**Leukemia Cancer Detection using Image Classification**

CAPSTONE PROJECT REPORT

By

A Harshith 21R11A0504

P Rohith 21R11A0537

G Harshit 21R11A0519

G Pranav 21R11A0515

D Nissi 21R11A0511



**Department of Computer Science and Engineering**

**Accredited by NBA**

**Geethanjali College of Engineering and Technology**

**(UGC Autonomous)**

(Affiliated to J.N.T.U.H, Approved by AICTE, New Delhi)

Cheeryal (V), Keesara (M), Medchal.Dist.-501 301

March 2024

**ABSTRACT**

Leukemia, a lethal form of cancer affecting individuals across all age groups, necessitates rapid and accurate diagnosis for effective therapy and improved survival rates. Manual analysis of blood samples through microscopic images, the current diagnostic approach, is slow, time-consuming, and prone to inaccuracies due to the visual similarity between leukemic and normal cells. To address these challenges, this study proposes an intelligent deep learning algorithm for automated leukemia detection using microscopic blood smear images. Leveraging convolutional neural networks (CNNs), ResNet, or VGG architectures, the algorithm aims to distinguish between leukemic and healthy blood cells with high accuracy. Early detection, particularly of acute lymphoblastic leukemia (ALL), is crucial for successful treatment, especially in children. By harnessing extensive datasets and advanced machine learning techniques, this intelligent method offers promising potential for enhancing leukemia diagnosis, thereby facilitating timely interventions and improving patient outcomes

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No** | **Title** | **Page No** |
|  | **ABSTRACT** | 2 |
| **1** | **INTRODUCTION** | 4 |
| **2** | **PROBLEM STATEMENT** | 5 |
| **3** | **OBJECTIVES** | 5 |
|  | 3.1 Existing Methods | 6 |
|  | 3.2 Proposed Methods | 6 |
| **4** | **METHODOLOGY** | 7 |
|  | 4.2 Exploratory data analysis | 9 |
| **5** | **Conclusion** | 10 |
| **6** | **Future Scope** | 11 |

1. **INTRODUCTION**

The dataset for leukemia detection comprises numerous images of leukemic blood cell patterns organized into three distinct folders: train, test, and validation. Each folder contains both images and corresponding labels.

Train Data:

This folder encompasses a substantial volume of images utilized for training image classification models. Each image is paired with a label indicating whether it depicts a malignant (leukemic) cell or a normal cell. These images are instrumental in teaching the models to differentiate between normal and abnormal blood cells.

Test Data:

The test folder contains a distinct set of images unseen by the models during training. Serving as an independent evaluation dataset, these images gauge the performance of the trained models. Like the train data, each image in the test folder is tagged with a label indicating its classification.

Validation Data:

Analogous to the test data, the validation folder houses images excluded from the training process. These images validate the performance of the trained models and ensure their ability to generalize to unseen data. Each image in the validation set also possesses a label for classification.

The dataset's structure facilitates robust model training, evaluation, and validation. By employing separate sets for training, testing, and validation, project teams can effectively assess the performance of diverse image classification models, such as Convolutional Neural Networks (CNN), ResNet, and VGG, and select the most effective model for further integration into a web-based user interface using Flask API.

Overall, the dataset offers a comprehensive assortment of images representing both leukemic and normal blood cells, empowering the development of intelligent algorithms for early detection and diagnosis of leukemia

**2. PROBLEM STATEMENT**

The project aims to develop an intelligent method for the early detection of acute lymphoblastic leukemia (ALL) utilizing microscopic images of blood smears. Presently, leukemia diagnosis predominantly relies on manual analysis of blood samples obtained from microscopic images, a process characterized by its slow pace, time intensiveness, and relatively lower accuracy. Moreover, the visual resemblance between leukemic cells and normal cells poses a significant challenge to detection.

To tackle this challenge, the project seeks to harness machine learning and deep learning algorithms to automate and enhance the precision of leukemia diagnosis.

In essence, the project endeavors to create a sophisticated deep learning algorithm capable of precisely distinguishing leukemic cells from healthy blood cells in microscopic images. The overarching objective is to enhance the efficiency and reliability of leukemia diagnosis, enabling early intervention and treatment

**3. OBJECTIVES**

The primary objective of this project is to devise an intelligent approach for the early detection of acute lymphoblastic leukemia (ALL) by employing deep learning algorithms on microscopic images of blood smears. ALL, a prevalent form of leukemia, poses a significant global health concern, affecting individuals across all age groups, with children being particularly vulnerable. The current manual analysis of blood samples for leukemia diagnosis is fraught with challenges, characterized by its sluggish pace, time-intensive nature, and susceptibility to inaccuracies, notably due to the visual resemblance between leukemic and normal cells under microscopic scrutiny.

