International Journal of Engineering Research and Science & Technology

Leukaemia BLOOD CANCER PREDICTION USING DEEP LEARNING

K.Pavani[1] , P.Mohan Rao[2]

#1 Assistant Professor &Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract: Blood cancer, particularly Leukemia, is a serious disease that affects the blood and bone marrow. Early detection of leukemia is very important for improving patient survival rates and enabling timely medical treatment. Traditionally, leukemia diagnosis is performed manually by medical experts through microscopic examination of blood smear images, which can be time-consuming and prone to human error. This project proposes an intelligent deep learning–based system for automatic detection of leukemia from microscopic blood smear images. The system uses a Convolutional Neural Network (CNN) based on the EfficientNetB0 architecture to extract high-level visual features from blood cell images. These extracted features are further processed using the XGBoost machine learning classifier to improve prediction accuracy and classification reliability.

The model is trained on a labeled dataset containing leukemia-infected and healthy blood cell images. Data preprocessing techniques such as image normalization, augmentation, and balanced sampling are applied to improve model generalization. The trained model is integrated into a REST API-based web system that allows Lab Technicians to upload blood smear images, Doctors to review predictions and approve reports, and Administrators to monitor system performance and analytics.

The proposed hybrid CNN-XGBoost approach improves diagnostic accuracy and provides fast automated analysis of blood smear images. This system can assist medical professionals in early leukemia detection and help reduce diagnostic workload in hospitals and laboratories.

Key words: Leukemia Detection, Blood Smear Images, Deep Learning, Convolutional Neural Networks (CNN), EfficientNetB0, XGBoost, Medical Image Analysis, Artificial Intelligence in Healthcare.

1. INTRODUCTION

Blood cancer, commonly known as Leukemia, is a serious medical condition that affects the blood and bone marrow. It occurs when abnormal white blood cells grow uncontrollably and interfere with the normal production of healthy blood cells. Early detection of leukemia is very important because it significantly increases the chances of successful treatment and survival.

Traditionally, leukemia diagnosis is performed manually by medical experts through microscopic examination of blood smear images. This process requires highly trained specialists and can be time-consuming. Additionally, manual diagnosis may sometimes lead to human errors, especially when analyzing large volumes of medical images.

With the rapid advancement of Artificial Intelligence (AI) and Deep Learning, automated medical image analysis has become possible. Deep learning models are capable of learning complex visual patterns from medical images and assisting doctors in disease detection. This project proposes an intelligent system for automatic leukemia detection using deep learning and machine learning techniques.

The system uses a Convolutional Neural Network (CNN) based on the EfficientNet architecture to extract important features from blood smear images. These extracted features are then processed by an XGBoost classifier to perform accurate classification of leukemia and normal cells. The proposed system also provides a web-based platform where lab technicians can upload blood smear images, doctors can verify diagnostic reports, and administrators can monitor system analytics.

This automated approach can support medical professionals by providing faster and more reliable leukemia detection. In addition to improving diagnostic accuracy, the proposed system also enhances the efficiency of medical workflows by providing automated analysis and report generation. The integration of deep learning with machine learning algorithms enables the system to learn complex patterns present in blood smear images and make reliable predictions.

By combining CNN-based feature extraction with the powerful classification capability of XGBoost, the system achieves improved performance compared to traditional machine learning approaches. Furthermore, the system is designed with a scalable architecture that supports multiple user roles including lab technicians, doctors, and administrators.

2. LITERATURE SURVEY

1) 1. R. D. Labati, V. Piuri, and F. Scotti (2011)

Title: *All-IDB: The Acute Lymphoblastic Leukemia Image Database for Image Processing*

Abstract:
This paper introduces the Acute Lymphoblastic Leukemia Image Database (ALL-IDB), a publicly available dataset designed to support research in automated leukemia detection and image processing. The database contains microscopic blood smear images collected from patients diagnosed with Acute Lymphoblastic Leukemia (ALL) and healthy individuals. The authors describe the image acquisition process, annotation methods, and dataset characteristics. The database serves as a benchmark for developing and evaluating computer-aided diagnostic systems, enabling researchers to design and compare image processing, machine learning, and classification algorithms for leukemia detection.

