparation Pipeline

The quality and preparation of input data significantly influence the effectiveness of DNN models. Hence, our data pre-processing pipeline incorporates a suite of advanced techniques specifically tailored to enhance microscopic images used for ALL identification. These techniques encompass but are not limited to Data Augmentation, Normalization, and Standardization. Specifically, data augmentation is implemented dynamically during the data loading process. Augmentation parameters, including rotation, zoom, and shear, are applied to the images on-the-fly during training, effectively expanding the diversity of the dataset and enhancing the model's ability to generalize. By strategically applying these pre-processing techniques, we aim to optimize the discriminative capabilities of our DNN-based system for detecting ALL while enhancing overall efficacy and accuracy.

3.3. Deep Neural Network Architecture

In pursuance of the research objectives articulated in this work, we have endeavored to integrate the following cutting-edge DNN algorithms.

- Convolutional Neural Network (ConvNet)
- MobileNetV2
- Residual Neural Network 50 (ResNet50)
- Visual Geometry Group 2019 (VGG19)

3.3.1. Convolutional Neural Network (ConvNet)

We have meticulously designed and optimized a deep and sophisticated Convolutional Neural Network (ConvNet) architecture with unparalleled complexity. The ConvNet model consists of 11 stacked layers, including 4 Convolutional Layers, 3 Max Pooling Layers, 2 Dropout Layers, a single Flatten Layer, and a Dense Layer equipped with SoftMax activation function for multiclass classification tasks. This model boasts an impressive total of 5,638,440 trainable parameters. These parameters facilitate a comprehensive understanding of the underlying data distribution, enabling the generation of high-quality insights into the targeted problem domain through nuanced learning patterns. The strategic configuration of the ConvNet architecture was optimized to strike a balance between computational efficiency and model accuracy, resulting in an exceptionally high-performing model for multiclass classification.

3.3.2. MobileNetV2 Architecture

We have devised an innovative MobileNetV2 architecture consisting of 81-layer depth incorporating elaborate assemblies of computational modules such as 13 Conv2D strata, 24 Batch Normalization-based Convolutional layers, 26 Rectified Linear Unit (ReLU) activation function-based layers, 12 Depth-wise Convolutional strata, 4 Zero-padding layers, a single Flatten layer and final Dense layer comprising four distinct units all utilizing SoftMax activation function. This network architecture boasts a total parameter count amounting to 34.5M, including trainable 32.25M and non-trainable parameters 2.25M. This remarkable degree of flexibility allows it to capture intricate patterns in high-dimensional data spaces, making it highly suitable for various applications such as computer vision or natural language processing requiring sophisticated & scalable ML models.

3.3.3. Residual Neural Networks 50 (ResNet50)

Our ResNet50 architecture comprises hierarchical layers totaling up to 149 arranged systematically across 42 Activation layers, 46 Convolutional layers, 45 Batch-wise Normalization, 16 Add mechanisms, 7 Zero Padding layers, a Flatten layer, and a Dense layer. With extensive parameterization totaling around 23 million parameters, out of which only about 400k are trainable while the remaining 23M are non-trainable, this provides exceptional representational power, enabling complex pattern identification tasks within high-dimensional datasets

3.3.4. Visual Geometry Group 2019 (VGG19)

The VGG19 convolutional neural network showcases impressive representational power across diverse computer vision tasks due largely to its highly-parameterized DL design consisting primarily of 16 distinct convolutional filters, five max pooling operations, a flatten operation stage followed by a dense output unit. Collectively contributing towards enhanced discriminative capabilities, boasting nearly 20 million learnable parameters inclusive of both

trainable & non-trainable components where 100k remain trainable while the remaining 20M stay non-trainable, respectively.

3.4. The Intelligent Telediagnosis Application

We have successfully designed and implemented an innovative system to expedite the detection of ALL subtypes by developing a state-of-the-art real-time telediagnosis web application. The abstract architecture of this Telediagnosis software is visually illustrated in Figure 4.

