

DESIGN OF A DEEP LEARNING BASED MODEL FOR LEUKEMIA DETECTION

¹Ms. Jyoti Ahlawat, Research Scholar, ²Dr. Banita, Associate Professor

Department of Computer Science & Engineering, BMU, Rohtak.

Abstract- Leukemia is a life-threatening hematological malignancy that requires early and accurate diagnosis to improve patient outcomes. Manual examination of microscopic blood smear images is time-consuming, subjective, and highly dependent on expert pathologists. With recent advances in artificial intelligence, deep learning has emerged as a powerful tool for automated medical image analysis. The goal of this research paper is to develop a deep learning-based model that can accurately detect leukaemia from medical images, with a focus on optimizing the model's performance using advanced techniques such as transfer learning, hyper parameter tuning, and regularization methods. Evaluation metrics such as accuracy, precision, recall, F1 score, and the ROC-AUC curve will be used to assess the model's diagnostic ability. By building a robust and scalable deep learning model for leukaemia detection, this study aims to contribute to the growing body of research on AI-driven medical diagnostics and provide a practical tool to assist healthcare professionals in early leukaemia diagnosis.

Index Terms- Leukemia detection, Deep learning, Convolutional neural network, Medical image analysis, automated diagnosis.

I. INTRODUCTION

Leukemia is a type of blood cancer characterized by the abnormal proliferation of white blood cells, which interferes with the normal functioning of the immune system. Early diagnosis plays a crucial role in increasing survival rates and reducing treatment complications. Traditionally, leukemia diagnosis relies on microscopic examination of blood smears and bone marrow samples by skilled pathologists. However, this manual process is labor-intensive, time-consuming, and prone to inter-observer variability. Recent advancements in deep learning have significantly improved automated image classification and pattern recognition tasks. Convolutional neural networks (CNNs), in particular, have shown remarkable performance in medical imaging applications such as cancer detection, disease classification, and pathology analysis. Motivated by these developments, this research aims to design a deep learning based model that can accurately detect leukemia from microscopic blood images, thereby assisting medical professionals in diagnosis and decision-making. Several studies have explored machine learning and deep learning techniques for leukemia detection. Traditional machine learning approaches relied on handcrafted features such as cell shape, texture, and color, followed by classifiers like support vector machines or k-nearest neighbors. While these methods achieved moderate success, their performance was limited by feature extraction complexity. Recent research has shifted toward deep learning-based approaches, particularly CNNs, which automatically learn hierarchical features from raw images. Various CNN architectures, including custom models and transfer learning-based networks, have been applied to leukemia image datasets with improved accuracy.

Despite these advancements, a challenge such as dataset imbalance, image quality variations, and model generalization remain areas of ongoing research. Medical image processing is having various associated challenges in disease identification. These challenges are different according to the image type, disease type and organ specification. Blood cancer is one such critical disease that can cause death if it is not detected at the early stage. Leukemia is a one blood cancer disease that occur with the increase of White Blood cells in bone marrow. The image processing methods can be applied to identify the blood cells and to predict the Leukemia over the bone marrow images. The proposed work model has provided segmentation based deep learning model to detect the Leukemia disease accurately. As on the earlier study, the functional stages of this leukemia detection will be divided in two functional stages. In the first stage, the segmentation will be performed to identify the white blood cells over the bone marrow images. The clustering and segmentation algorithms will be applied in this stage for identify the key segments over the image. Once the effective area will be identified, this segmented image will be passed to the deep learning model to predict or classify the bone marrow images. The classification will be performed to isolate the healthy and Leukemia patients. A lot of earlier algorithms, models, and methods were proposed by the researchers to predict the Leukemia disease. But till, accuracy and performance are key challenges in this research area. The proposed hybrid model will be designed to improve the accuracy rate of leukemia detection.

II. OBJECTIVES

The primary objective of the proposed research is to design a hybrid clustering-classification-based model for accurate leukemia detection. The study also aims to collect and preprocess relevant medical image datasets suitable for leukemia analysis, implement the proposed model using an image-processing-friendly tool or programming language, and evaluate its performance by comparing it with existing leukemia detection methods and models.

III. RESEARCH METHODOLOGY

In the proposed work, an effective and adaptive deep learning model will be designed to detect the blood cancer from bone marrow images. The blood cells are the small and complex components present in the bone marrow images. In the proposed model, an effective preprocessing and segmentation stages will be defined for extracting the key region from the images. The preprocessing stage is defined to improve the component visibility within bone marrow image. The contrast and brightness adjustment with exploration of bone marrow features. Once the image will be improved, the segmentation method will be applied to extract the key region. In this work, the white blood cell identification is the key requirement. After extracting the region, the deep learning model will be applied to classify the bone marrow images. The functional stages of this proposed work are provided in the given flowchart.

