Lung and Colon Cancer Detection

Project Overview

- **Objective:**
  - Detect lung and colon cancer using advanced image classification techniques.

- **Dataset:**
  - Source: Kaggle
  - Total Images: 25,000 colored images
  - Classes:
    1. Colon adenocarcinoma
    2. Colon benign tissue
    3. Lung adenocarcinoma
    4. Lung benign tissue
    5. Lung squamous cell carcinoma

Project Overview

- **Approach:**
  - Developed and evaluated four models using CNN and DenseNet architectures.
- **Team Composition:**

Ahmed Said Emam	CNN ‘Gray Images’
Mohammed Ahmed	CNN ‘Colored Images’
Ahmed Mohammed Elsayed	DenseNet ‘Gray Images’
Mohammed Ahmed Ismail	DenseNet ‘Colored Image’
Ahmed Maher	MLOps
Mohammed Abdalkader	Deployment ‘Flask’

Introduction

- **Background:**
  - Cancer remain a leading cause of mortality globally, with lung and colon cancer being among the most common.
- **Lung Cancer Statistics:**
  - Leading cause of cancer deaths worldwide, approximately 1.8 million deaths in 2020.
  - 5-year survival rate is around .
- **Colon Cancer Statistics:**
  - Third most common cancer globally, with over 1.9 million new cases in 2020.
  - 5-year survival rate is approximately
- **Importance of the Project:**
  - Leveraging machine learning and deep learning for early detection.
  - Utilize image data from scans to develop robust models for classifying cancer types.
  - Enhancing diagnostic accuracy and reducing the burden on healthcare systems.

Model Architectures

- **Models Developed:**
  - **Convolution Neural Network:**
    - Grayscale images.
    - Colored images.
  - **DenseNet pre-trained:**
    - Grayscale (Trained from scratch).
    - Colored (Fine-tuned pre-trained model).

- **Input Shape:**
  - **200x200** pixels for DenseNet models.
  - **200x200** pixels for CNN models

Model 1: CNN on Colored Images

- **Architecture:**
  - Convolution layers, pooling, ReLU activation, full connected layers, softmax for output.

Model: "sequential"

Layer (type)	Output Shape
conv2d (Conv2D)	(None, 200, 200, 32)
max_pooling2d (MaxPooling2D)	(None, 100, 100, 32)
conv2d_1 (Conv2D)	(None, 100, 100, 64)
max_pooling2d_1 (MaxPooling2D)	(None, 50, 50, 64)
conv2d_2 (Conv2D)	(None, 50, 50, 128)
max_pooling2d_2 (MaxPooling2D)	(None, 25, 25, 128)
conv2d_3 (Conv2D)	(None, 25, 25, 256)
max_pooling2d_3 (MaxPooling2D)	(None, 12, 12, 256)
conv2d_4 (Conv2D)	(None, 12, 12, 512)
max_pooling2d_4 (MaxPooling2D)	(None, 6, 6, 512)
flatten (Flatten)	(None, 18432)
dense (Dense)	(None, 256)
dense_1 (Dense)	(None, 5)

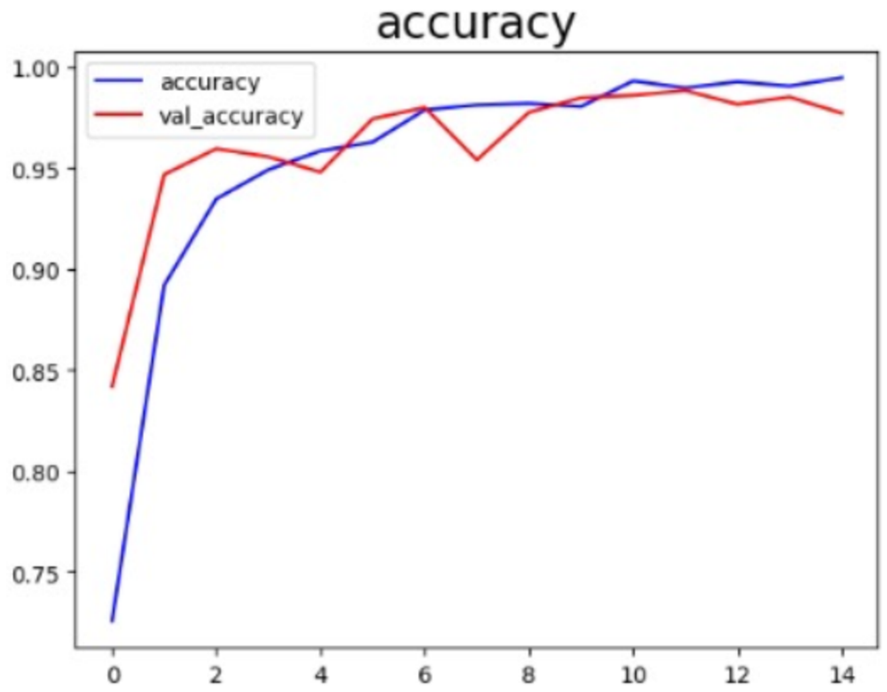
Total params: 6,288,709 (23.99 MB)

