BLOOD CANCER IDENTIFICATION USING DEEP LEARNING TECHNIQUE

Mrs. K. Sunanda [1], Mudumbai Sahithi[2], Emmadi Leelavardhan Raj[3], Manupati Anil[4], Gatla Bhanu Teja[5]

[1] Assistant Professor Department of CSE-Data Science, Sreyas Institute of Engineering and Technology, [2,3,4,5] Department of CSE, Sreyas Institute of Engineering and Technology

Abstract. Early diagnosis of a disease is an important aspect of management; however, blood cancer creates a health hazard in these times. "Blood Cancer Identification using Deep Learning Technique" aims to develop a state-of-the-art diagnostic system, parlaying a hybrid ensemble technique to augment the blood cancer identification process's precision and accuracy. The platform should strive for a smooth user experience with an intuitive UI so that users can interact with the platform by uploading diagnostic images for accurate prediction for their health status. This hybrid approach ensures the performance of other limitations such as overfitting, data imbalance, and model generalizations. Preprocessing of blood smear images is then done, augmenting the data to increase its diversity before fine tuning the ensemble with the advanced optimization techniques. The major objective of this project is to provide a worthy solution that will facilitate the early detection and accurate identification of blood cancer and its types to a great extent, making it a landmark in medical diagnostics. The system achieves an impressive **accuracy of 99.77%**, making it a highly effective tool for clinical support.

Keywords: Blood Cancer Detection, Deep Learning Techniques, Diagnostic Images, Blood Smear Images, Medical Diagnostics

1 Introduction

Early diagnosis and therapy support blood cancer patient survival. Traditionally, the diagnosis of blood cancers, particularly subtypes such as Acute Lymphoblastic Leukemia (ALL), has been reliant on the manual, microscopic blood smear analysis, which is a time-intensive, subjective process subject to inter- and intra-anatomic variability and human error. This project aims to develop an innovative solution to change the diagnostic workflow through a Hybrid Ensemble Deep Learning System that combines the power of EfficientNetB3, transfer learning, and ensemble modelling approaches to achieve high-accuracy blood smear image classification. An innovative component of this development is the complex CNN-based System to detect significant features, lessen manual involvement and achieve a level of multiplicative differentiation (L1, L2, L3) by recognition of semantic characteristics in scales of medical images. The project has built in sophisticated data augmentation processes, Adamax optimization and softmax probability based classifications to ensure applicability of solutions in multi-classification problems like those found in medical data and particularly blood cancers because of overfitting and under sampling issues with data imbalance. Due to the structured web interface component of the project, the application allows for motion to capture still images to be uploaded, real-time feedback on diagnostic class and visual representation in the user experience making it easily available for clinical diagnosis at the front line of medical practitioners and with diagnostic labs. This framework aims to enable precision, speed, and scalability in leukemia diagnosis, contributing to a large step towards the advancement of AI-assisted healthcare.

2 Objectives

The primary objective of this study is to create a robust diagnostic paradigm that integrates deep learning methods to facilitate the timely and precise detection of blood cancer, with a particular emphasis on Acute Lymphoblastic Leukemia (ALL). Leukemia represents a serious epidemiological concern that is characterized by a disruption of hematopoiesis, evidenced by producing excessive quantities of various abnormal white blood cells. The diagnosis is mainly based on the interpretation of peripheral blood smear photographs made by a qualified hematologist under a microscope. This manual analysis provides limited potential for consistency, replicability, and speed, and is especially problematic for detecting subtle morphological changes in the earliest stages of cancer. For these reasons, this project describes a hybrid ensemble deep learning framework that will increase diagnostic accuracy and analysis time efficiency. The diagnostic framework will utilize convolutional neural networks (CNNs) to automatically extract complicated features from blood smear photographs, thus alleviating the burden associated with conventional handcrafted feature-engagement. The framework central to this design will include an EfficientNetB3 using transfer learning, which will very much improve classification accuracy and lower training time. The project will also determine various ALL subtypes (L1, L2, L3) which can be critical for ensure a proper treatment pathway. By automating the potentially fraught manual diagnostic process, the framework will help to eliminate human error, minimize time to diagnosis, and improve reliability especially in settings with limited resources. The intended contribution of the proposed diagnostic framework is to produce a scalable, accurate and efficient design to aid healthcare practitioners in clinical decision-making, health systems and empowerment, and patient outcomes.

