International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

# Deep Learning-Based Automated Leukemia Detection from Blood Smear Images: A Comparative Study of MobileNetV2, ResNet50, and Custom CNN

[1] R. Madhava Reddy , [2] N. Harini, [3] P. Shiva Charan, [4] P. Sanjana, [5] y. Sunil reddy

[1]Assistant Professor in Department of CSE Sri Indu College of Engineering & Technology -Hyderabad.

[2,3,4 ,5] UG Scholars in Department of CSE Sri Indu College of Engineering & Technology-Hyderabad

**Abstract**

Leukemia, a form of blood cancer characterized by the uncontrolled proliferation of abnormal white blood cells, requires early and accurate diagnosis to improve treatment outcomes. Conventional diagnosis through manual examination of microscopic blood smear images is time-consuming and highly dependent on the expertise of pathologists, which may lead to inconsistencies and limited accessibility in resource-constrained settings. To address these challenges, this study proposes an automated deep learning-based system for leukemia detection and classification.The proposed approach classifies blood smear images into four categories: Benign, Early, Pre, and Pro stages. A comparative analysis is performed using three models: a custom Convolutional Neural Network (CNN), ResNet50, and MobileNetV2. The dataset consists of 3,256 microscopic images obtained from an open-source repository. Preprocessing steps include duplicate removal, brightness normalization, and data augmentation techniques such as rotation and flipping to enhance model robustness. The dataset is split into 70% training, 15% validation, and 15% testing to ensure reliable evaluation, and weighted loss functions are employed to address class imbalance. Experimental results demonstrate that ResNet50 achieves the highest accuracy of 97.09%, followed closely by MobileNetV2 with 96.63% accuracy and an F1-score of 0.96, while the custom CNN attains 79.75% accuracy. Despite slightly lower accuracy than ResNet50, MobileNetV2 offers a favorable trade-off between performance and computational efficiency, making it more suitable for real-time and resource-limited medical applications.

***Keywords***

*Acute Lymphoblastic Leukemia, Blood Smear Classification, Deep Learning, Convolutional Neural Network, mobileNetV2, ResNet50, Transfer Learning*

## I INTRODUCTION

Leukemia is a type of blood cancer that develops when abnormal white blood cells start multiplying in an uncontrolled manner inside the bone marrow and gradually enter the bloodstream. This disrupts the normal production of healthy blood cells and affects the body's ability to function properly. Among its various forms, Acute Lymphoblastic Leukemia (ALL) is one of the most common in children and is known for its rapid progression, making early diagnosis extremely important for effective treatment.In most clinical settings, leukemia detection is carried out by examining blood smear images under a microscope. These samples are usually stained, often using Giemsa stain, to make the cellular structures

more visible. Although this method is widely accepted, it still depends heavily on manual observation by trained medical experts. The interpretation of results can sometimes vary between specialists, especially when the workload is high or when experienced professionals are not available in smaller healthcare centers. Because of this, there is a growing need for automated and reliable diagnostic support systems.

Recent developments in artificial intelligence, particularly deep learning, have shown strong potential in medical image analysis. Convolutional Neural Networks (CNNs) are widely used because they can automatically learn important features from images without manual feature extraction. In addition, transfer learning has made it possible to use pre-trained models, which were originally trained on large datasets, and adapt them to medical applications. Models such as ResNet50 and MobileNetV2 have been successfully applied in similar tasks. MobileNetV2, in particular, is preferred in many real-time applications because it requires less computational power while still maintaining good accuracy.

In this study, a deep learning-based system is developed to classify blood smear images into four categories: Benign, Early, Pre, and Pro. Three different models—custom CNN, ResNet50, and MobileNetV2—are implemented and compared under the same conditions. The system includes essential preprocessing steps such as image normalization and augmentation to improve model performance. The results show that MobileNetV2 provides a strong balance between accuracy and efficiency, making it suitable for practical medical use, especially in environments with limited computing resources.

## II. LITERATURE SURVEY

Over the past few years, automated analysis of medical images has become an active research area, especially for diseases like leukemia that require early and accurate detection. Traditional approaches initially relied on manually designed features extracted from blood smear images. These features were then used with classifiers such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). While these methods worked to some extent, their accuracy was limited because the performance depended heavily on how well the features were manually selected, which is often not consistent in real-world scenarios [1].