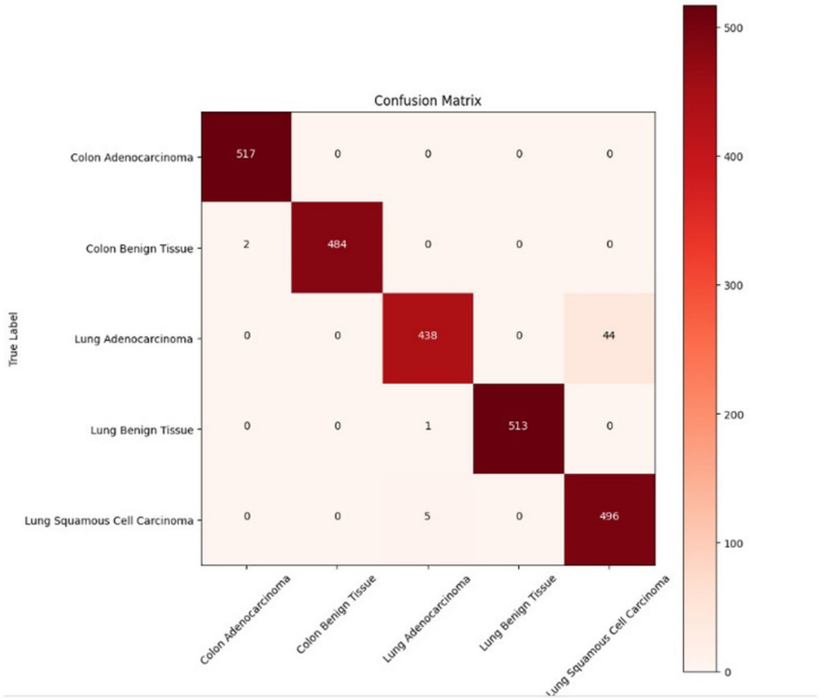
Trainable params: 6,288,709 (23.99 MB)

Non-trainable params: 0 (0.00 B)

Model 1: CNN on Colored Images

- **Performance Metrics:**
  - Accuracy: 98% on test data.





