Lung Cancer Prediction using CT medical imaging

- *Nisarg Patel (1225464851)*

Introduction

INTRODUCTION

- As lung cancer is among the top three cancers worldwide, its early-stage detection and prediction play the most crucial role in cancer prevention.
- In this project, we proposed an efficient model that accurately predicts whether a patient is suffering from lung disease or not, based on a CT scan image.
- However, for any AI-based system to predict cancer accurately, it should get access to a reasonable amount of patients' digital data such as low-dose computed tomography (LDCT) and computed tomography (CT).
- The authors have implemented a specific type of image augmentation and pre-processing on a relatively small CT image dataset and have trained ensembled Convolutional Neural Network (CNN) models that performs better.

OUR GOAL

- Dataset Accessibility
- Model Performance

Literature Review

Literature Review

- A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification
  - KNN → 80% accuracy on 813 random samples
- Lung cancer screening with low-dose CT scans using a deep learning approach
  - DeepScreener → 78.2% accuracy (trained on 3000 samples)
- Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN)
  - U-Net → 86.6% accuracy , AUC → 0.83
- Automatic Lung Cancer Prediction for Chest X-ray Images Using the Deep Learning Approach
  - DenseNet-121 → 74.43% accuracy

Dataset Description

Dataset Description

- 1000 CT scans images with size (224,224)
- 4 categories (1 normal CT scan, 3 category of lung cancer)

Category	Training	Validation	Testing	
Adenocarcinoma	195	23	120	**338**
Large Cell	115	21	51	**187**
Squamous Cell	155	15	90	**260**
Normal	148	13	54	**215**
	613	**72**	**315**	**1000**

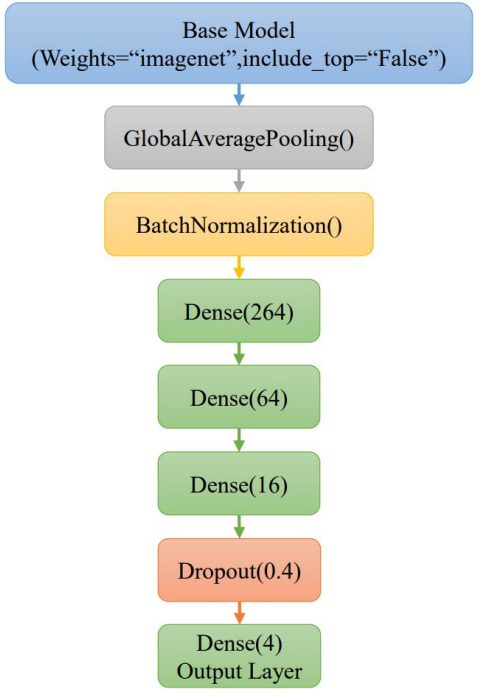
Pre-Processing

Pre-Processing

- ImageDataGenerator `from tensorflow.keras.preprocessing.image import ImageDataGenerator`
  - Pixel Scaling: rescale = 1. /255
  - Image Augmentation:
    - Rotation Range: 15
    - Shear Range: 0.2
    - Zoom Range: 0.2
    - Horizontal flip: True
    - Width shift range: 0.1
    - Height shift range: 0.1
    - Fill Method: ‘Nearest’

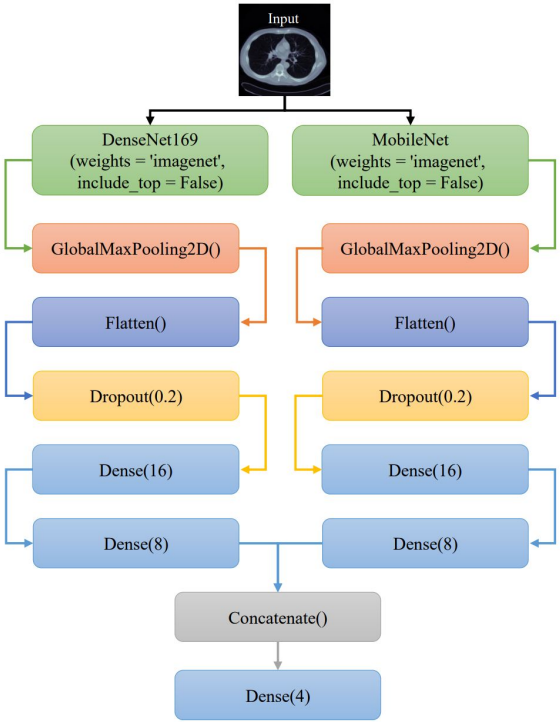
EXPERIMENTS

Base Model & performance



Model	Loss	Accuracy	Recall	Precision
DenseNet169	**0.547894**	**0.853968**	0.93968	0.45963
DenseNet201	0.759781	0.765079	0.87619	0.57381
EfficientNetB4	0.995534	0.55873	0.82857	0.4579
InceptionV3	0.624704	0.793651	**0.94286**	0.44461
MobileNet	0.559497	0.825397	0.93651	**0.65121**
ResNet50	1.123878	0.514286	0.6	0.32643
VGG19	0.706449	0.72381	0.77778	0.58612
Xception	0.828387	0.669841	0.75238	0.59102