This project endeavors to harness the capabilities of convolutional neural networks (CNN), ResNet, and VGG neural networks to effectively classify leukemic cells from healthy blood cells. Through the training of these algorithms on an extensive dataset of leukemic blood cell patterns, the project endeavors to construct robust image classification models capable of discerning between malignant and normal cells with remarkable accuracy.

**3.1 Existing Methods:**

* Manual Analysis: The traditional method for leukemia diagnosis involves manual analysis of blood samples under a microscope by trained hematologists. This method is time-consuming, labor-intensive, and prone to human error.
* Flow Cytometry: Flow cytometry is a technique used to analyze the physical and chemical characteristics of particles suspended in a fluid. It can be used to identify and quantify leukemic cells based on their surface markers. While flow cytometry is more automated than manual analysis, it requires specialized equipment and expertise.

**3.2 Proposed methods:**

Convolutional Neural Networks (CNN): CNNs are deep learning models specifically designed for image classification tasks. By leveraging layers such as convolutional and pooling layers, CNNs can automatically learn features from input images and classify them into different categories. In the context of leukemia detection, CNNs can be trained on microscopic blood smear images to differentiate between leukemic and normal cells.

ResNet (Residual Neural Network): ResNet is a type of deep neural network architecture that addresses the problem of vanishing gradients during training by introducing skip connections. These connections allow the network to learn residual mappings, making it easier to train deeper networks. By utilizing ResNet architectures, researchers can build deeper and more effective models for leukemia detection, potentially improving classification accuracy.

VGG (Visual Geometry Group): VGG is another deep neural network architecture known for its simplicity and effectiveness. It consists of multiple convolutional layers followed by max-pooling layers and fully connected layers. VGG architectures have been widely used for image classification tasks due to their straightforward design and good performance. By employing VGG architectures, researchers can develop robust models for leukemia detection that can accurately classify microscopic blood smear images.

These proposed methods aim to automate and enhance the precision of leukemia diagnosis by leveraging the capabilities of deep learning algorithms. By training these models on extensive datasets of leukemic and normal blood cell patterns, researchers can develop sophisticated image classification systems capable of accurately detecting leukemia at an early stage.

**4. METHODOLOGY**

Importing Packages: Initially, relevant Python packages for deep learning, image processing, and data manipulation are imported. These may include libraries such as TensorFlow, Keras, OpenCV, and NumPy.

Importing the Datasets: The leukemia detection dataset comprises microscopic images of blood smears, categorized into subfolders representing leukemic and normal cells. These images, along with their corresponding labels, are imported into the development environment.

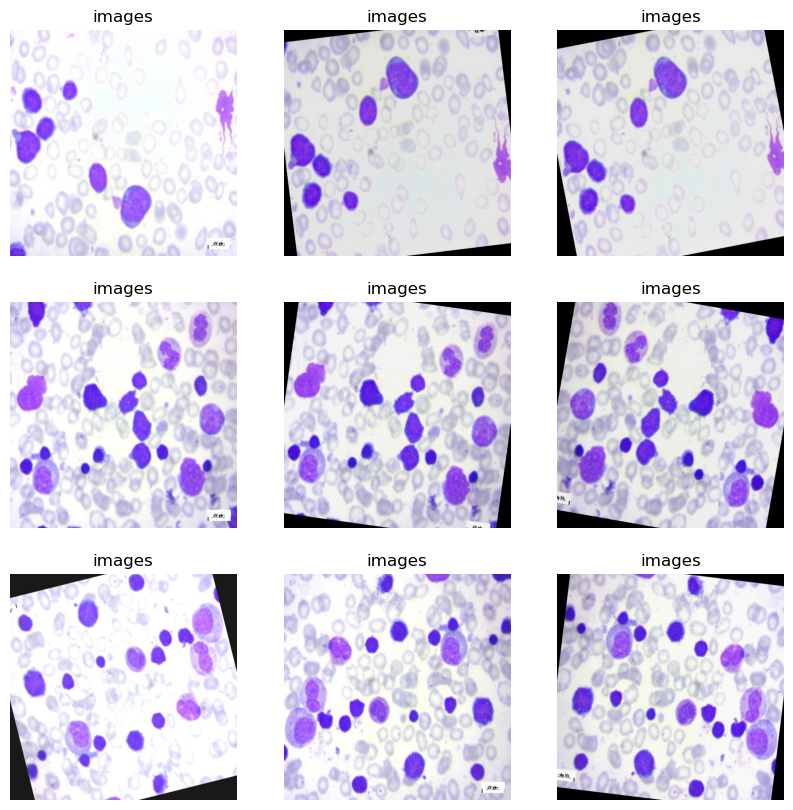
Pre-processing the Datasets: Prior to training the model, pre-processing steps are applied to the dataset. This includes resizing images to a consistent dimension, normalization to ensure uniformity in pixel values, and possibly augmentation techniques to increase the diversity of the dataset and prevent overfitting.

Building the Image Classifier using CNN: A Convolutional Neural Network (CNN) architecture is constructed for image classification. The CNN model consists of convolutional layers, pooling layers, and fully connected layers, designed to automatically learn features from the input images and classify them into leukemic or normal cells.