3 Scope

This research is focused on the design and implementation of an artificial intelligence-driven system for diagnosing blood cancer, i.e., Acute Lymphoblastic Leukemia (ALL) detection from stained microscopic blood smear images. It is proposed to overcome the drawbacks of existing diagnostic techniques—low accuracy, low generalization, and no subtype classification—through a hybrid ensemble of deep neural networks for feature extraction and classification in an automatic way. The platform has the ability to correct several image inconsistencies such as non-uniform staining and lighting using sophisticated preprocessing methods. The model uses transfer learning so that it can function effectively even with lower-labeled data. The platform features a simple web frontend for image upload, subtype predictions, confidence level viewing. The ultimate goal is to present a cost-effective, secure, and scalable diagnostic tool appropriate for both high-end modern hospitals and resource-poor settings.

4 Methodology

In this research, we introduce an overall methodology with data fetching, preprocessing, model building, training, testing, and launching. Microscopic blood smear images were initially downloaded from publicly accessible sources such as Kaggle, with reference to healthy cells and leukemic categories such as Pre-B, Pro-B, and early Pre-B cells. Images were preprocessed with resizing to 128x128 pixels, normalization, and data augmentation (including rotation, flip, and brightness modifications) to achieve diversity in the dataset and address class imbalance. Labels were encoded numerically, and a custom data generator was used for batch processing in training. The model architecture uses EfficientNetB3 with high accuracy and computation efficiency choice. Transfer learning was applied by starting with the pre-trained weights of ImageNet and thereafter fine-tuning them on the leukemia data. A hybrid ensemble strategy was used to aggregate predictions from more than one checkpoint with the application of soft voting in order to be more robust. Adamax optimizer and categorical cross-entropy loss were applied to train the model while accuracy, precision, recall, F1-score, and confusion matrix were applied to test the model. Finally, the trained model was deployed through a Django-based web interface so users can upload images and receive diagnostic outcomes in real time. This method is guaranteed to make the system useful and accessible for real-world implementation.

5 System Architecture

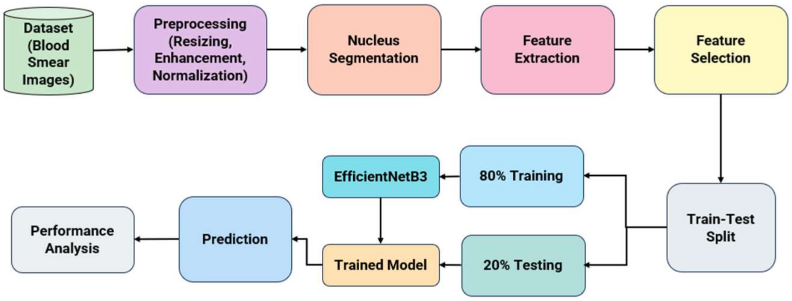


Fig. 5 System Architecture

5.1 Algorithms

- **EfficientNetB3:** EfficientNetB3 is used for feature extraction, is a convolutional neural network (CNN) with great balanced scaling of depth, width, and resolution requiring high accuracy and few parameters. It is employed in this project to automatically extract deep features from blood smear images. Employing transfer learning based on pre-training on ImageNet, it is used to speed up training and enhance classification performance.
- **Adamax Optimizer:** Adamax is one of the variants of the Adam optimizer designed for use in data that is of high dimension. Adamax learns the learning rate adaptively for each parameter and thereby ensures making stable and efficient convergence during training. The optimizer improves the model's performance particularly in tasks of complicated classification such as determining the subtype of leukemia.
- **Softmax Activation :** Softmax is used for the last layer of the neural network for transforming raw output scores into a probability distribution over multiple classes (e.g., L1, L2, L3, and Normal).It enables the model to predict in a probabilistic manner and pick the highest likely class for every input image.
- **Categorical Cross-Entropy :** This loss function computes difference between output class probability distribution and actual class labels in multi-class classification. It is minimized for improving model accuracy during training. It avoids the model learning to predict high confidence to an incorrect class and punishes wrong predictions.

5.2 Dataset

The sample images taken from the dataset named **Acute Lymphoblastic Leukemia (ALL) image dataset** (downloaded from Kaggle).

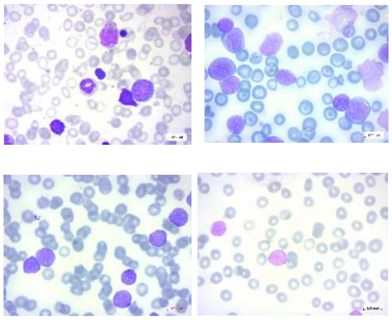
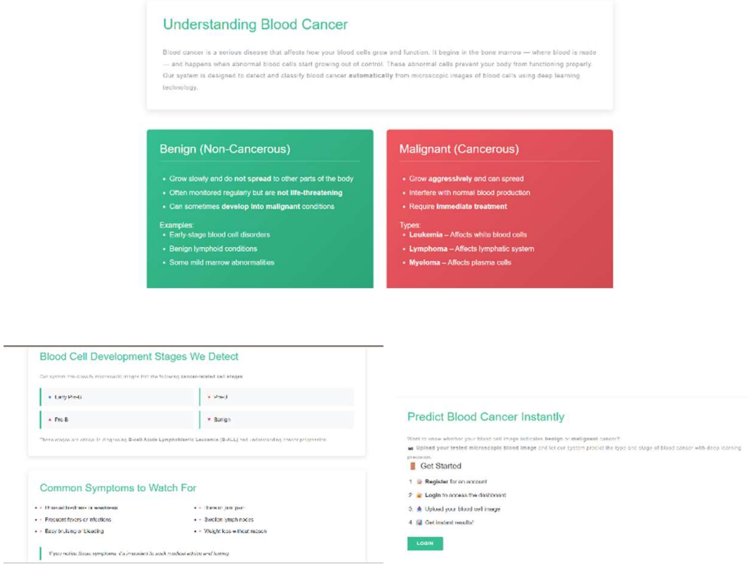