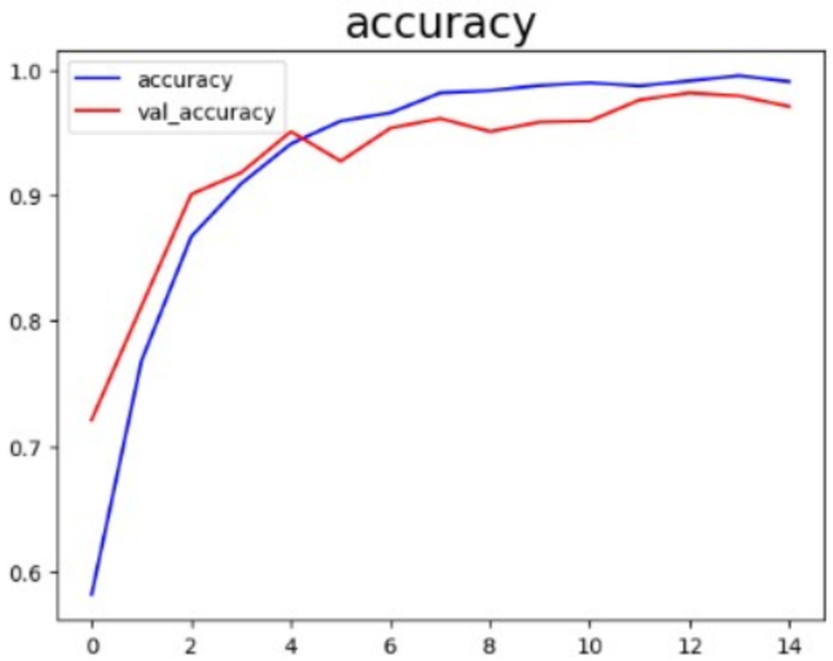
Model 2: CNN on Grayscale Images

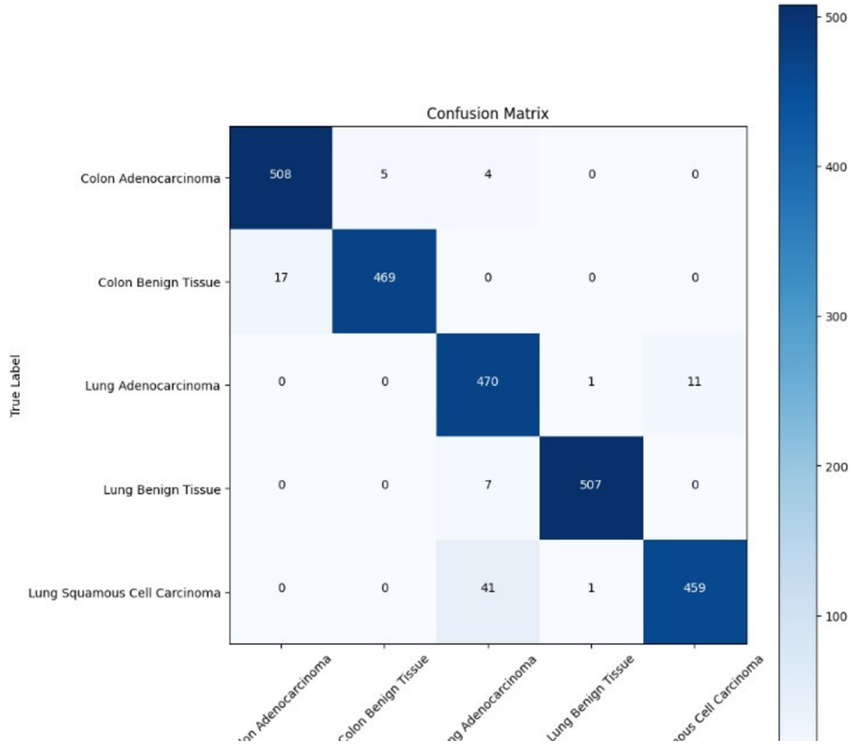
- **Architecture:**
  - Similar to Model 1 but adapted for grayscale images.
  - Grayscale input (derived from colored images).

- **Key Difference:**
  - Processing grayscale channels in input layers.

Model 2: CNN on Grayscale Images

- **Performance Metrics:**
  - Accuracy: 96% on test data.





DenseNet Model on Colored Images

- **Architecture:**
  - DenseNet pre-trained model (imageNet) fine-tuned for colored data.

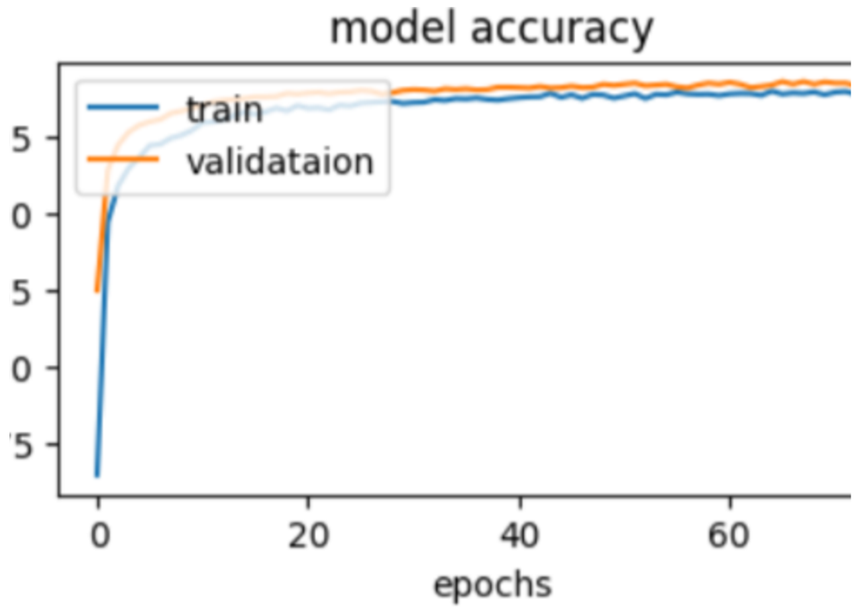
- **Preprocessing:**
  - Images resized to pixels
  - Rescaling: 1./255

