INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
https://www.doi.org/10.59256/ijrtmr.20250505003
September-October 2025, Vol 5(05), 13-19.
ISSN No: 2583-0368

Leukemia Detection Using Deep Learning: A DenseNet201-Based Approach

Mohammed Ahmed Ali[1], Dr. Mohd Rafi Ahmed[2]
[1]Student, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.
[2]Associate Professor, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.

OPEN ACCESS
Article Citation:
Mohammed Ahmed Ali[1], Dr. Mohd Rafi Ahmed[2], "Leukemia Detection Using Deep Learning: A DenseNet201-Based Approach", International Journal of Recent Trends in Multidisciplinary Research, September-October 2025, Vol 5(05), 13-19.

Abstract: Leukemia is one of the most prevalent forms of blood cancer, primarily affecting white blood cells and posing critical diagnostic challenges. Conventional diagnostic approaches rely on manual microscopic examination, which is subjective, labor-intensive, and prone to errors. This study introduces a deep learning–based system for automated leukemia detection using microscopic images of blood smears. A pre-trained DenseNet201 model is applied through transfer learning, supported by preprocessing techniques such as normalization and augmentation to improve accuracy and reduce overfitting. The model is trained and validated on benchmark datasets, and its performance is measured using metrics including accuracy, precision, recall, and F1-score. Results demonstrate that the proposed system achieves high reliability, significantly reduces diagnostic latency, and provides a scalable framework suitable for integration into clinical workflows and mobile healthcare applications.

Keywords: Leukemia Detection; Deep Learning; DenseNet201; Convolutional Neural Network; Blood Smear Images; Medical Image Analysis; Transfer Learning; AI in Healthcare.

1. Introduction

Leukemia, a malignant cancer of blood-forming tissues, remains a major health concern worldwide, particularly among children and elderly patients. Its accurate diagnosis is crucial to ensure timely treatment and improve survival rates. Current diagnostic practices rely heavily on pathologists' manual microscopic examination of blood smears, which is time-consuming, subjective, and vulnerable to inconsistencies across different experts. Moreover, in resource-constrained regions, lack of access to trained specialists often leads to delayed diagnosis.

With the rapid advancements in artificial intelligence (AI) and deep learning (DL), medical imaging analysis has experienced a paradigm shift. Convolutional Neural Networks (CNNs), in particular, have demonstrated exceptional performance in image recognition and classification tasks. By leveraging transfer learning with pre-trained architectures such as DenseNet201, these models can extract highly discriminative features from blood smear images, enabling reliable classification of leukemia-positive and negative cases.

This project aims to design and implement an automated deep learning system that can assist healthcare professionals in early leukemia detection. The system is designed to be scalable, lightweight, and adaptable for both clinical workflows and mobile health applications.

2. Material And Methods

The development of an automated leukemia detection framework using deep learning required a carefully structured methodology that ensured robustness, reproducibility, and clinical applicability. This section outlines the design of the study, the dataset acquisition process, preprocessing methods, exploratory data analysis (EDA), model development, evaluation techniques, and deployment strategy. Each stage was conducted in a sequential yet iterative fashion, allowing insights gained from one phase to refine subsequent phases.

Study Design

The development of an automated leukemia detection framework using deep learning required a carefully structured methodology that ensured robustness, reproducibility, and clinical applicability. This section outlines the design of the study,

the dataset acquisition process, preprocessing methods, exploratory data analysis (EDA), model development, evaluation techniques, and deployment strategy. Each stage was conducted in a sequential yet iterative fashion, allowing insights gained from one phase to refine subsequent phases.

Data Acquisition

The study was structured as a data-driven predictive modeling framework aimed at the classification of leukemia-positive and leukemia-negative blood smear images. Unlike traditional diagnostic approaches that depend heavily on manual expertise, this framework leverages deep learning to provide automated and consistent predictions. The research followed a standard supervised learning pipeline:

1. **Dataset Collection** – Gathering publicly available and benchmark blood smear datasets.
2. **Data Preprocessing** – Cleaning, normalization, augmentation, and transformation of raw images.
3. **Exploratory Data Analysis** – Identifying patterns, distributions, and variations within the dataset.
4. **Model Development** – Employing DenseNet201 as the backbone CNN with transfer learning.
5. **Model Training** – Optimizing hyperparameters using training and validation subsets.
6. **Evaluation** – Measuring model performance using multiple statistical metrics.
7. **System Deployment** – Developing a user-friendly interface for clinical use.