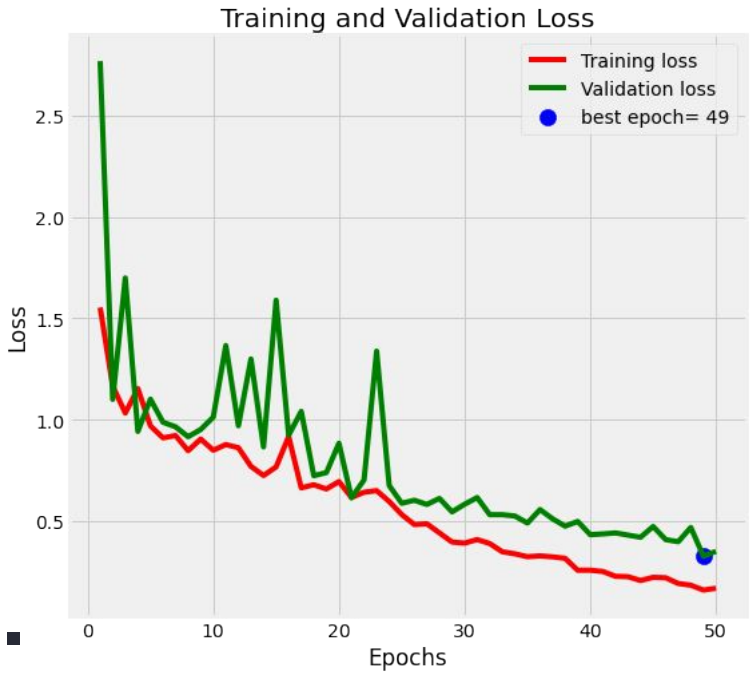
Proposed models & environment setting

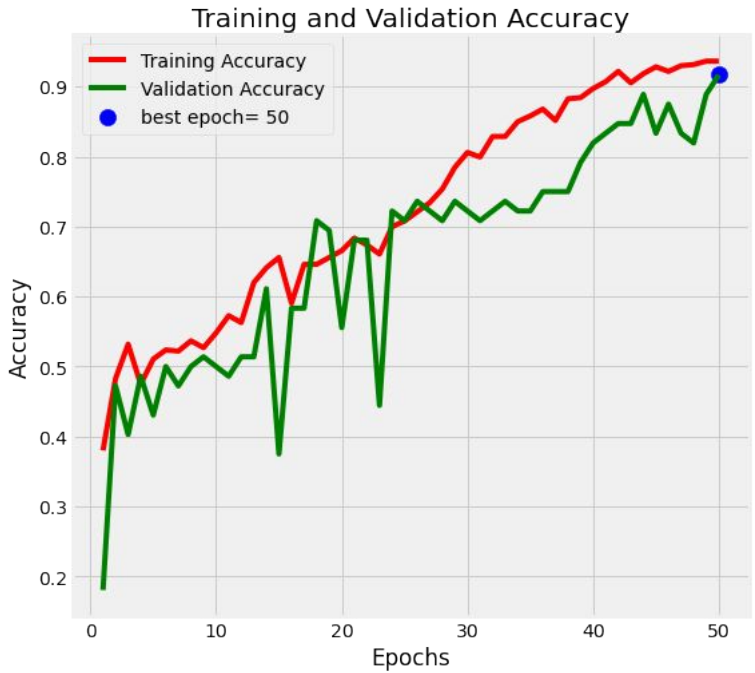


Parameters	Values
Programming Language	Python 3.8.0
Platform	Google Colab
Framework	TensorFlow
Batch size	8
Epochs	50
Optimizer	Adam
Loss	Categorical cross-entropy
Metrics	Accuracy, Recall, Precision
ModelCheckpoint	monitor = Validation Accuracy

RESULT & ANALYSIS

Result & Analysis





Result & Analysis

Model	Loss	Accuracy	Recall	Precision
DenseNet169	0.547894	0.853968	0.93968	0.45963
MobileNet	0.559497	0.825397	0.93651	0.65121
InceptionV3	0.624704	0.793651	0.94286	0.44461
Proposed	**0.344487**	**0.873015**	**1**	0.58558

CONCLUSION

- Solved two major challenges
  - Small dataset
  - Better performance

- Detection of 3 lung cancer's categories with a highest recall on a test (hold out) dataset through proposed model.

Thank You!

Any questions?