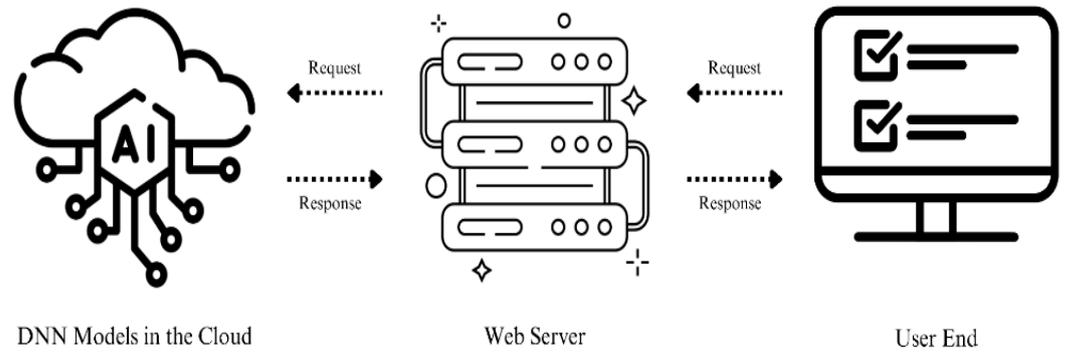


Figure 4. The architecture of the telediagnosis application.

The Web App first collects ALL Images from the patient or clinician, who constitutes the user end. The cloud promptly executes various image processing modules and then discerns which Deep Neural Network (DNN) model would be most optimal for analysis based on these processed images. The selected DNN model subsequently employs its predictive analysis functionalities to determine the subtype of ALL present within each input microscopic image file within an average of 3 seconds. These resultant findings are then effectively transferred back to the user end, where they can be easily accessed and comprehended through intuitive UI elements, as shown in Figure 5. This developed web app promises to significantly accelerate diagnosis across all subtypes of Acute Lymphoblastic Leukemia (ALL), resulting in expedited treatment interventions that ultimately lead to improved patient outcomes. To achieve this goal, we have written our application using Python 3.8 and Flask micro-framework for building web applications while also utilizing Render for API calls and cloud operations. Furthermore, we have employed TensorFlow-cpu alongside Keras & OpenCV libraries, among others, when developing our DL modules. We plan to reduce the computational time further and increase the robustness of the application by scaling our system setup.

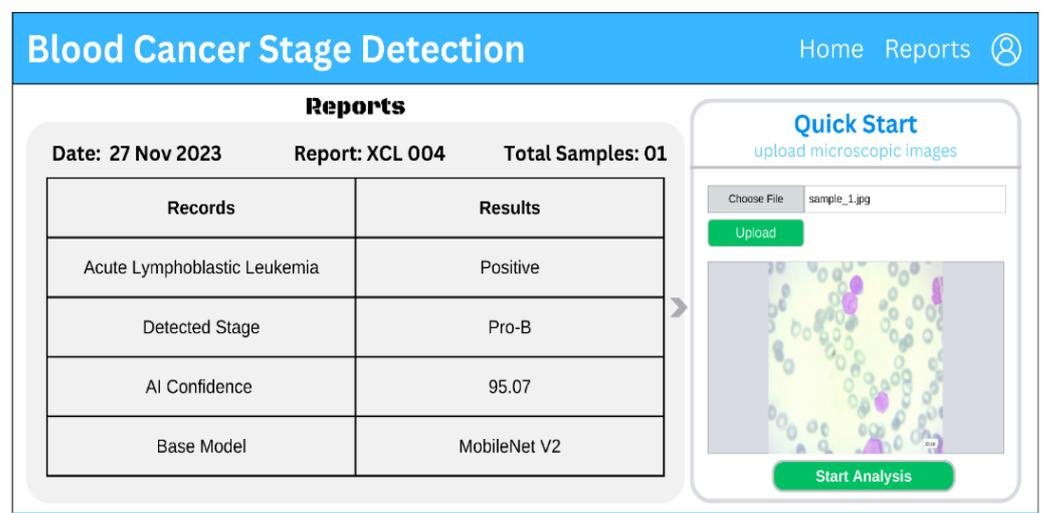


Figure 5. Microscopic image file upload and result page.