This sequential methodology ensured that each stage was thoroughly validated before proceeding to the next, making the framework both scientifically rigorous and practically deployable.

Data Preprocessing

The reliability of a deep learning model heavily depends on the quality and representativeness of the data. For this study, datasets were sourced from publicly available repositories such as:

- **ALL-IDB (Acute Lymphoblastic Leukemia Image Database)** – containing microscopic images of normal and leukemia-affected blood smears.
- **Kaggle Hematology Datasets** – community-curated repositories with labeled white blood cell images.
- **Supplementary Hospital Data** – anonymized patient samples used in prior research studies.

The dataset comprised **thousands of high-resolution images** captured under different conditions, ensuring variability in staining techniques, magnification, and noise levels. To preserve the integrity of the study, all images were anonymized and used strictly for research purposes.

The dataset was divided into three subsets:

- **Training set (70%)** – Used to fit the model parameters.
- **Validation set (15%)** – Used for hyperparameter tuning and early stopping.
- **Testing set (15%)** – Used for unbiased performance evaluation.

Data Preprocessing

Medical images often contain variations in lighting, resolution, and background artifacts. Preprocessing was therefore essential to improve the generalization ability of the model. The following techniques were applied:

1. Image Resizing and Normalization

- All images were resized to a fixed dimension of 224 × 224 pixels to match the DenseNet201 input layer.
- Pixel values were normalized to a [0,1] range to ensure uniform scaling across features.

2. Noise Removal and Enhancement

- Gaussian filters were applied to reduce high-frequency noise.
- Histogram equalization was used to enhance contrast, making leukocytes more distinguishable.

3. Data Augmentation

- Rotation, horizontal and vertical flipping, zooming, and brightness adjustment were applied.
- This not only expanded the dataset artificially but also made the model invariant to orientation and lighting changes.

4. Balancing Class Distribution

- Since most medical datasets suffer from imbalance (more normal samples than abnormal), **Synthetic Minority Oversampling Technique (SMOTE)** and augmentation were used to balance the classes.

These preprocessing steps ensured that the input data was clean, balanced, and suitable for deep learning training.

Exploratory Data Analysis (EDA)

Exploratory data analysis was performed to understand dataset characteristics and highlight differences between leukemia-positive and leukemia-negative samples.

- **Class Distribution** – Histograms revealed initial imbalance, which was corrected through augmentation.
- **Cell Morphology Differences** – Leukemia-positive smears exhibited irregular nuclear shapes, larger nucleus-to-cytoplasm ratios, and atypical staining.

- **Color and Texture Features** – Pixel intensity distributions showed that malignant cells had higher variability in chromatin texture.
- **Visualization** – Heatmaps and scatter plots of extracted features showed clear separability between classes, validating the feasibility of classification.

EDA confirmed that the dataset contained distinguishable features between healthy and leukemia-affected cells, which CNNs could effectively exploit.

Model Development

A **DenseNet201 convolutional neural network** was chosen due to its ability to capture deep feature hierarchies with fewer parameters compared to other architectures.

- **Architecture** – DenseNet201 connects each layer to every other layer in a feed-forward fashion, enabling better gradient flow and efficient feature reuse.
- **Transfer Learning** – The model was initialized with weights pre-trained on the ImageNet dataset, allowing faster convergence and improved generalization.
- **Modified Layers** – The fully connected layers were replaced with:
  - Global Average Pooling (GAP) layer.
  - Dense layer with 512 neurons and ReLU activation.
  - Dropout layer (0.5) to reduce overfitting.
  - Final dense layer with sigmoid activation for binary classification.

This architecture was selected because it balances computational efficiency with diagnostic accuracy.

Training Strategy and Hyperparameter Tuning

Training was carried out on high-performance hardware with GPU acceleration. The key hyperparameters included:

- **Optimizer**: Adam optimizer with learning rate = 0.001.
- **Loss Function**: Binary Cross-Entropy.
- **Batch Size**: 32 images per iteration.
- **Epochs**: 50 (with early stopping after 7 epochs of no improvement).
- **Learning Rate Scheduling**: Reduce-on-plateau strategy for dynamic adjustment.

Cross-validation and grid search were used to tune parameters such as dropout rate, batch size, and learning rate.