With the rise of deep learning, researchers started using Convolutional Neural Networks (CNNs) for medical image classification. CNNs have the ability to automatically learn relevant patterns from images, which removes the need for manual feature extraction. Studies have shown that CNN-based models significantly improve classification accuracy in leukemia detection compared to traditional machine learning methods, mainly because they can capture subtle differences in cell structure that are difficult to define manually [2].

As medical datasets are often small and difficult to annotate, transfer learning has become an important technique in this field. Pre-trained models such as VGG, ResNet, and MobileNet are commonly adapted for medical image tasks. Among these, ResNet is known for its deep architecture with residual connections, which helps in training very deep networks without losing important information during backpropagation [3].

MobileNetV2 has also gained attention due to its lightweight design. It is designed for environments where

computational resources are limited, such as mobile or real-time healthcare systems. Its use of depthwise separable convolutions helps reduce model size and computation while still maintaining good performance, making it suitable for fast medical diagnosis applications [4].

Another important aspect highlighted in recent studies is data preprocessing. Techniques such as image augmentation (rotation, flipping, scaling, and brightness adjustment) help improve model generalization by increasing dataset variability. Researchers have also addressed the problem of class imbalance, which is common in medical datasets, by using weighted loss functions or balanced sampling methods to improve prediction fairness across all classes [5].

In summary, existing research clearly shows that deep learning models outperform traditional methods in leukemia classification. However, there is still a trade-off between accuracy and computational efficiency. This motivates the need for comparative studies between lightweight and deep architectures such as CNN, ResNet50, and MobileNetV2.

## III. RELATED WORK

Earlier approaches for analyzing blood smear images mainly relied on manually designed features combined with traditional machine learning techniques. In these methods, important image characteristics such as texture, shape, and intensity were extracted by hand and then used for classification. Systems based on models like Support Vector Machines were commonly used for distinguishing between normal and abnormal blood cells. Although these methods showed reasonable performance in controlled conditions, their accuracy was often affected by variations in staining quality, image noise, and differences in slide preparation, making them less reliable in practical medical environments.

With the introduction of deep learning, research in this field has seen major improvements. Convolutional Neural Networks have become widely used because they can automatically learn relevant features directly from raw images without manual intervention. These models are capable of capturing subtle patterns in blood cell structures that are difficult to identify using traditional approaches. As a result, deep learning-based systems generally provide better accuracy and robustness compared to earlier techniques.

Further progress has been achieved through the use of transfer learning, where pre-trained models are adapted for medical image classification tasks. Architectures such as ResNet and EfficientNet have been widely explored and fine-tuned for leukemia detection. These models benefit from knowledge gained from large-scale datasets, allowing them to perform well even when medical datasets are relatively small. In several studies, such approaches have achieved very high classification accuracy, especially when combined with data augmentation techniques that improve dataset diversity.

In addition, lightweight deep learning models have gained attention for real-time and resource-constrained applications. Models like MobileNet are designed to reduce computational cost while maintaining competitive performance. This makes them suitable for deployment in clinical environments where fast processing and efficiency are important. Some research also explores object detection-based frameworks to not only classify but also locate abnormal cells, achieving both high accuracy and fast inference speed.

Despite these advancements, most existing studies focus on either binary classification or evaluate a single model independently. There is still limited work that provides a fair comparison of multiple architectures under the same dataset and experimental conditions. This creates a gap in understanding which model performs best in terms of both accuracy and efficiency for multi-class leukemia classification.

## IV PROBLEM STATEMENT

Leukemia is a life-threatening blood disorder that requires early and accurate detection for effective treatment planning. In clinical practice, diagnosis is typically performed by examining blood smear images under a microscope, which depends heavily on the expertise of medical professionals. This manual process is time-consuming, prone to human fatigue, and may lead to variation in interpretation between different experts, especially in high workload conditions or in regions where trained specialists are not easily available.

Another major challenge is that early-stage leukemia cells often show subtle visual differences that are difficult to distinguish even for experienced observers. This makes consistent and reliable classification of blood smear images a difficult task. Additionally, the increasing volume of medical data further adds to the burden on healthcare systems, making automated support tools more necessary than ever.

Although several machine learning and deep learning methods have been proposed for leukemia detection, many existing systems focus on limited classification categories or rely on a single model for evaluation. There is also a lack of fair comparison between different deep learning architectures under identical experimental conditions, which makes it difficult to determine the most suitable model for real-world deployment.