4. Results and Discussion

This study aimed to employ multiple DNNs for the identification of Acute Lymphoblastic Leukemia subtypes through the analysis of microscopic blood smear images. Initially, we individually trained four distinct DNN models and recorded their respective validation accuracies. Subsequently, the performance of these models was evaluated using the test dataset with the built-in model evaluator method provided by Keras. The most effective model, as determined by its accuracy in the evaluation stage, was selected for integration into the telediagnosis web application. Furthermore, the selected model underwent rigorous testing again using the test data to assess its classification performance, specifically focusing on metrics such as precision and recall pertinent to the medical domain. Detailed outcomes regarding the performance of the trained models throughout different phases are delineated in subsequent subsections of this study.

4.1. MobileNetV2 Model

Examining the optimized MobileNetV2 model encompasses assessing its performance across various metrics, including training validation accuracy and loss. Remarkable accuracy scores of 0.9312 and 0.9799 have been attained for the corresponding phases. Furthermore, the model's proficiency is underscored by the recorded loss metrics during training and validation, standing at 0.7753 and 0.1792, respectively. In Figure 6, a graphical representation is provided, illustrating the relationship between training accuracy and validation accuracy, as well as training loss and validation loss of the MobileNetV2 model, while showcasing robust generalization capabilities.

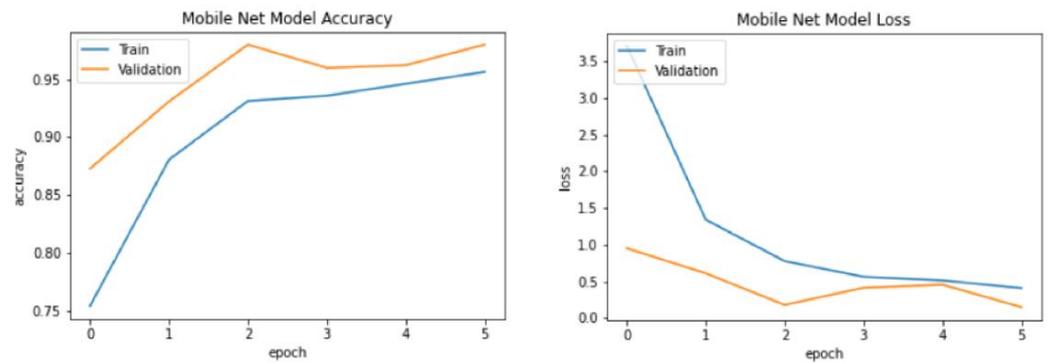


Figure 6. Accuracy and Loss of the MobileNetV2 Model during Training and Validation.

4.2. Convolutional Neural Network (ConvNet) Model

The optimized ConvNet model demonstrates notable accuracy and minimal loss measures across its training and validation stages, as depicted in Figure 7. This graphical representation offers insight into its performance. The observed accuracy rates and low loss values across these phases confirm the model's proficiency in delivering precise and dependable predictions.