2) 2. S. Rehman, S. Abbas, T. Saba, M. F. Khan, and U. Tariq (2018)

Title: *Classification of Acute Lymphoblastic Leukemia Using Deep Learning*

Abstract:
This study presents a deep learning-based framework for the classification of Acute Lymphoblastic Leukemia (ALL) using microscopic blood smear images. The proposed system employs convolutional neural networks (CNNs) to automatically extract discriminative features from blood cell images without requiring manual feature engineering. Image preprocessing and augmentation techniques are applied to improve model robustness and reduce overfitting. Experimental results demonstrate that the deep learning approach achieves high classification accuracy and outperforms conventional machine learning methods, highlighting the effectiveness of CNNs for leukemia diagnosis.

3) 3. R. K. Vogado, R. M. S. Veras, F. H. D. Araujo, et al. (2018)

Title: *Leukemia Diagnosis in Blood Smear Images Using Deep Learning*

Abstract:
This paper proposes an automated leukemia diagnosis system based on deep learning techniques for the analysis of peripheral blood smear images. The methodology integrates image preprocessing, segmentation, and convolutional neural network architectures to identify leukemia-infected cells. The model learns complex visual patterns directly from raw image data, reducing the dependency on handcrafted features. Experimental evaluations show promising diagnostic performance, demonstrating that deep learning can provide reliable support for medical professionals in leukemia screening and diagnosis.

4) 4. S. Rawat and N. Wang (2017)

Title: *Deep Convolutional Neural Networks for Image Classification*

Abstract:
This paper provides a comprehensive review of deep convolutional neural networks (CNNs) and their applications in image classification tasks. The authors discuss the architecture of CNNs, including convolutional layers, pooling operations, activation functions, and optimization techniques. The study highlights the superior performance of CNNs in extracting hierarchical image features and achieving state-of-the-art classification accuracy across various domains. Challenges such as computational complexity and overfitting are also addressed, along with strategies to improve model efficiency and generalization.

5) 5. T. Chen and C. Guestrin (2016)

Title: *XGBoost: A Scalable Tree Boosting System*

Abstract:
This paper introduces XGBoost, a scalable and efficient gradient boosting framework designed for high-performance machine learning applications. The proposed system incorporates regularization techniques, parallel processing, and optimized tree-learning algorithms to improve prediction accuracy while reducing computational cost. XGBoost effectively handles large-scale datasets and complex feature interactions, making it suitable for classification, regression, and ranking tasks. Extensive experiments demonstrate its superior performance compared to existing boosting methods, establishing XGBoost as one of the most widely used machine learning algorithms in predictive analytics and data science.

3..METHODOLOGY

i) Proposed Work:

In this project, an intelligent deep learning-based system is proposed for the automatic detection of leukemia using microscopic blood smear images. The system utilizes an EfficientNet-based Convolutional Neural Network (CNN) to automatically extract significant visual features from blood cell images, enabling accurate representation of cellular patterns associated with leukemia. The extracted deep features are then provided to an XGBoost classifier, which performs the final classification and predicts whether the given blood smear image corresponds to leukemia-infected cells or normal blood cells. To enhance the model's performance and generalization capability, image preprocessing techniques such as resizing, normalization, and data augmentation are employed. Furthermore, the system incorporates a role-based access mechanism that supports multiple user categories, including administrators, doctors, and lab technicians, ensuring secure and efficient management of medical data. The proposed platform also generates structured diagnostic reports based on prediction outcomes and enables doctors to review, verify, and approve the generated reports through a user-friendly web interface, thereby supporting accurate and timely clinical decision-making..

ii) System Architecture:

The architecture of the Blood Cancer Leukemia Detection System follows a structured approach to analyze microscopic blood smear images and classify them as leukemia or normal cells using deep learning and machine learning techniques.