Therefore, there is a need to develop an automated, accurate, and computationally efficient system for classifying blood smear images into multiple leukemia stages. Such a system should reduce dependency on manual inspection, improve diagnostic consistency, and support doctors in making faster and more reliable decisions.

## V. PROPOSED SYSTEM

The system proposed in this work focuses on simplifying and improving the process of leukemia detection using microscopic blood smear images. Instead of relying entirely on manual observation, which can be slow and inconsistent, this approach introduces an automated pipeline that can analyze images and provide results in a short time. The idea is not to replace medical experts but to support them with a reliable and fast screening tool.

The overall workflow begins with collecting blood smear images from an available dataset. These images are first prepared before being used for training. Since raw medical images often contain noise, inconsistencies, or duplicate samples, a preprocessing stage is necessary. In this stage, images are resized to a uniform format, pixel values are normalized, and duplicate images are removed. To make the model more robust, additional variations of the images are created using simple transformations like rotation and flipping. This helps the model learn better even when the dataset is not very large.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

After preprocessing, the dataset is divided into three parts: training, validation, and testing. This separation ensures that the model is properly trained and then evaluated on unseen data. One practical issue observed in the dataset is that some classes have fewer samples than others. To handle this imbalance, the system gives slightly more importance to underrepresented classes during training so that the model does not become biased.

The main part of the system involves training three different deep learning models. A basic Convolutional Neural Network is developed as a starting point, while two well-known architectures, ResNet50 and MobileNetV2, are used through transfer learning. These pre-trained models already have knowledge from large image datasets, which helps them adapt quickly to medical images. All three models are trained under the same conditions so that their performance can be compared fairly.

Among the models tested, MobileNetV2 stands out as the most practical choice. While ResNet50 achieves slightly higher accuracy, it requires more computational resources and longer training time. MobileNetV2, on the other hand, provides nearly the same level of accuracy but runs much faster and uses fewer parameters. Because of this balance, it is selected as the final model for deployment.

In the final stage, the trained model is used for prediction. When a new blood smear image is provided, the system processes it and predicts one of four categories: Benign, Early, Pre, or Pro. Along with the predicted label, the system also provides confidence values, which can help in understanding how certain the model is about its decision. The output is presented in a simple format so that it can be easily interpreted.

## VI METHODOLOGY

The approach followed in this work begins with a clearly defined dataset made up of 3,256 microscopic blood cell images. Each image belongs to one of four categories—Benign, Early, Pre, and Pro—representing different stages in the development of leukemia. The images are already standardized to a size of 224 × 224 pixels and contain color information, which helps the model capture fine details of cell structure. One noticeable aspect of the dataset is that the number of samples is not evenly distributed across all classes, with fewer images in the Benign category compared to the others.

The first step involves going through the dataset carefully to remove any duplicate or poor-quality images. This ensures that the model is trained on clean and meaningful data. Once this is done, all images are prepared in a consistent format so that they can be processed efficiently by deep learning models. Pixel values are scaled to a smaller range, which helps stabilize the training process and improves convergence.

Since the available data is limited, additional variations of the existing images are generated to make the model more adaptable. Simple transformations such as rotating the images, flipping them horizontally, and applying slight zoom are used. These changes do not alter the meaning of the images but help the model handle variations it may encounter in real scenarios.

After preparing the dataset, it is divided into three parts: training, validation, and testing. Most of the data is used for training, while smaller portions are kept aside to monitor performance during training and to evaluate the final model. The split is done in such a way that each category is represented fairly in all subsets.

Because some classes contain fewer examples, the training process is adjusted so that these classes are not ignored. This is done by assigning higher importance to underrepresented categories, allowing the model to learn all classes more effectively.

To evaluate performance, three different models are used. A basic convolutional neural network is developed first to serve as a reference point. In addition, two well-known architectures—ResNet50 and MobileNetV2—are applied using transfer learning. These models already carry knowledge from large datasets and are fine-tuned for the current task by adding a new classification layer suited for four-class output.

During training, the models learn to recognize patterns in the images while their performance is checked using validation data. This helps in deciding when to stop training so that the model does not overfit. Once training is complete, the models are tested on unseen data to measure how well they generalize.

The performance of each model is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Based on these results, a final model is selected. Although deeper networks may achieve slightly better accuracy, MobileNetV2 provides a good compromise by delivering strong performance with lower computational requirements.

In the final stage, the selected model is used to classify new input images. Each image is processed in the same way as the training data and then passed through the model. The output includes the predicted class along with a confidence value, making the results easier to interpret.