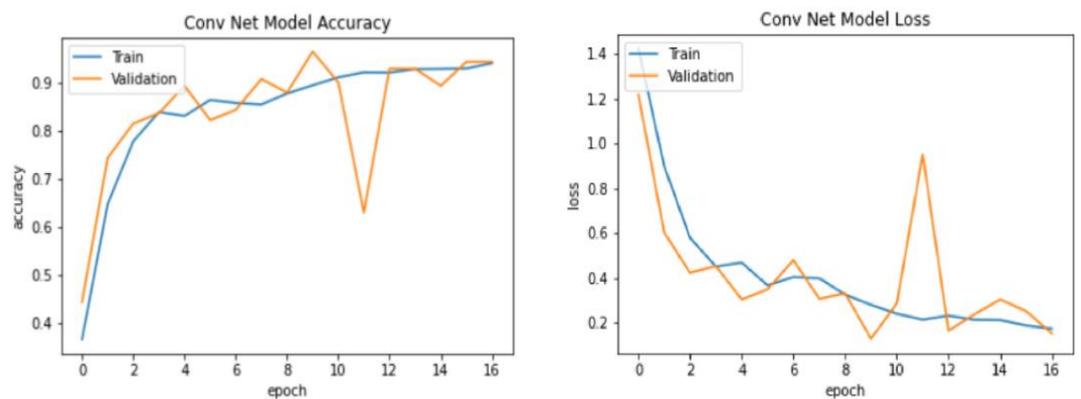


Figure 7. Accuracy and Loss of the ConvNet Model during Training and Validation.

4.3. Residual Neural Network 50 (ResNet50) Model

The assessment of the streamlined ResNet50 model's performance, particularly concerning its accuracy and loss across training and validation datasets, is presented in Figure 8. This figure offers a succinct analysis of both training and validation accuracy, alongside the depiction of training and validation loss. Such visual representations offer valuable insights into the model's performance, aiding comprehension of its behavior nuances and facilitating further optimization of its architecture.

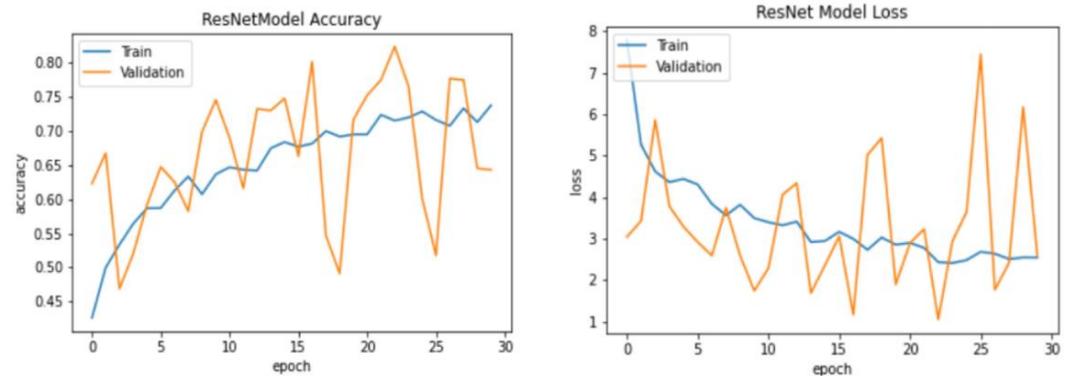


Figure 8. Accuracy and Loss of the ResNet50 Model during Training and Validation.

4.4. Visual Geometry Group 2019 (VGG19) Model

The performance evaluation of the optimized VGG19 model entails thoroughly examining its accuracy and loss metrics throughout the training and validation phases. Furthermore, Figure 9 illustrates the accuracy and loss of the model on both training and validation datasets. These graphical representations highlight the model's efficacy in achieving minimal loss values, thus substantiating its potential as a reliable model for applications in healthcare.