For this project, SDLC provides the foundational framework for organizing each stage of development—from dataset preparation and model training to backend integration and user interaction—ensuring that every component functions cohesively within the overall system.

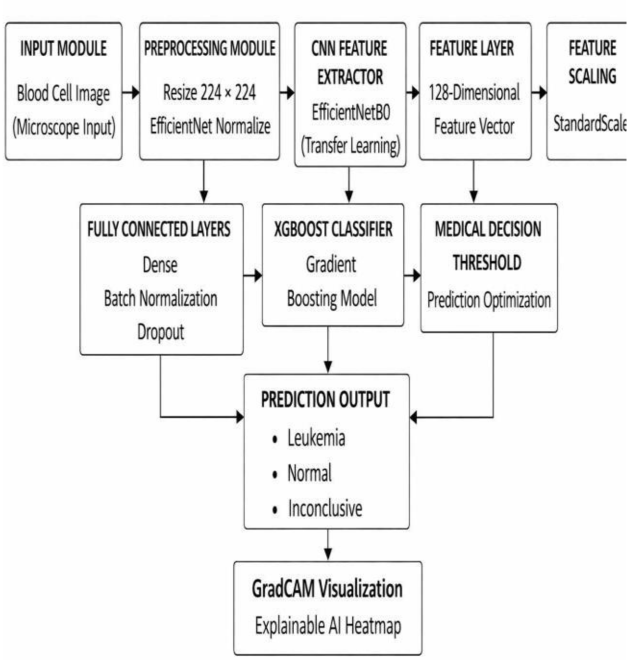


Fig1 proposed architecture

iii) Modules:

Module 1: Data Cleaning

Remove corrupted, duplicate, and irrelevant blood smear images.

Verify that each image is correctly labeled as leukemia-affected or normal.

Filter out images with poor quality, excessive noise, or unclear cell structures.

Improve dataset quality and consistency for model training.

Module 2: Data Visualization

Analyze the distribution of leukemia and normal blood cell images.

Use visualization techniques such as bar charts, pie charts, and histograms.

Identify dataset characteristics and class balance.

Ensure the dataset is well-organized before model training.

Module 3: Data Transformation

Convert blood smear images into numerical pixel values.

Resize images to a fixed input size.

Normalize pixel intensities to improve model performance.

Prepare images for processing by the CNN model.

Module 4: Dataset Splitting

Divide the dataset into training and testing sets.

Use the training set to teach the model.

Use the testing set to evaluate model performance on unseen data.

Ensure proper assessment of the leukemia detection model.

Module 5: Preprocessing

Resize, normalize, and enhance blood smear images.

Improve image quality for deep learning analysis.

Prepare images as input to the CNN model.

Module 6: Feature Extraction

Process preprocessed images using the CNN model.

Automatically extract important features such as cell shape, texture, and abnormal patterns.

Generate feature representations for classification.

Module 7: Classification and Prediction

Pass extracted features to the XGBoost classifier.

Analyze feature patterns and classify images.

Predict whether the blood smear image represents leukemia-affected cells or normal blood cells.

iv) Algorithms:

Convolutional Neural Network (CNN)

In this project, a Convolutional Neural Network (CNN) is used to analyze microscopic blood smear images for leukemia detection. The CNN automatically learns and extracts important visual features from the images, such as cell shape, texture, and abnormal patterns. Unlike traditional image processing methods, CNNs do not require manual feature extraction, as they can identify relevant patterns directly from the image data. The extracted features help the system distinguish between leukemia-affected blood cells and normal blood cells, improving the effectiveness of the detection process.

EfficientNet Architecture

The proposed system uses an EfficientNet-based CNN architecture for feature extraction. EfficientNet is designed to efficiently learn meaningful visual information from images while maintaining high performance. After preprocessing, the blood smear images are provided as input to the EfficientNet model, which extracts deep features representing important characteristics of the blood cells. These features are then used for further analysis and classification. The use of EfficientNet helps improve the quality of feature extraction and supports accurate leukemia detection.