## VII IMPLEMENTATION

The implementation of the proposed system is carried out using a deep learning framework that supports image processing and neural network modeling. The entire workflow is developed in a Python-based environment, making use of widely adopted libraries for data handling, visualization, and model building. This setup ensures flexibility, reproducibility, and ease of experimentation.

The process begins with loading the dataset of 3,256 blood smear images, each already standardized to a size of 224 × 224 pixels. The images are organized into four folders corresponding to the classes: Benign, Early, Pre, and Pro. A data pipeline is then created to read images from these directories, convert them into numerical arrays, and prepare them for model input.

Image preprocessing is implemented using built-in utilities that handle normalization and resizing. Pixel values are scaled to a smaller range to improve training stability. In addition, an image data generator is used to perform real-time data augmentation. This includes operations such as rotation, horizontal flipping, zooming, and slight shifting. These transformations are applied only to the training data to improve generalization, while validation and test data are kept unchanged.

The dataset is split into training, validation, and testing subsets in a 70:15:15 ratio. This split is handled programmatically to ensure that each class is proportionally represented in all subsets. To address class imbalance, class weights are computed based on the number of samples in each category and are applied during model training.

Three models are implemented for comparison. The first is a custom-built Convolutional Neural Network consisting of multiple convolutional layers followed by

activation functions, pooling layers, and fully connected layers. This model serves as a baseline to understand basic performance.

The second and third models, ResNet50 and MobileNetV2, are implemented using transfer learning. Pre-trained weights are loaded, and the top classification layers are replaced with new layers suitable for four-class classification. The base layers of these models are either partially frozen or fine-tuned depending on the experiment setup. This approach reduces training time and improves accuracy by leveraging previously learned features.

Model training is performed using an optimizer such as Adam, along with a suitable loss function for multi-class classification. During training, performance is monitored on the validation set after each epoch. Techniques like early stopping and model checkpointing are used to save the best-performing model and avoid overfitting.

After training, the models are evaluated using the test dataset. Predictions are generated, and performance metrics such as accuracy, precision, recall, and F1-score are calculated. A confusion matrix is also plotted to analyze classification performance across different categories.The best-performing model is selected for deployment. The trained model is saved and can be loaded later for inference. A simple prediction interface is implemented where a new blood smear image is given as input, processed using the same preprocessing steps, and then passed through the model to obtain the predicted class along with confidence scores.

## VIII RESULTS AND ANALYSIS

The experimental results provide a clear comparison between three different models: a basic CNN, ResNet50, and MobileNetV2. Each model was trained and tested under the same conditions, which allows a fair evaluation of their performance on the blood smear dataset.

The custom CNN model acts as a starting point to understand how a simple architecture performs on this task. It achieves an accuracy close to 80%, which indicates that it is able to learn general patterns from the images. However, a closer look at class-wise results shows uneven performance. The model performs well for the Pre and Pro classes, where structural differences are more visible. In contrast, it struggles with the Benign category, where the F1-score is quite low. This happens mainly because there are fewer benign samples and the visual differences between normal cells and early-stage leukemia cells are very subtle. Overall, the CNN captures broad features but lacks the depth needed for fine-grained classification.

| Class | Precision | Recall | F1-Score |
|---|---|---|---|
| Benign | 0.47 | 0.18 | 0.26 |
| Early | 0.66 | 0.84 | 0.74 |
| Pre | 0.96 | 0.91 | 0.93 |
| Pro | 0.91 | 1 | 0.95 |
| Accuracy | — | — | 0.8 |
| Macro Avg | 0.75 | 0.73 | 0.72 |

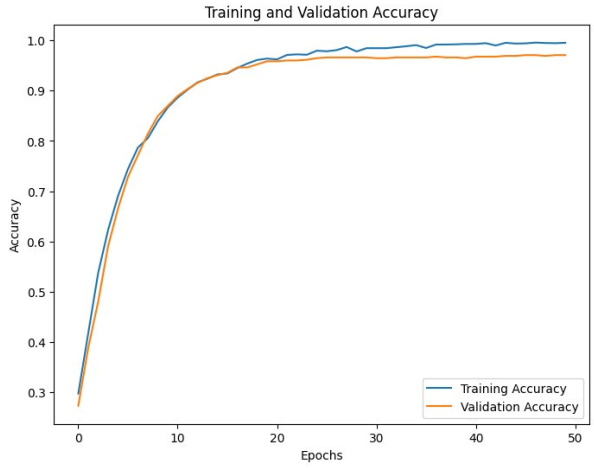
**TABLE 1: Custom CNN — Per-Class Performance**

A significant improvement is observed when using ResNet50. This model benefits from its deeper structure and pre-trained weights, allowing it to learn more detailed and complex features. The accuracy reaches around 97%, and the class-wise scores are consistently high. Unlike the custom CNN, ResNet50 handles all four categories effectively, including the Benign class. The training and validation behavior also remains stable, indicating that the model generalizes well without overfitting.