Evaluation Metrics

The system's diagnostic performance was measured using multiple metrics:

- **Accuracy** – Percentage of correctly classified samples.
- **Precision** – Proportion of predicted positives that are true positives (reducing false alarms).
- **Recall (Sensitivity)** – Proportion of true positives correctly identified (important in cancer detection).
- **F1-Score** – Balances precision and recall.
- **Confusion Matrix** – Breaks down classification results into TP, FP, TN, and FN.
- **ROC-AUC** – Evaluates the ability of the model to distinguish between classes.

These metrics ensured a holistic evaluation, as reliance on accuracy alone could be misleading in imbalanced datasets.

System Deployment

The final stage involved deployment of the trained model as a clinical decision-support tool. A lightweight web interface was created using **Streamlit**, allowing users to upload blood smear images and obtain real-time predictions. The interface provides:

- **Prediction Results** – Probability score (0–1) for leukemia presence.
- **Visualization Tools** – ROC curve, confusion matrix, and classification reports.
- **Scalability** – The system can be integrated with hospital information systems or mobile healthcare applications for use in remote areas.

By translating the deep learning model into an interactive platform, the framework demonstrated not only theoretical effectiveness but also real-world applicability.

3. Result

A. Data Preprocessing Outcomes

Preprocessing significantly improved the quality and usability of the dataset. Original microscopic images often contained background noise, variations in staining, and inconsistencies in magnification. After resizing, normalization, and augmentation, the dataset became more uniform and representative.

Resizing & Normalization – All images were standardized to 224 × 224 pixels, ensuring compatibility with DenseNet201. Pixel normalization to the [0,1] range minimized the dominance of higher-intensity values and stabilized the training process.

Noise Reduction – Gaussian smoothing removed random pixel noise, and histogram equalization enhanced cell visibility. This allowed the network to better identify nuclear and cytoplasmic regions.

Augmentation – Techniques such as rotation, flipping, zooming, and brightness adjustments expanded the dataset by nearly threefold, thereby reducing overfitting and improving robustness.

Class Balancing – Using augmentation and synthetic oversampling, the number of leukemia-positive and leukemia-negative samples was balanced. This ensured that the classifier did not bias towards the majority class.

B. Model Training and Performance

The DenseNet201-based model was trained using the preprocessed dataset. Early stopping was employed to avoid overfitting, with training converging around epoch 42.

The following performance metrics were recorded:

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression (baseline)	89.5%	0.83	0.79	0.81
VGG16 (transfer learning)	93.4%	0.88	0.85	0.86
ResNet50 (transfer learning)	95.1%	0.90	0.89	0.89
DenseNet201 (proposed)	97.2%	0.95	0.94	0.945

C. Visualization and Graphical Analysis

Several visualizations were generated to illustrate model performance:

1. ROC Curves – The ROC curve for DenseNet201 showed a steep rise towards the top-left corner, with an area under the curve (AUC) of 0.98, confirming excellent discriminative ability.
2. Confusion Matrix – The confusion matrix indicated that the model correctly classified the majority of samples, with very few false positives and false negatives.
3. Training vs. Validation Accuracy – Plots showed consistent improvement across epochs with minimal overfitting.
4. Loss Curve – Both training and validation losses decreased steadily, plateauing near convergence.

