Blood Cell Cancer Detection

Prepared by: Name: Abeer Fatima
Roll Number: AI-292008

Project Title: Blood Cell Cancer Detection

Abstract

The Blood Cell Cancer Detection project leverages an image-based deep learning model to classify blood cell cancer into four categories: **Benign**, **[Malignant] Pro-B**, **[Malignant] Pre-B**, and **[Malignant] early Pre-B**. A Convolutional Neural Network (CNN) was employed to achieve accurate and efficient predictions. Extensive preprocessing, exploratory data analysis (EDA), and normalization techniques were applied to ensure optimal performance. The model demonstrated high accuracy during training and testing phases and was subsequently deployed using Streamlit for user interaction.

Methodology

Data Preparation

1. **Exploratory Data Analysis (EDA):**
    - Visual inspection of dataset.
    - Identification of class distribution.
2. **Image Resizing:**
    - Standardized all images to a uniform size to ensure consistency during training.
3. **Class Mapping:**
    - Applied mapping techniques to assign appropriate labels to each class.
4. **Normalization:**
    - Normalized pixel values to range between 0 and 1 for computational efficiency.

Model Architecture

A CNN model was designed with the following specifications:

- **Input Layer:** Preprocessed image data.
- **Convolutional Layers:** Extracted features using filters.
- **Pooling Layers:** Reduced dimensionality while preserving essential features.
- **Fully Connected Layers:** Combined features for classification.
- **Output Layer:** Produced probabilities for the four classes.

Training and Validation

- The dataset was split into training, validation, and testing subsets.
- The model was trained over 10 epochs with a batch size of 32.
- The Adam optimizer and categorical cross-entropy loss function were used.

Performance Metrics

- **Accuracy** and **loss** were tracked for training, validation, and testing phases.
- Random images were uploaded and tested to ensure robustness.

Results

Training Performance

Epoch	Accuracy	Loss	Validation Accuracy	Validation Loss
1	0.4431	4.0125	0.7091	0.7205
2	0.9231	0.2135	0.8728	0.3237
3	0.9760	0.0956	0.9483	0.2207
4	0.9341	0.1879	0.7996	0.6872
5	0.9884	0.0437	0.8642	0.4142
6	0.9899	0.0429	0.8405	0.5894

Final Metrics

Metric	Value
Training Accuracy	96.45%
Training Loss	0.0925
Validation Accuracy	94.83%
Validation Loss	0.2207
Test Accuracy	94.17%
Test Loss	0.2577

Deployment

- The trained model was saved as an **H5** file.
- The model was deployed on **VS Code** using **Streamlit**.
- Users can upload random images for prediction through a user-friendly interface.
- The model provided accurate predictions for all test cases.

Conclusion

The Blood Cell Cancer Detection model achieved high accuracy and demonstrated excellent generalization capabilities. It is a robust tool for early detection and classification of blood cell cancer types, which can aid in medical diagnosis and treatment planning. The successful deployment using Streamlit makes the model accessible for real-world applications.

Future Enhancements

- Increase the dataset size for better generalization.
- Fine-tune the model hyperparameters for improved performance.
- Integrate additional image augmentation techniques.
- Develop a mobile application for wider accessibility.