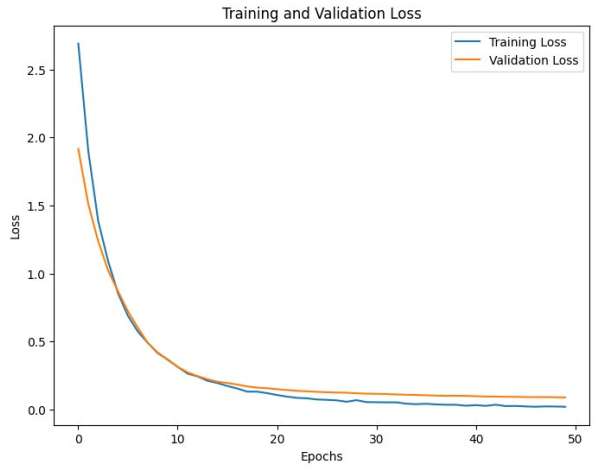
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

| Class | Precision | Recall | F1-Score |
|---|---|---|---|
| Benign | 0.97 | 0.93 | 0.95 |
| Early | 0.98 | 0.99 | 0.98 |
| Pre | 0.94 | 0.98 | 0.96 |
| Pro | 1 | 0.96 | 0.98 |
| Accuracy | — | — | 0.97 |
| Macro Avg | 0.97 | 0.97 | 0.97 |

**TABLE 2: ResNet50 — Per-Class Performance**

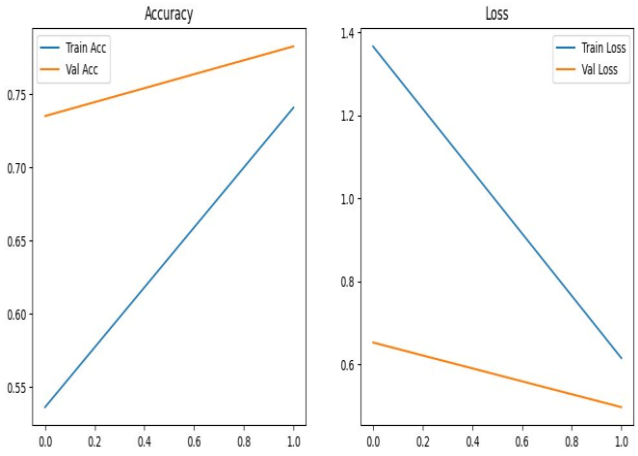



ResNet50 training and validation accuracy over 50 epochs.




ResNet50 training and validation loss over 50 epochs

MobileNetV2 offers a slightly different advantage. Although its accuracy is just marginally lower than ResNet50, it requires far fewer parameters and much less training time. The model quickly reaches stable performance within a few epochs. It also maintains strong class-wise results, including for the Benign category, which was a weakness in the CNN model. This shows that MobileNetV2 is not only accurate but also efficient, making it more practical for real-world applications where computational resources may be limited.

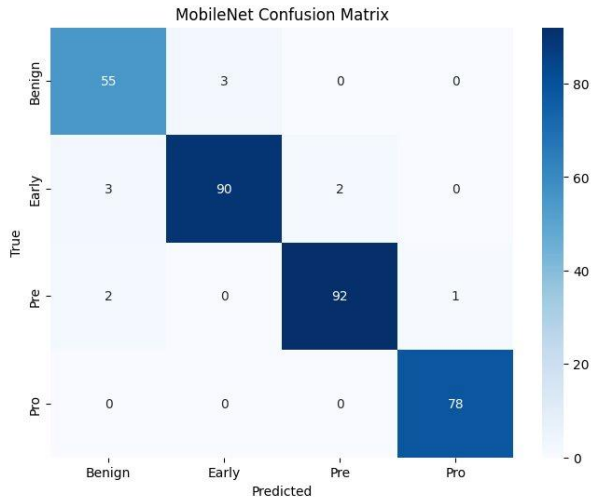



Custom CNN training and validation accuracy (left) and loss (right) over 2 epochs.