Fig. 5.2.1 Data Set

5.3 Data Preprocessing

Data preprocessing is an important step here normalizing and optimizing the blood smear images to facilitate training a deep learning model. The original data images from various online storage sources are plagued with inconsistencies in resolution, lighting, staining properties, and noise. The preprocessing pipeline starts with resizing all input images to a standard resolution of 128x128 pixels in order to introduce homogeneity to the dataset. Pixel values are also normalized to [0,1] range in order to increase training convergence and stability of the model.Noise reduction filters are then applied to sharpen images and eliminate background noise. To address the issue of sparse data and imbalance, advanced data augmentation techniques are employed. These involve horizontal and vertical flipping, rotation with random angles, zooming, shifting, and brightness/contrast adjustment. These not only expand the dataset size but also allow the model to learn invariant features, hence making it more generalizable. The class names, initially in string format ('Benign', '[Malignant] Pre-B'), are converted into numerical values using one-hot encoding for neural network loss function compatibility. Apart from augmentation and normalization, images are also quality-checked with good-staining and in-focus images only being used during training. This is done through simple threshold checks on histogram distribution and image sharpness. The data is then divided into training (80%) and testing (20%) sets for unbiased testing. During training, a dedicated data generator is applied for generating mini-batches of data, allowing memory-efficient usage and online augmentation. Overall, this all-encompassing preprocessing pipeline puts the dataset through to be highly prepared for thorough model training so that the network gets consistent, varied, and high-quality input mimicking real-world diagnostic scenarios.

5.4 Results



Understanding Blood Cancer

Blood cancer is a serious disease that affects how your blood cells grow and function. It begins in the bone marrow — where blood is made — and happens when abnormal blood cells start growing out of control. These abnormal cells prevent your body from functioning properly. Our system is designed to detect and classify blood cancer **automatically** from microscopic images of blood cells using deep learning technology.

Benign (Non-Cancerous)

- Grow slowly and do **not spread** to other parts of the body
- Often monitored regularly but are **not life-threatening**
- Can sometimes **develop into malignant** conditions

Examples:
- Early-stage blood cell disorders
- Benign lymphoid conditions
- Some mild marrow abnormalities

Malignant (Cancerous)

- Grow **aggressively** and can spread
- Interfere with normal blood production
- Require **immediate treatment**

Types:
- **Leukemia** – Affects white blood cells
- **Lymphoma** – Affects lymphatic system
- **Myeloma** – Affects plasma cells

Blood Cell Development Stages We Detect

Our system can classify microscopic images into the following cancer-related cell stages:

- Early Pre-B
- Pre-B
- Pro-B
- Benign

These stages are critical in diagnosing **B-cell Acute Lymphoblastic Leukemia (B-ALL)** and understanding cancer progression.

Common Symptoms to Watch For

- Unusual tiredness or weakness
- Frequent fevers or infections
- Easy bruising or bleeding
- Bone or joint pain
- Swollen lymph nodes
- Weight loss without reason

If you notice these symptoms, it's important to seek medical advice and testing.

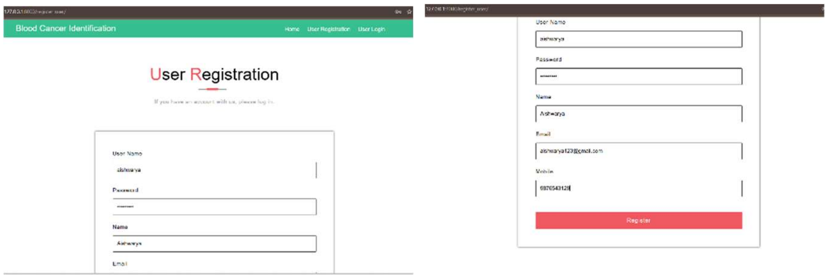
Predict Blood Cancer Instantly

Want to know whether your blood cell image indicates **benign** or **malignant** cancer?
Upload your tested microscopic blood image and let our system predict the type and stage of blood cancer with deep learning precision.