XGBoost Classifier

After feature extraction, the extracted deep features are passed to an XGBoost classifier for final classification. XGBoost analyzes the feature patterns obtained from the EfficientNet model and predicts whether the blood smear image belongs to a leukemia-affected patient or a normal individual. The classifier uses the extracted information to make accurate predictions and improve classification reliability. By combining EfficientNet-based feature extraction with XGBoost classification, the proposed system provides an effective approach for automated leukemia detection from microscopic blood smear images..

4. EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

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

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{ the AP of class } k$$

$$n = \text{ the number of classes}$$

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

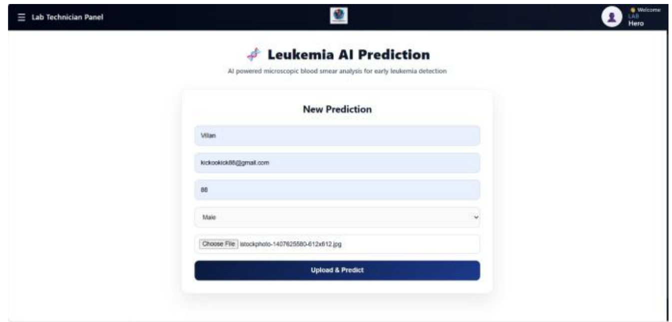
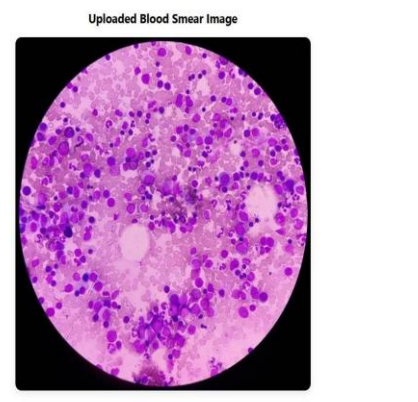


Fig 2:Choosing an image



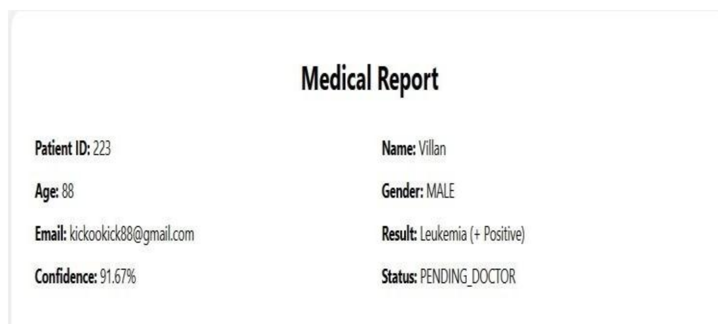


Fig 3:Medical Report

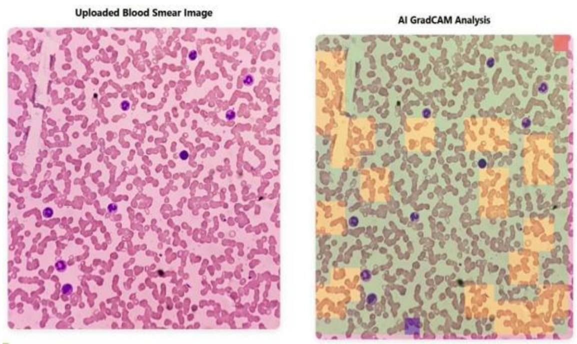


Fig 4:Results

5.CONCLUSION

In this project, a deep learning-based system is developed for the detection of leukemia from microscopic blood smear images. The proposed approach combines the feature extraction capability of Convolutional Neural Networks (CNN) with the classification power of the XGBoost algorithm to achieve accurate and reliable leukemia detection. The CNN model is employed to automatically extract important visual features such as cell shapes, textures, and abnormal structures present in blood smear images. These extracted features are then provided to the XGBoost classifier, which analyzes the feature patterns and predicts whether the given image represents leukemia-affected cells or normal blood cells.

