Project Title: Leukemia Classification with Deep Learning

Project Overview:

The project aims to classify leukemia images using three different deep learning models:

1. **Basic CNN (Convolutional Neural Network)**
2. **VGG16** (Transfer Learning from a pre-trained model)
3. **ResNet50** (Another transfer learning model)

The goal is to determine the best-performing model for leukemia classification based on accuracy and other performance metrics.

Dataset Handling:

- **Dataset Source:** The dataset was loaded from a specified Google Drive directory, containing various classes of leukemia images.
- **Dataset Split:** The data was split into training, validation, and test sets using a 60-20-20 ratio.
- **Data Augmentation:** Augmentation techniques such as rotation, shifts, shear, zoom, and horizontal flipping were applied to enhance generalization.

Models Trained:

1. **CNN (Convolutional Neural Network):**
    - A simple CNN architecture with three convolutional layers, followed by max-pooling, and fully connected layers.
    - Optimized using the Adam optimizer and regularized with dropout.
    - Trained for 10 epochs with early stopping.
2. **VGG16 (Transfer Learning):**
    - A pre-trained VGG16 model was fine-tuned for this task. Initially, the convolutional layers were frozen, and only the classifier was trained.
    - Later, additional layers were unfrozen to improve performance via fine-tuning.
    - Optimized using the Adam optimizer.
3. **ResNet50 (Transfer Learning):**
    - A pre-trained ResNet50 model, initially frozen, and then fine-tuned for classification.
    - Similar to VGG16, this model uses a combination of frozen and trainable layers to leverage learned features.

Evaluation Metrics:

- **Test Accuracy:** The primary metric to assess model performance on unseen data.
- **Confusion Matrix and Classification Report:** To provide detailed insights into classification errors and evaluate precision, recall, and F1-scores.

Results:

The following test accuracies were recorded for the three models:

- **CNN Model**: 61.67%
- **VGG16 Model**: 81.25%
- **ResNet50 Model**: 18.75%

The **VGG16 model** achieved the highest accuracy of **81.25%**, outperforming the CNN and ResNet50 models.

Analysis:

1. **CNN Model:** Achieved a moderate test accuracy of 61.67%. The relatively simple architecture may have struggled with the complexity of the leukemia dataset.
2. **VGG16 Model:** The best-performing model with 81.25% accuracy. Transfer learning from a model pre-trained on ImageNet allowed for effective feature extraction, making it well-suited for this task.
3. **ResNet50 Model:** Underperformed with an accuracy of only 18.75%, likely due to over-complexity and insufficient fine-tuning for this specific dataset.

Conclusion:

The **VGG16 model** with transfer learning is the optimal choice for the leukemia classification task, with the highest test accuracy of **81.25%**. While the CNN model had moderate success, ResNet50 struggled to generalize effectively.

Recommendations:

1. **Further Fine-tuning:** Additional tuning of the VGG16 model could push performance even higher.
2. **Augmentation & Dataset Expansion:** More diverse data augmentation or expanding the dataset could improve the CNN and ResNet50 models.
3. **Hyperparameter Optimization:** Testing different optimizers, learning rates, and regularization techniques might enhance performance across all models.

This report summarizes the results of leukemia image classification using deep learning models, highlighting the **VGG16** model as the top performer.