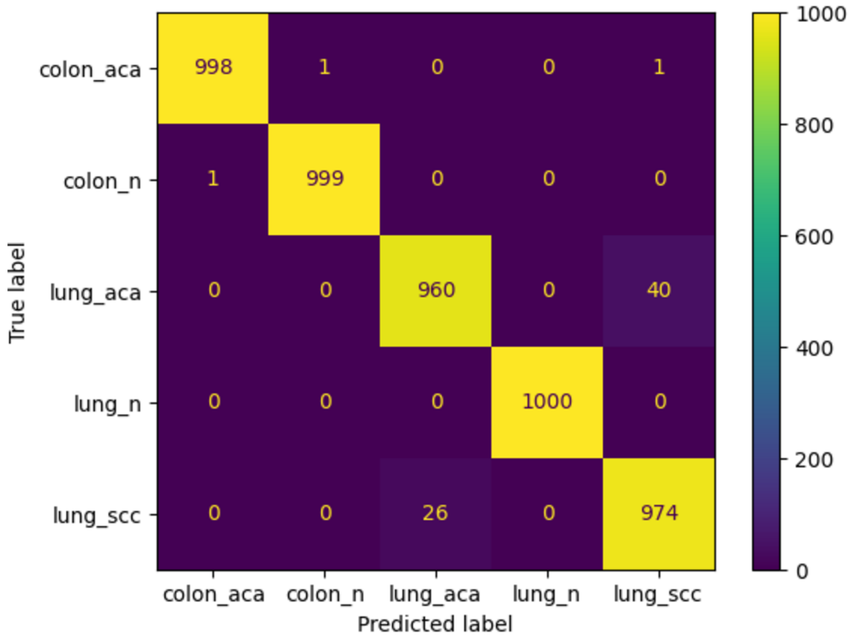
- **Optimizer:** Adamax (learning rate = 0.001)

- **Epochs:** 100 (Early Stopping applied)

DenseNet Model on Colored Images

- **Performance Metrics:**
  - Accuracy: 98% on test data.





DenseNet Model on Grayscale Images

- **Architecture:**
  - DenseNet pre-trained model trained from scratch on grayscale data.

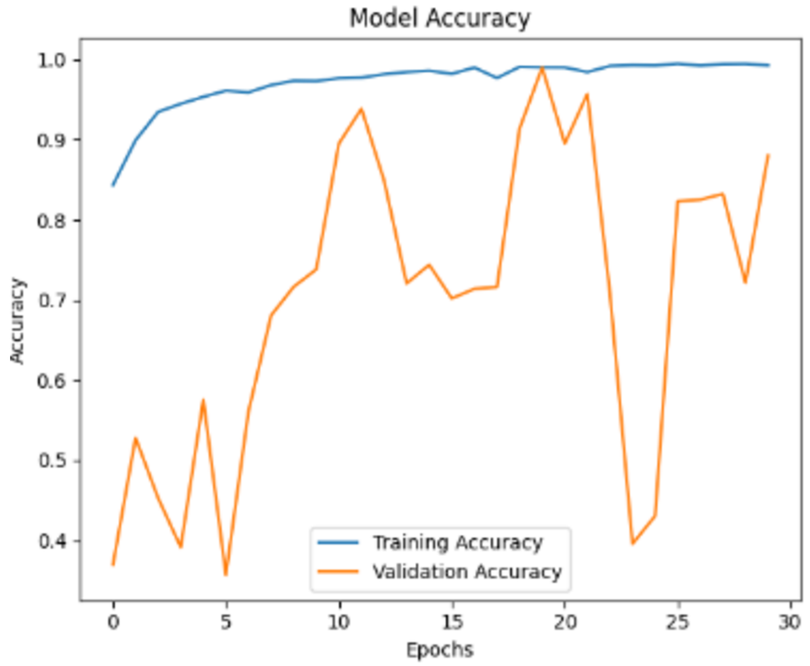
- **Preprocessing:**
  - Image resized to pixels
  - Rescaling: 1./255