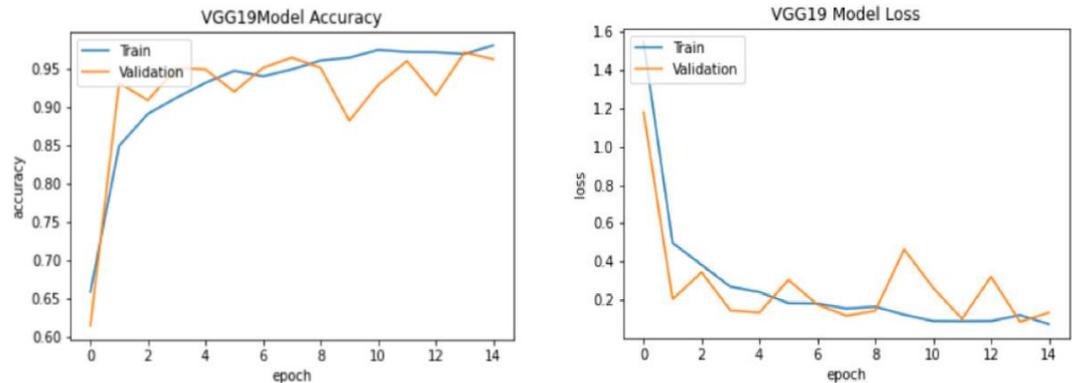


Figure 9. Accuracy and Loss of the VGG19 Model during Training and Validation.

4.5. Evaluation

In the conducted experiment, MobileNetV2, ConvNet, and VGG19 demonstrated the highest performance, achieving accuracies of 97.99%, 96.43%, and 96.43%, respectively. Subsequently, the models were evaluated with test data utilizing Keras's built-in model evaluator, which provided insights into accuracy and loss functions. Notably, MobileNetV2 exhibited the highest accuracy of 97.42% during this evaluation phase, prompting its selection for integration into the tele diagnosis web application. The training, validation, and evaluation scores of the DNN models are summarized in Table 1.

Further assessment of the chosen MobileNetV2 model using test data yielded a test accuracy of 97%, along with macro-average precision, recall, and F1-score of 97%, 96%, and 96%, respectively. The detailed classification report of the best and selected MobileNetV2 model is depicted in Figure 10. Additionally, Figure 11 illustrates the confusion metrics associated with the chosen model.

Table 1. Accuracy and Loss of different DNN models in various states.

Metrics	States	MobileNetV2	ConvNet	VGG19	ResNet50
Accuracy	Training	<u>0.9312</u>	0.8939	0.9488	0.7151
	Validation	0.9799	0.9643	<u>0.9643</u>	0.8237
	Evaluation	0.9742	0.9128	<u>0.9613</u>	0.8526
Loss	Training	0.7753	0.2799	0.1577	2.4317
	Validation	0.1792	0.1282	0.1184	<u>0.1054</u>
	Evaluation	0.2351	<u>0.2309</u>	0.099	0.8412

	precision	recall	f1-score	support
Benign	0.94	0.89	0.92	76
Early	0.95	0.98	0.97	148
Pre	0.99	0.98	0.99	145
Pro	0.98	0.99	0.98	121
accuracy			0.97	490
macro avg	0.97	0.96	0.96	490
weighted avg	0.97	0.97	0.97	490

Figure 10. Classification Report of the Best Performing (MobileNetV2) Model.