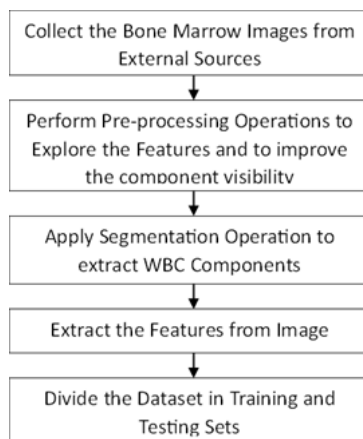


Figure : Expected Steps in the Proposed Model

Proposed Deep Learning Model

The proposed model is based on convolutional neural network architecture. It consists of multiple convolutional layers for feature extraction, followed by pooling layers to reduce spatial dimensions. Fully connected layers are used for classification, and a softmax activation function is applied in the output layer to generate class probabilities.

Key components of the model include:

- Convolutional layers with ReLU activation
- Max-pooling layers for dimensionality reduction
- Dropout layers to prevent over fitting
- Fully connected layers for classification

The proposed study would primarily focus on reviewing the methods and findings of existing Leukaemia detection. We found certain gaps in the traditional methods used to predict Leukaemia throughout our research. Findings from this study pave the path for the use of edge-based Hybrid CNNs in prediction of Leukaemia. The task at hand is to put the given method into action. The proposed technique and algorithm have been compared to the gold standard. Relative to the already employed methods for breast cancer prediction, the proposed research is said to be more efficient.

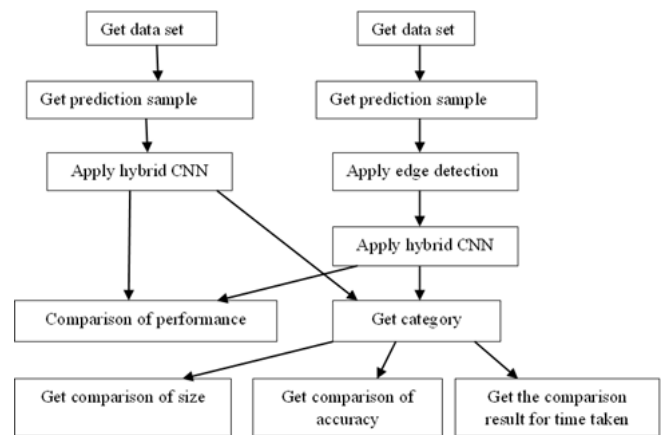


Figure: Process Flow of Work

IV. PROCESS FLOW OF PROPOSED WORK

Step 1: Image sample would be collected.

Step 2: Apply tradition CNN classifier in order to check time and space consumption

Step 3: Apply edge detection on image set.

Step 4: Apply proposed hybrid CNN model in order to check time and space consumption.

Step 5: Compare space consumption and performance of conventional and proposed work.

Classification of Data Set

Designing a deep learning-based model for leukaemia detection involves several steps, including data classification, model selection, and performance evaluation. Here's a structured approach to classification and model design for this task:

Dataset Preparation

- **Data Collection:** Obtain a dataset that contains relevant features for leukaemia detection, such as blood cell counts, genetic data, or medical imaging.
- **Data Cleaning:** Remove any duplicate entries, handle missing values, and correct any inconsistencies in the dataset.
- **Data Labelling:** Ensure that the dataset is properly labelled, distinguishing between leukaemia cases and healthy controls, or different types of leukaemia if applicable.

Data Pre-processing

- **Normalization/Standardization:** Scale the data to ensure that all features contribute equally to the model's performance.
- **Data Augmentation:** For imaging data, use techniques like rotation, flipping, or zooming to artificially increase the size of the dataset and improve model robustness.
- **Feature Selection:** Use techniques like Principal Component Analysis (PCA) or Recursive Feature Elimination (RFE) to select the most relevant features for the model.

Model Design

- **Choose an Architecture:** Depending on the data type, you can choose different architectures:
 - For tabular data: Fully connected neural networks (FCNN).
 - For imaging data: Convolutional Neural Networks (CNN).
 - For sequential data (like genetic sequences): Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM) networks.

Define the Model:

- **Input layer** (size corresponding to the features).
- **Hidden layers** with activation functions (ReLU, sigmoid, etc.).
- **Output layer** (softmax for multi-class classification or sigmoid for binary classification).