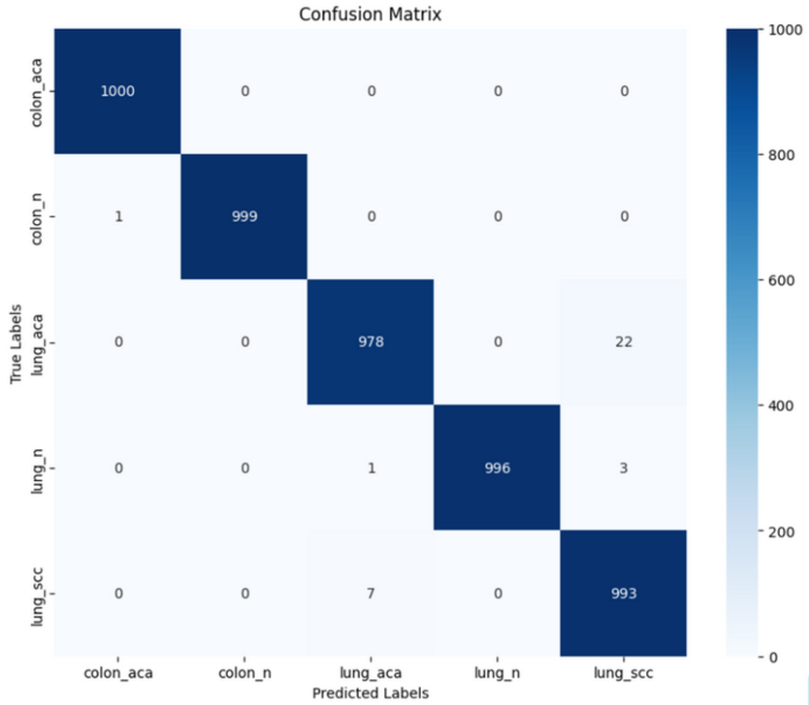
- **Optimizer:** Adamax (Learning rate = 0.001)

- **Epochs:** 100 (Early Stopping applied)

DenseNet Model on Grayscale Images

- **Performance Metrics:**
  - Accuracy: 99% on test data.





Model Comparison

- **Metric for Comparison:**
  - Accuracy, precision, recall

Metric	CNN (gray)	CNN (colored)	DenseNet (gray)	DenseNet (colored)
Accuracy	96%	98%	99%	98%
Precision	90%	97%	98%	96%
Recall	99%	98%	99%	97%

  - **Best Model:** DenseNet on grayscale images.
    - **Accuracy: 99%** on test data

Deployment

- **Tools Used:**
  - **Flask** for deployment
  - **MLOps Process:** Tested sample cancer images on each model to determine the best performance.

- **Result:**
  - DenseNet (grayscale) selected for deployment due to highest accuracy 99%

Challenged & Solutions

- Challenges:
  - Computational resources: Managing model training time.
  - Preprocessing different image sizes: Handled with rescaling for each model.

- **Solutions:**
  - Utilized pre-trained model for efficiency.
  - Optimized models using Early Stopping.

Future Work

- Expand dataset to include more cancer types.
- Investigate additional transfer learning models (e.g., EfficientNet).
- Implement cloud-based deployment for better scalability.

Conclusion

- **Summary**: successfully built a model for lung and colon cancer detection. DenseNet (Grayscale) provided the best accuracy at .
- **Takeaway:** Pre-trained models (DenseNet) are highly effective for cancer detection, especially when paired with grayscale images.


Thank You