Get Started

1. **Register** for an account
2. **Login** to access the dashboard
3. Upload your blood cell image
4. Get instant results!

Fig. 5.4.1 Home Page

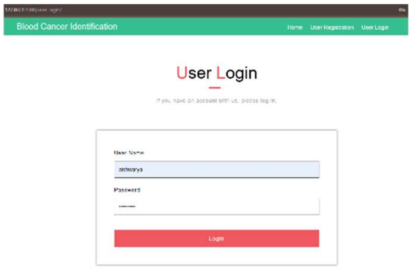


User Registration

If you have an account with us, please log in.

User Name: aishwarya
Password: ••••••••
Name: Aishwarya
Email: aishwarya123@gmail.com
Mobile: 6876543129

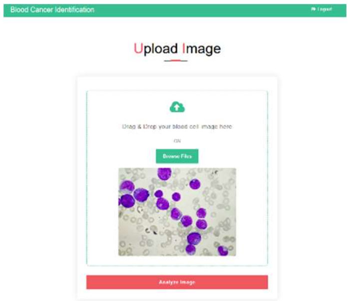
Fig. 5.4.2 User Registration



If you have an account with us, please log in.

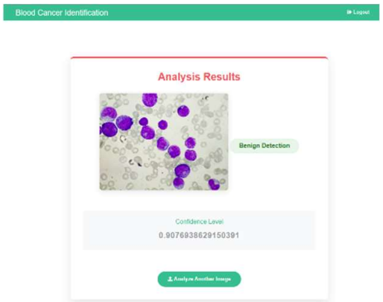
User Name: aishwarya

Fig. 5.4.3 User Login



Drag & Drop your blood cell image here

Fig. 5.4.4 Upload Image



Benign Detection

Confidence Level
0.9076938629150391

Fig. 5.4.5 Analysis Result

6 Conclusion

This study presents a deep learning-based method for improving blood cancer diagnosis through semantic visual inspection of microscopic blood smears. Conventional diagnosis heavily depend on manual interpretation, which is inefficient and subject to human bias. The system proposes an improved and scalable process of diagnosing Acute Lymphoblastic Leukemia (ALL) and its variants through the use of a hybrid ensemble model developed using EfficientNetB3 combined with transfer learning.Through extensive preprocessing and data enhancement, the system maintains consistency and enhanced generalization. Utilizing Adamax optimizer it maintains efficient model training and accurate predictions. Also, access via a simple web application expands access to make it simple for clinicians and diagnosing laboratories to simply access quick, accurate information through simple upload of smear photos. This platform not only saves time for diagnosis but also increases confidence and transparency in clinical judgment. Overall, this solution has a valuable position in AI-assisted healthcare, towards making early and reliable detection of blood cancer affordable and effective across all medical environments.

7 References

[1] Blood Cancer Identification using Hybrid Ensemble Deep Learning Technique. (2023, March 2). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10084996

[2] Detection of Blood Cancer-Leukemia using K-means Algorithm. (2021, May 6). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9432244

[3] *Identification of leukemia diseases based on microscopic human blood cells using image processing. (2018, October 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8579387*

[4] Blood cancer detection using improved machine learning algorithm. (2023, August 10). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10245375

[5] Jawahar, M., Anbarasi, L. J., Narayanan, S., & Gandomi, A. H. (2024). An attention-based deep learning for acute lymphoblastic leukemia classification. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-67826-9

[6] Alsalem, M., Zaidan, A., Zaidan, B., Hashim, M., Madhloom, H., Azeez, N., & Alsyisuf, S. (2018). A review of the automated detection and classification of acute leukaemia: Coherent taxonomy, datasets, validation and performance measurements, motivation, open challenges and recommendations. *Computer Methods and Programs in Biomedicine*, *158*, 93–112. https://doi.org/10.1016/j.cmpb.2018.02.005

[7] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. a. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

[8] Deshpande, N. M., Gite, S., & Aluvalu, R. (2021). A review of microscopic analysis of blood cells for disease detection with AI perspective. *PeerJ Computer Science*, *7*, e460. https://doi.org/10.7717/peerj-cs.460

[9] Anilkumar, K., Manoj, V., & Sagi, T. (2021). Automated detection of B cell and T cell acute lymphoblastic leukaemia using deep learning. *IRBM*, *43*(5), 405–413. https://doi.org/10.1016/j.irbm.2021.05.005

[10] Das, P. K., & Meher, S. (2021). An efficient deep Convolutional Neural Network based detection and classification of Acute Lymphoblastic Leukemia. *Expert Systems With Applications*, *183*, 115311. https://doi.org/10.1016/j.eswa.2021.115311