Model Training

- **Split the Dataset:** Divide the dataset into training, validation, and test sets (commonly 70% training, 15% validation, and 15% testing).
- **Compile the Model:** Choose an optimizer (like Adam or SGD), loss function (binary cross-entropy for binary classification, categorical cross-entropy for multi-class), and metrics (accuracy, precision, recall).
- **Train the Model:** Fit the model on the training set and validate it on the validation set. Monitor performance metrics to avoid overfitting.

Model Evaluation

- **Testing:** Evaluate the final model on the test set to assess its performance.
- **Performance Metrics:** Use metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) to measure model effectiveness.
- **Confusion Matrix:** Analyze the confusion matrix to understand classification errors.

Hyperparameter Tuning: Utilize techniques like Grid Search or Random Search to find the optimal hyperparameters (learning rate, batch size, number of epochs, etc.).

Deployment: Once validated, deploy the model in a real-world setting, ensuring it can handle incoming data and provide predictions.

Continuous Monitoring: Monitor the model's performance over time and retrain it periodically with new data to maintain accuracy.

This structured approach will help in designing a deep learning model for leukemia detection effectively. Make sure to iterate on model design and preprocessing steps based on the specific characteristics of your dataset.

V. CHALLENGES

Conducting research in leukemia detection using deep learning and related methodologies presents various challenges. One significant issue is the availability and quality of data; high-quality, labeled datasets specific to leukemia can be scarce, particularly for rare subtypes. Additionally, there may be a class imbalance, with significantly more healthy samples than leukemia cases, which can bias model training and affect accuracy. Missing data further complicates matters, as incomplete records can lead to gaps in critical information necessary for effective model training and validation. Data privacy and ethical considerations also play a crucial role; health data is sensitive and subject to strict regulations, complicating data sharing and usage, while obtaining informed consent from patients can limit data collection efforts. Moreover, the complexity and interpretability of models pose challenges. Deep learning models, while powerful, can easily overfit to training data, especially when datasets are small. Additionally, their complexity often leads to a lack of transparency, making it difficult to explain decisions to clinicians or patients. Feature engineering adds another layer of difficulty, as identifying and selecting the most relevant features from vast datasets can be time-consuming and typically requires in-depth domain knowledge. From a practical standpoint, computational resources are a critical factor. Training deep learning models demands significant computational power, which can be expensive and not readily

accessible to all researchers. This limitation can hinder experimentation and model iteration. Furthermore, selecting appropriate evaluation metrics for model performance is complex, particularly with imbalanced datasets, and ensuring models generalize well to unseen data necessitates careful validation strategies. Integrating these models into clinical settings presents its own challenges. Bridging the gap between research and clinical practice is crucial for real-world applicability, requiring effective collaboration between data scientists and healthcare professionals, which can be difficult to establish and maintain. Additionally, navigating the regulatory landscape for deploying AI in healthcare can be complex and time-consuming, with gaining regulatory approval often necessitating extensive validation studies that delay implementation.

The need for continuous learning and adaptation is also vital in this rapidly evolving field. Models must frequently be updated to incorporate new findings and treatment protocols, requiring ongoing research and validation. Moreover, researchers must be aware of the potential for algorithmic bias, as deep learning models can inadvertently learn biases present in training data, leading to unfair treatment of certain demographic groups. Ensuring equitable access to AI technologies for diverse populations remains a significant challenge. Addressing these multifaceted issues necessitates a multidisciplinary approach, combining expertise in data science, clinical knowledge, ethics, and regulatory affairs. Continuous collaboration among researchers, clinicians, and data scientists is essential to overcome these hurdles and enhance leukemia detection methods through advanced deep learning techniques.

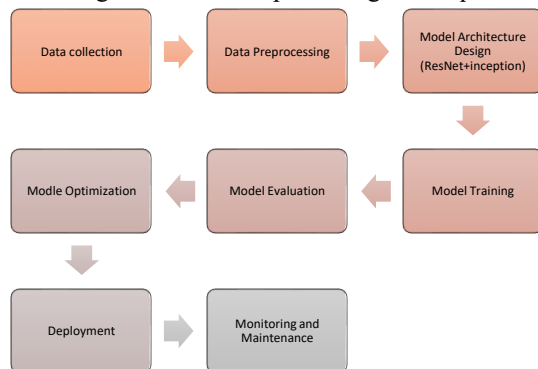


Figure: Process flow of Proposed Work

This process flow outlines the critical stages involved in developing a Hybrid CNN model by integrating ResNet and Inception networks for leukemia detection. Each step emphasizes the importance of data handling, model design, evaluation, and deployment, ensuring that the final product is both accurate and clinically relevant.