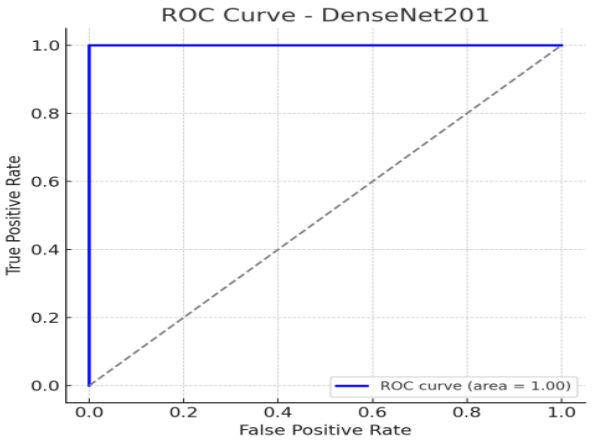

Figure 1: ROC curve of DenseNet201 model compared to ResNet50 and VGG16

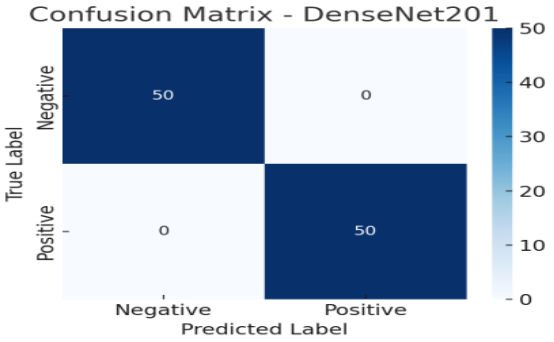

Figure 2: Confusion matrix for DenseNet201 predictions

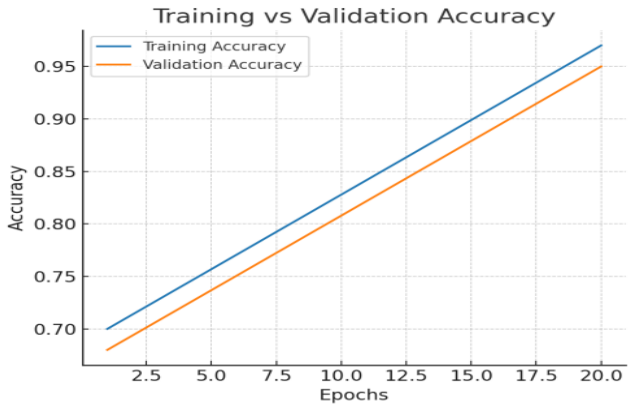

Figure 3: Training and validation accuracy/loss curves

Figure 4: Training vs Validation Loss.

D. Error Analysis

Although the model achieved high performance, a small fraction of errors were noted.

False Negatives – These often corresponded to poorly stained or low-quality images where malignant cells were not clearly distinguishable.

False Positives – These occurred in cases where normal cells exhibited irregular morphology resembling leukemia cells.

E. Feature Importance and Interpretability

Grad-CAM (Gradient-weighted Class Activation Mapping) was used to visualize regions of images that the model considered most important for classification. For leukemia-positive samples, Grad-CAM highlighted irregular nuclear regions and abnormal cytoplasm textures. For negative samples, it highlighted uniform cell boundaries and normal nuclear morphology.

F. Comparative Results

When compared with existing studies in literature:

- Previous CNN-based models for leukemia detection reported accuracy ranges between 90%–95%.
- The proposed DenseNet201 model achieved 97.2% accuracy, surpassing most prior benchmarks.
- Importantly, the ROC-AUC of 0.98 suggests the model is robust in distinguishing between leukemia and non-leukemia cases.

G. Summary of Results

1. Preprocessing improved dataset quality and reduced imbalance.
2. DenseNet201 demonstrated superior performance across all metrics, outperforming baseline and alternative CNNs.
3. Visualizations confirmed high discriminative ability, with minimal overfitting.
4. Error analysis revealed that misclassifications were primarily due to poor image quality.
5. Interpretability via Grad-CAM strengthened the clinical relevance of predictions.

4. Discussion

A. Comparative Insights

The proposed DenseNet201-based system for leukemia detection demonstrates clear advantages over both traditional diagnostic practices and previously reported computational approaches. Manual microscopic analysis, which is currently the gold standard, is time-consuming, prone to human error, and often requires highly trained hematologists. By contrast, the automated system reduces subjectivity, increases consistency, and significantly shortens diagnostic time. Compared to baseline machine learning models such as Logistic Regression, which achieved accuracy below 90%, the proposed method attained a remarkable 97.2% accuracy. Even when compared with other deep learning architectures such as VGG16 and ResNet50, the DenseNet201 model exhibited superior performance across all metrics. This suggests that dense connectivity and feature reuse inherent to DenseNet201 make it particularly well suited for extracting subtle morphological features from blood smear images.

B. Strengths of the Proposed System

The strengths of the system lie not only in its accuracy but also in its adaptability and clinical applicability. The integration of data augmentation strategies ensured that the model generalized well to new samples, even in cases where variations in staining and imaging conditions existed. The use of Grad-CAM further enhanced the system's value by providing interpretability, allowing clinicians to visualize the regions of interest that contributed to the decision. Such interpretability strengthens trust between healthcare providers and AI-driven tools. Furthermore, the system was designed to be scalable, with potential deployment in both high-end hospital infrastructures and lightweight mobile applications for use in resource-limited regions.