The model is trained and evaluated using a publicly available leukemia image dataset. Experimental results demonstrate that the proposed CNN-XGBoost model achieves good accuracy and reliability in classifying blood smear images. The developed system can assist medical professionals by providing faster and more accurate leukemia detection, thereby reducing the time required for manual analysis and supporting the early diagnosis of leukemia.

Furthermore, the use of an EfficientNet-based CNN architecture enhances the system's ability to learn complex patterns from microscopic blood smear images. By combining deep learning feature extraction with the machine learning capabilities of XGBoost, the proposed model improves prediction performance and provides better generalization when analyzing unseen images. The system also incorporates preprocessing techniques such as image resizing, normalization, and data augmentation to improve image quality and enhance model training. These preprocessing steps enable the model to learn more robust and discriminative features from the dataset.

Overall, the proposed CNN–XGBoost-based leukemia detection framework demonstrates the significant potential of artificial intelligence in medical image analysis. The system provides a reliable, efficient, and automated approach for leukemia detection and can serve as a valuable decision-support tool for healthcare professionals in clinical environments..

6. FUTURE SCOPE

- Although the proposed leukemia detection system provides reliable performance, there are several opportunities for further improvement and expansion. Training the model with a larger and more diverse dataset of blood smear images can significantly enhance the accuracy, robustness, and generalization capability of the system. A dataset containing images collected from different laboratories, microscopes, and patient conditions would enable the model to learn more representative patterns and reduce the risk of overfitting.

•

• The current system performs binary classification to determine whether a blood smear image is leukemia-affected or normal. In future work, the framework can be extended to support multi-class classification for identifying different types of leukemia, such as Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), and other blood-related disorders. Such an enhancement would increase the clinical usefulness of the system and provide more detailed diagnostic information.

•

• Further improvements can also be achieved by exploring advanced deep learning architectures, including EfficientNet variants, Vision Transformers (ViT), and hybrid deep learning models. These architectures have shown promising performance in medical image analysis and may provide superior feature extraction capabilities and classification accuracy when trained on larger datasets.

•

• In addition, the proposed system can be expanded into a comprehensive clinical decision support platform that provides detailed diagnostic reports, confidence scores, visual explanations, and treatment-related recommendations. Such a system can assist healthcare professionals in making faster and more informed decisions, reduce the workload of laboratory experts, and improve the overall efficiency of leukemia diagnosis in hospitals and diagnostic centers

REFERENCES

1. Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing SystemsElie Krevat, Elliot Cuzzillo. Improving Off-line Handwritten Character Recognition with Hidden Markov Models.

2. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the International Conference on Machine Learning (ICML).

3. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

4. Litjens, G., et al. (2017). A Survey on Deep Learning in Medical Image Analysis. Medical Image Analysis.

5. Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification. Neural Computation, 2017.

6. National Cancer Institute. Leukemia Overview. Available at: https://www.cancer.gov/types/leukemia.

7. Matek C, Schwarz S, Marr C, Spiekermann K. Human-level recognition of blast cells in acute myeloid leukemia with deep learning. Nature Machine Intelligence, 2019.

8. Espana-Boquera S, Castro-Bleda MJ, Gorbe-Moya J, Zamora-Martinez FB (2011) Improving offline handwritten text recognition with hybrid HMM/ANN models. IEEE Trans Pattern Anal Mach Intell 33(4) TensorFlow Documentation. Image Classification using Convolutional Neural Networks. Available at: https://www.tensorflow.org.

9. Scikit-learn Documentation. Machine Learning in Python. Available at: https://scikitlearn.org.

10. XGBoost Documentation. XGBoost Machine Learning Library. Available at: https://xgboost.readthedocs.io..

Author Profiles

Ms.K.Pavani Working as Assistant & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with MCA ,M. Tech in Computer Science .She has 10 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada,NTR District. Her area of interest includes AI ML, etc.

P.Mohan Rao is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in BCA(Bachelor of computer applications) from GBR Degree college of Anaparthi. His area of interest are DBMS and Deep Learning with Python.