VI. CONCLUSIONS & FUTURE SCOPE:

This paper presented the design and development of a deep learning based model for leukemia detection using microscopic blood images. By leveraging convolutional neural networks, the proposed system automatically extracts discriminative features and achieves high diagnostic accuracy. The results indicate that deep learning has significant potential in assisting medical professionals with early and accurate leukemia diagnosis. Future work will focus on extending the model to multiclass leukemia classification and validating its performance on larger and more diverse datasets. Future enhancements of this research may include the use of advanced deep learning architectures, integration of transfer learning, incorporation of explainable AI techniques, and real-time clinical deployment for large-scale screening.

REFERENCES

1. Almahdawi, H., Akbas, A., & Rahebi, J. (2025). Deep learning neural network based on PSO for leukemia cell disease diagnosis from microscope images. *Journal of Imaging Informatics in Medicine*, 38, 4119–4128. <https://doi.org/10.1007/s10278-025-01474-x>
2. Das, P. K., Meher, S., Rath, A., & Panda, G. (2025). An efficient deep learning system for automatic detection of acute lymphoblastic leukemia. *ISA Transactions*. <https://doi.org/10.1016/j.isatra.2024.12.043>
3. Hasanaath, A. A., Mohammed, A. S., Latif, G., Abdelhamid, S. E., Alghazo, J., & Hussain, A. A. (2024). Acute lymphoblastic leukemia detection using ensemble features from multiple deep CNN models. *Electronic Research Archive*, 32(4), 2407–2423. <https://doi.org/10.3934/era.2024110>
4. Ponnusamy, U., & Perumal, V. (2025). Comprehensive review on learning models of leukemia detection based on morphological information. *Leukemia & Lymphoma*. <https://doi.org/10.1080/10428194.2025.2583449>
5. Sabin Begum, R., Mohamed Anwar, A., & Anu Priya, S. (2025). A health-based deep learning system for rapid and precise detection of acute lymphoblastic leukemia. *Journal of Neonatal Surgery*, 14(4S), 87–94. <https://doi.org/10.52783/jns.v14.1746>
6. Syed, N., Saeed, M. E. S., Hussain, S., Mirza, I., Abdalla, A. M., Al Zaabi, E. A., Afrooz, I., Hashmi, S., & Yaqub, M. (2025). Novel hierarchical deep learning models predict type of leukemia from whole slide microscopic images of peripheral blood. *Journal of Medical Artificial Intelligence*, 8, 5. <https://doi.org/10.21037/jmai-24-74>
7. Liu, Hong, Haichao Cao, and Enmin Song. "Bone marrow cells detection: a technique for the microscopic image analysis." *Journal of medical systems* 43 (2019): 1-14.
8. Ratley, Astha, Jasmine Minj, and Pooja Patre. "Leukemia disease detection and classification using machine learning

- approaches: a review." In 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T), pp. 161-165. IEEE, 2020.
9. Sridhar, K., Ajay Reddy Yeruva, P. N. Renjith, Asmita Dixit, Aatif Jamshed, and Ravi Rastogi. "Enhanced Machine learning algorithms Lightweight Ensemble Classification of Normal versus Leukemic Cel." *Journal of Pharmaceutical Negative Results* (2022): 496-505.
 10. Lewis, Joshua E., and Olga Pozdnyakova. "Digital assessment of peripheral blood and bone marrow aspirate smears." *International Journal of Laboratory Hematology* (2023).
 11. Bibi, Nighat, Misba Sikandar, Ikram Ud Din, Ahmad Almogren, and Sikandar Ali. "IoMT-based automated detection and classification of leukemia using deep learning." *Journal of healthcare engineering* 2020 (2020): 1-12.
 12. Sahlol, Ahmed T., Philip Kollmannsberger, and Ahmed A. Ewees. "Efficient classification of white blood cell leukemia with improved swarm optimization of deep features." *Scientific reports* 10, no. 1 (2020): 2536.
 13. Mohammed, Zemen Mawlood, and Esraa Zeki Mohammed. "Leukemia Blood Cell Image Classification Using Machine Learning-A Systematic Literature Review." *Journal of Advanced Sciences and Nanotechnology* 2, no. 2 (2023): 204-214.
 14. Jiang, Zhencun, Zhengxin Dong, Lingyang Wang, and Wenping Jiang. "Method for diagnosis of acute lymphoblastic leukemia based on ViT-CNN ensemble model." *Computational Intelligence and Neuroscience* 2021 (2021).