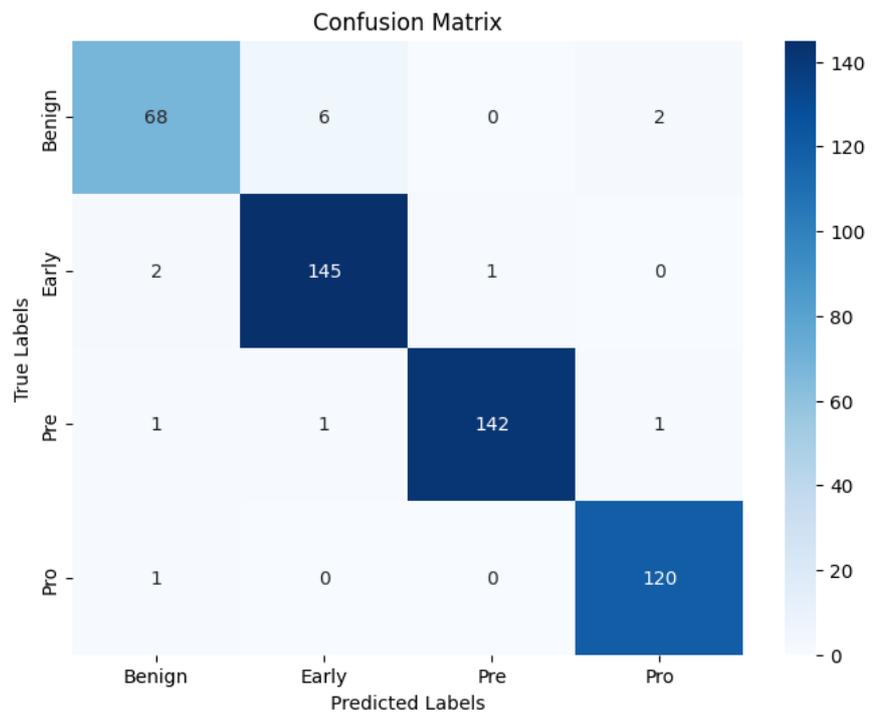


Figure 11. Confusion Matrix of the Best Performing (MobileNetV2) Model.

Based on the results above, MobileNetV2 shows superiority over other models such as ConvNet and VGG19, although VGG19 has better training accuracy, MobileNetV2 has more consistent performance between training and evaluation. Apart from that, looking at the data in the confusion matrix, it appears that the dataset is not balanced, so we need to pay attention to the recall, precision, and f1 measurements. It can be seen that the values of accuracy, recall, precision, and f1 are relatively equal in the unbalanced dataset. In theory, MobileNetV2 also features a lighter and more efficient architectural design, which is important for real-time applications such as teleradiology. MobileNetV2 uses fewer and more efficient convolution operations, enabling faster image processing without requiring large computing resources. This makes it more practical to apply to web-based systems that require fast and accurate

responses. Additionally, MobileNetV2's efficiency in managing test data shows that the model is better at generalizing learning from training data to real situations, an important aspect of medical diagnostic applications.

4.6. Comparison

Table 2 provides a comparative analysis between the methodology employed in this study and recent related works, emphasizing the robustness of MobileNetV2. This comparison underscores the potential of the optimized model to address various practical applications by consistently delivering reliable and accurate predictions.

Table 2. Comparative study on test accuracy with recent related works.

Algorithms	Dataset	Accuracy
You Only Look Once (YOLO) V4[3]	ALL-IDB1	96.06%
Alex Krizhevsky Network (AlexNet)		89.40%
Dense Convolutional Network (DenseNet)121		86.90%
Residual Neural Network (ResNet) 18	ISBI C-NMC 2019	91.70%
Visual Geometry Group (VGG) 16		92.40%
SqueezeNet		93.20%
MobileNet V2 [5]		95.80%
Residual Neural Network (ResNet) 101		81.49%
Residual Neural Network (ResNet) 5	Shanghai Children's Medical Center (SCMC)	79.82%
Residual Neural Network (ResNet) 50		80.73%
Ensembling ResNets [13]		82.93%
GoogleNet	The American Society of Hematology	96.06%
Convolutional Neural Network [20]		94.69%
LeNet [21]	Bone Marrow Laboratory of Taleqani Hospital, Iran	95.75%
Method of this paper	Bone Marrow Laboratory of Taleqani Hospital, Iran	97.00%

The method proposed in this study achieved a high accuracy of 97.00% on the dataset from the Bone Marrow Laboratory of Taleqani Hospital, Iran, one of the highest accuracies among various algorithms and datasets in related studies. This shows the method's effectiveness in overcoming challenges in recognizing Acute Lymphoblastic Leukemia (ALL) from microscopic images. Although tested on different datasets, this high accuracy indicates the adaptability and generalizability of the method, which is very important in medical diagnostic applications. This shows that the approach can take advantage of recent advances in deep learning technology to improve accuracy and reliability in medical diagnosis, especially in classifying leukemia subtypes.