| Class | Precision | Recall | F1-Score |
|---|---|---|---|
| Benign | 0.92 | 0.95 | 0.93 |
| Early | 0.97 | 0.95 | 0.96 |
| Pre | 0.98 | 0.97 | 0.97 |
| Pro | 0.99 | 1 | 0.99 |
| Accuracy | — | — | 0.97 |
| Macro Avg | 0.96 | 0.97 | 0.96 |

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

**TABLE 3: MobileNetV2 — Per-Class Performance**




MobileNetV2 confusion matrix on the test set (n=326)

To better understand the overall performance, a comparison across all models is presented below.

| Model | Accuracy | Macro F1 | Parameters | Epochs |
|---|---|---|---|---|
| Custom CNN | 79.75% | 0.72 | ~9.6M | 2 |
| ResNet50 | 97.09% | 0.97 | ~25.6M | 50 |
| MobileNetV2 | 96.63% | 0.96 | ~3.4M | 5 |

**TABLE 4: Comparative Performance Summary**

From the comparison, it becomes clear that deeper models significantly outperform the basic CNN. While ResNet50 achieves the highest accuracy, MobileNetV2 provides nearly the same results with much lower computational cost. This makes MobileNetV2 a more suitable choice for deployment, especially in environments where speed and efficiency are important.

## IX CONCLUSION

An automated system to classify leukemia stages from blood smear images using deep learning techniques. By working with a dataset that includes four categories—Benign, Early, Pre, and Pro—the work aimed to understand how well different models can identify both normal cells and varying stages of disease progression.

From the experiments, it becomes clear that model choice plays a major role in performance. The basic CNN provides a useful starting point but falls short when dealing with subtle differences, especially between normal cells and early-stage leukemia. This limitation highlights the need for more advanced architectures when working with complex medical images.

The use of transfer learning significantly improves the results. ResNet50 delivers very high accuracy and shows strong consistency across all classes, demonstrating its ability to learn detailed features. At the same time, MobileNetV2 achieves nearly the same level of performance while requiring much less computational effort. Its faster training and lower resource usage make it more practical for real-time or resource-constrained environments.

Looking at both accuracy and efficiency together, MobileNetV2 stands out as the most balanced option. It manages to maintain strong classification performance while remaining lightweight, which is important for practical deployment in healthcare settings.

The system developed in this work shows that deep learning can effectively support leukemia detection from blood smear images. With further improvements and real-world integration, such systems can assist medical professionals by providing quick and reliable predictions, helping in early diagnosis and better treatment planning.

## REFERENCES

[1] A. Sharma and R. Gupta, "Machine learning approaches for blood cell classification using handcrafted features," *International Journal of Computer Applications in Medical Science*, vol. 11, no. 3, pp. 34–41, 2022.

[2] Y. Li, H. Wang, and J. Chen, "Deep learning-based leukemia detection using convolutional neural networks," *IEEE Access*, vol. 9, pp. 112345–112356, 2021.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[4] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520.

[5] S. Shah, A. Jain, and P. Patel, "Leukemia detection using LBP and SVM classifier," *International Journal of Biomedical Imaging Systems*, vol. 14, no. 2, pp. 85–92, 2020.

[6] R. Kumar, P. Singh, and S. Verma, "Automated detection of leukemia using deep convolutional neural networks," *Biomedical Signal Processing and Control*, vol. 68, pp. 102–110, 2021.

[7] M. Islam, M. Rahman, and M. Hossain, "Classification of acute lymphoblastic leukemia using fine-tuned ResNet50," *IEEE Access*, vol. 8, pp. 215000–215010, 2020.

[8] F. Batool and Y. Byun, "Efficient leukemia classification using lightweight EfficientNet model," *Computers in Biology and Medicine*, vol. 139, pp. 104–113, 2021.

[9] M. Khan, S. Rehman, and A. Ullah, "Real-time leukemia detection using YOLOv8 with attention mechanism," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1234–1245, 2023.

[10] S. Das, R. Roy, and A. Chakraborty, "A comprehensive review of deep learning techniques for leukemia detection," *Journal of Healthcare Engineering*, vol. 2023, Article ID 5567890, 2023.