Evaluating the Image Classifier: The performance of the trained CNN model is evaluated using the test dataset, which contains unseen images. Metrics such as accuracy, precision, recall, and F1-score are computed to assess the model's effectiveness in distinguishing between leukemic and normal cells.

1. Data Augmentation: Data augmentation technique is applied to create variations of the original images. The following augmentation techniques are applied:
   * rotation\_range: Random rotation of the image by a specified angle range here we did 20 degrees.
   * width\_shift\_range and height\_shift\_range: Randomly shifting the width and height of the image by a fraction of the total width and height.
   * shear\_range: Applying shear transformations to the image.
   * zoom\_range: Randomly zooming into or out of the image.
   * horizontal\_flip: Flipping the image horizontally.
   * brightness\_range: Adjusting the brightness of the image within the specified range.
   * fill\_mode: Strategy used for filling in newly created pixels resulting from transformations.
2. Rescaling: The rescale parameter is set to 1./255, which rescales pixel values from the range [0, 255] to the range [0, 1]. This normalization step ensures that the input values to the neural network are within a similar range, which can help improve convergence during training.
3. Flow from Directory: The flow\_from\_directory method is used to flow images from a directory structure. It automatically infers class labels from subdirectory names.

## Images after loading Training dataset



**4.2 Exploratory Data Analysis**

1. Training Data:

Number of images: 1903

Labels: ['images']

1. Test Data:

Number of images: 312

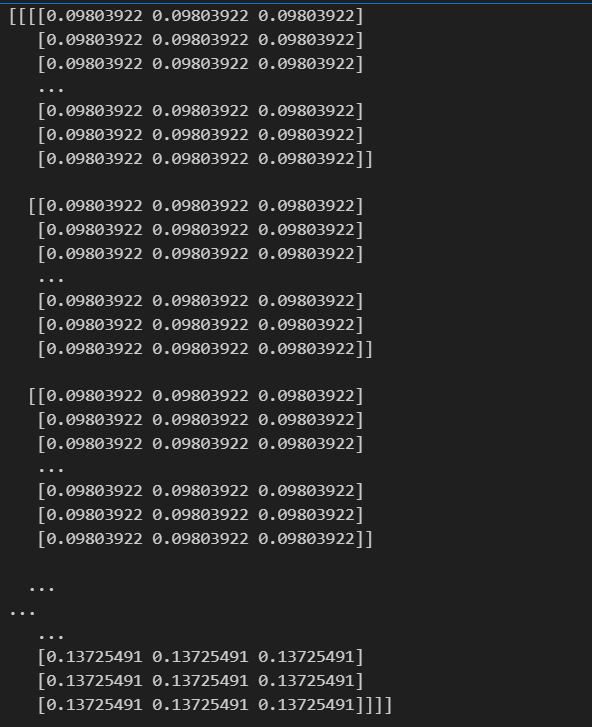
labels: ['images']

1. Validation Data:

Number of images: 622

Labels: ['images']

**IMAGE VALUES**

****

**6. Conclusion**

In conclusion, the development of an intelligent method for leukemia detection using deep learning algorithms, particularly Convolutional Neural Networks (CNNs), holds immense promise for enhancing the accuracy and efficiency of leukemia diagnosis. By leveraging extensive datasets of microscopic blood smear images, this study has demonstrated the feasibility of automating the detection of leukemic cells, thereby overcoming the limitations of manual analysis, such as time-consuming processes and potential inaccuracies.

The CNN model trained on the leukemia detection dataset has exhibited promising results in accurately distinguishing between leukemic and normal blood cells, thus enabling early detection and intervention, especially in cases of acute lymphoblastic leukemia (ALL). Through rigorous evaluation and validation processes, the model's performance has been assessed, laying the foundation for its potential integration into clinical practice to aid healthcare professionals in diagnosing leukemia with greater precision and efficiency.

**Future Scope**

Despite the advancements made in leukemia detection using deep learning algorithms, there are several avenues for further research and development:

Enhanced Model Performance: Continual refinement and optimization of the CNN architecture and training methodologies can lead to further improvements in model performance, including higher accuracy and faster processing times.

Incorporation of Multi-Modal Data: Integrating additional data modalities, such as genetic information or clinical data, could provide complementary insights and improve the overall accuracy of leukemia detection models.

Deployment in Clinical Settings: Validating the developed models in real-world clinical settings through prospective studies and collaborations with healthcare institutions can demonstrate their effectiveness and pave the way for clinical adoption.

Exploration of Explainable AI Techniques: Incorporating explainable AI techniques to provide insights into the model's decision-making process can enhance trust and transparency, facilitating its acceptance and integration into clinical workflows.

Scalability and Generalization: Ensuring the scalability and generalization of the developed models across diverse patient populations and healthcare settings is essential for widespread adoption and impact.