5. Conclusions

The severity of ALL has resulted in a growing number of fatalities, while survivors are left with reduced vitality. Despite the increasing use of AI-based intelligent systems for diagnosing this disease, numerous challenges remain, mainly due to ALL blast cells' varied patterns, shapes, and textures. In light of this, we have designed and optimized multi-DNN models and a tele-diagnostic web app to provide efficient diagnosis assistance. This is a critical step in the fight against ALL to ensure accessible and affordable healthcare. While many researchers have attempted to apply DNNs to detect ALL Leukemia, a comprehensive approach and generalized model are still lacking. Thus, we plan to develop a robust methodology capable of performing image analysis tasks, including identifying regions of interest and approximating blast cell counts, which a broad class will implement in the future.

Author Contributions: Conceptualization: MTH Khan Tusar and MT Islam.; Methodology: MTH Khan Tusar, AH Sakil, and MT Islam.; software: MTH Khan Tusar.; validation: MTH

Khan Tusar and AH Sakil.; formal analysis: MTH Khan Tusar.; investigation: MT Islam.; resources: MTH Khan Tusar.; data curation: MTH Khan Tusar.; writing—original draft preparation: MTH Khan Tusar.; writing—review and editing: MTH Khan Tusar, MNHN Khandaker and MM Hossain.; visualization: MTH Khan Tusar, MNHN Khandaker and MM Hossain.; supervision: MT Islam.

Funding: This research received no external funding.