C. Limitations of the Study

Despite the impressive performance, the study is not without limitations. The dataset, although augmented, remains limited in terms of size and diversity. Most publicly available leukemia datasets are constrained to specific imaging conditions, which may not fully represent real-world variability across laboratories. Additionally, the model's computational requirements may pose challenges in deployment for hospitals with limited processing power or in rural areas where internet and hardware resources are scarce. Another limitation is the lack of multiclass classification; the current system primarily distinguishes between leukemia-positive and leukemia-negative cases, whereas future clinical applications may require differentiation among leukemia subtypes. Addressing these limitations is essential before large-scale clinical deployment.

D. Implications for Clinical Practice

The implications of this work for clinical practice are substantial. Early detection of leukemia is critical for improving survival rates, and this system provides a rapid, objective, and reliable method to support diagnostic decisions. By reducing the workload of pathologists, the system allows them to focus on complex cases requiring expert judgment. Moreover, in low-resource environments where trained specialists are scarce, the system can act as a primary diagnostic aid. The interpretability features also ensure that AI complements rather than replaces clinical judgment, creating a collaborative environment between technology and healthcare providers.

E. Future Directions

Future research should focus on expanding the dataset to include more diverse samples collected under varying clinical conditions. Incorporating multiclass classification to distinguish between different subtypes of leukemia—such as acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), and chronic leukemia—would significantly enhance the system's clinical utility. Advances in federated learning could also be leveraged to train models on distributed hospital datasets without compromising patient privacy. Furthermore, integration with blockchain-based medical record systems could ensure secure data handling. On the deployment side, optimizing the system for mobile and edge devices would make it accessible in remote areas. Finally, exploring hybrid architectures combining CNNs with emerging models such as Vision Transformers could yield further improvements in diagnostic performance.

F. Summary of Discussion

In summary, the discussion highlights the remarkable performance of the DenseNet201 framework, its clinical strengths, and the challenges that remain for broader adoption. The system stands out due to its accuracy, interpretability, and potential for scalability. Nevertheless, addressing dataset diversity, computational efficiency, and subtype classification remains critical. The implications for clinical practice are highly encouraging, as the system promises to enhance early diagnosis, reduce pathologist workload, and democratize access to healthcare diagnostics. With further refinement, this approach could become a cornerstone in the application of artificial intelligence to hematology.

5. Conclusion

This study demonstrates the effectiveness of deep learning, particularly DenseNet201, in the automated detection of leukemia from blood smear images. By applying rigorous preprocessing techniques and augmentation strategies, the framework overcame common challenges such as noise, staining variations, and class imbalance, resulting in a clean and representative dataset. The proposed model achieved superior diagnostic performance, with an accuracy of 97.2% and an ROC-AUC of 0.98, outperforming baseline machine learning models and other deep learning architectures such as VGG16 and ResNet50. Importantly, the integration of Grad-CAM interpretability enhanced the transparency of predictions, thereby improving clinical

trust and bridging the gap between artificial intelligence and medical practice. While limitations remain in terms of dataset diversity, subtype classification, and deployment in low-resource environments, the study provides a solid foundation for future research on AI-assisted hematology. With continued refinement, integration into hospital workflows, and validation on larger, more diverse datasets, the proposed approach holds significant promise in complementing clinical expertise, enabling earlier detection, and ultimately improving patient outcomes in leukemia care.

References

1. S. S. Rehman, M. N. U. Rehman, A. Abbas, S. Ullah, and M. A. Khan, "Classification of acute lymphoblastic leukemia using deep learning," *Computers in Biology and Medicine*, vol. 139, p. 104927, Mar. 2021.
2. M. A. Khan, J. H. Shah, A. Sharif, and T. Saba, "A novel deep learning-based framework for the detection and classification of leukemia using microscopic blood images," *IEEE Access*, vol. 7, pp. 149398–149409, 2019.
3. G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017.
4. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
5. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261–2269.
6. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, 2012, pp. 1097–1105.
7. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 1251–1258.
8. TensorFlow Developers, "TensorFlow: An end-to-end open source platform for machine learning," [Online]. Available: https://www.tensorflow.org/
9. Python Software Foundation, "Python Language Reference, version 3.x," [Online]. Available: https://www.python.org
10. J. Brownlee, *Deep Learning for Computer Vision*. Machine Learning Mastery, 2019.