Data Availability Statement: The publicly available dataset can be found at the following URL: <https://www.kaggle.com/datasets/mehradaria/leukemia>.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] F. Al-Hafiz, S. Al-Megren, and H. Kurdi, “Red blood cell segmentation by thresholding and Canny detector,” *Procedia Comput. Sci.*, vol. 141, pp. 327–334, 2018, doi: 10.1016/j.procs.2018.10.193.
- [2] I. Rahadi, M. Choodoung, and A. Choodoung, “Red blood cells and white blood cells detection by image processing,” *J. Phys. Conf. Ser.*, vol. 1539, no. 1, p. 012025, May 2020, doi: 10.1088/1742-6596/1539/1/012025.
- [3] R. Khandekar, P. Shastry, S. Jaishankar, O. Faust, and N. Sampathila, “Automated blast cell detection for Acute Lymphoblastic Leukemia diagnosis,” *Biomed. Signal Process. Control*, vol. 68, p. 102690, Jul. 2021, doi: 10.1016/j.bspc.2021.102690.
- [4] B. J. Bain, “Diagnosis from the Blood Smear,” *N. Engl. J. Med.*, vol. 353, no. 5, pp. 498–507, Aug. 2005, doi: 10.1056/NEJMra043442.
- [5] S. Ramaneswaran, K. Srinivasan, P. M. D. R. Vincent, and C.-Y. Chang, “Hybrid Inception v3 XGBoost Model for Acute Lymphoblastic Leukemia Classification,” *Comput. Math. Methods Med.*, vol. 2021, pp. 1–10, Jul. 2021, doi: 10.1155/2021/2577375.
- [6] 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, Feb. 2022, doi: 10.3390/s22041629.
- [7] S. Hosseini *et al.*, “Menstrual blood contains immune cells with inflammatory and anti-inflammatory properties,” *J. Obstet. Gynaecol. Res.*, vol. 41, no. 11, pp. 1803–1812, Nov. 2015, doi: 10.1111/jog.12801.
- [8] S. Tahura and M. Hussain, “Treatment Refusal and Abandonment in Pediatric Patients with Acute Lymphoblastic Leukemia in Bangladesh,” *Int. J. Sci. Res. (IJ)*, vol. 6, no. 8, pp. 643–645, 2017, doi: 10.21275/3071703.
- [9] L. Bigorra, A. Merino, S. Alférez, and J. Rodellar, “Feature Analysis and Automatic Identification of Leukemic Lineage Blast Cells and Reactive Lymphoid Cells from Peripheral Blood Cell Images,” *J. Clin. Lab. Anal.*, vol. 31, no. 2, p. e22024, Mar. 2017, doi: 10.1002/jcla.22024.
- [10] P. E. J. van der Meijden and J. W. M. Heemskerk, “Platelet biology and functions: new concepts and clinical perspectives,” *Nat. Rev. Cardiol.*, vol. 16, no. 3, pp. 166–179, Mar. 2019, doi: 10.1038/s41569-018-0110-0.
- [11] J.-N. Eckardt *et al.*, “Deep learning detects acute myeloid leukemia and predicts NPM1 mutation status from bone marrow smears,” *Leukemia*, vol. 36, no. 1, pp. 111–118, Jan. 2022, doi: 10.1038/s41375-021-01408-w.
- [12] T. A. M. Elhassan, M. S. M. Rahim, T. T. Swee, S. Z. M. Hashim, and M. Aljurf, “Feature Extraction of White Blood Cells Using CMYK-Moment Localization and Deep Learning in Acute Myeloid Leukemia Blood Smear Microscopic Images,” *IEEE Access*, vol. 10, pp. 16577–16591, 2022, doi: 10.1109/ACCESS.2022.3149637.
- [13] M. Zhou *et al.*, “Development and Evaluation of a Leukemia Diagnosis System Using Deep Learning in Real Clinical Scenarios,” *Front. Pediatr.*, vol. 9, Jun. 2021, doi: 10.3389/fped.2021.693676.
- [14] J. Liu *et al.*, “A deep learning method and device for bone marrow imaging cell detection,” *Ann. Transl. Med.*, vol. 10, no. 4, pp. 208–208, Feb. 2022, doi: 10.21037/atm-22-486.
- [15] C. Mondal *et al.*, “Ensemble of Convolutional Neural Networks to diagnose Acute Lymphoblastic Leukemia from microscopic images,” *Informatics Med. Unlocked*, vol. 27, p. 100794, 2021, doi: 10.1016/j.imu.2021.100794.
- [16] J. Hossain, M. T. Islam, and M. T. H. K. Tusar, “Streamlining Brain Tumor Classification with Custom Transfer Learning in MRI Images,” in *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, May 2023, pp. 522–526. doi: 10.1109/SIST58284.2023.10223507.
- [17] M. T. Haque Khan Tusar, M. T. Islam, and F. I. Raju, “Detecting Chronic Kidney Disease (CKD) at the Initial Stage: A Novel Hybrid Feature-selection Method and Robust Data Preparation Pipeline for Different ML Techniques,” in *2022 5th International Conference on Computing and Informatics (ICCI)*, Mar. 2022, pp. 400–407. doi: 10.1109/ICCI54321.2022.9756094.
- [18] M. Aria, M. Ghaderzadeh, D. Bashash, H. Abolghasemi, F. Asadi, and A. Hosseini, “Acute Lymphoblastic Leukemia (ALL) image dataset.” Kaggle, 2021. doi: 10.34740/kaggle/dsv/2175623.
- [19] M. Ghaderzadeh, M. Aria, A. Hosseini, F. Asadi, D. Bashash, and H. Abolghasemi, “A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images,” *Int. J. Intell. Syst.*, vol. 37, no. 8, pp. 5113–5133, Aug. 2022, doi: 10.1002/int.22753.
- [20] M. O. Aftab, M. Javed Awan, S. Khalid, R. Javed, and H. Shabir, “Executing Spark BigDL for Leukemia Detection from Microscopic Images using Transfer Learning,” in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, Apr. 2021, pp. 216–220. doi: 10.1109/CAIDA51941.2021.9425264.
- [21] B. S. A. A. K. Datchanamorthy, H. M. A. A. S., and Kavya Varshini M, “Acute Lymphoblastic Leukemia Detection Using Sequential, Lenet and Vggnet Models,” *Pakistan Hear. J.*, vol. 56, no. 1, 2023, [Online]. Available: <https://www.pkheartjournal.com/index.php/journal/article/view/1195>