Adaovi Project

Domain: Data Science

Batch: March, 2024.

Project Title: Lung Cancer Detection

Submitted by

ATCHAYAH S

GOVERNMENT COLLEGE OF ENGINEERING,

TIRUNELVELI.

PROJECT DESCRIPTION:

In healthcare, early detection and accurate diagnosis of diseases are crucial for effective treatment and improved patient outcomes. However, the sheer volume and complexity of medical data make it challenging for healthcare professionals to analyse and interpret all relevant information accurately.

PROPOSED SOLLUTION:

The objective of this project is to develop a Convolutional Neural Network (CNN) model to accurately classify lung cancer images into four categories: squamous cell carcinoma, normal, large cell carcinoma, and adenocarcinoma. The model should achieve a high validation accuracy and generalize well to new, unseen data.

DATA COLLECTION:

Collected the chest CT scan images dataset from the website "kaggle.com". The images are in jpg or png to fit the model. Data contain 3 chest cancer types which are Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and 1 folder for the normal cell

Data folder is the main folder that contain all the step folders inside Data folder are test, train, valid. Training set is **70%**, testing set is **20%**, validation set is **10%**.

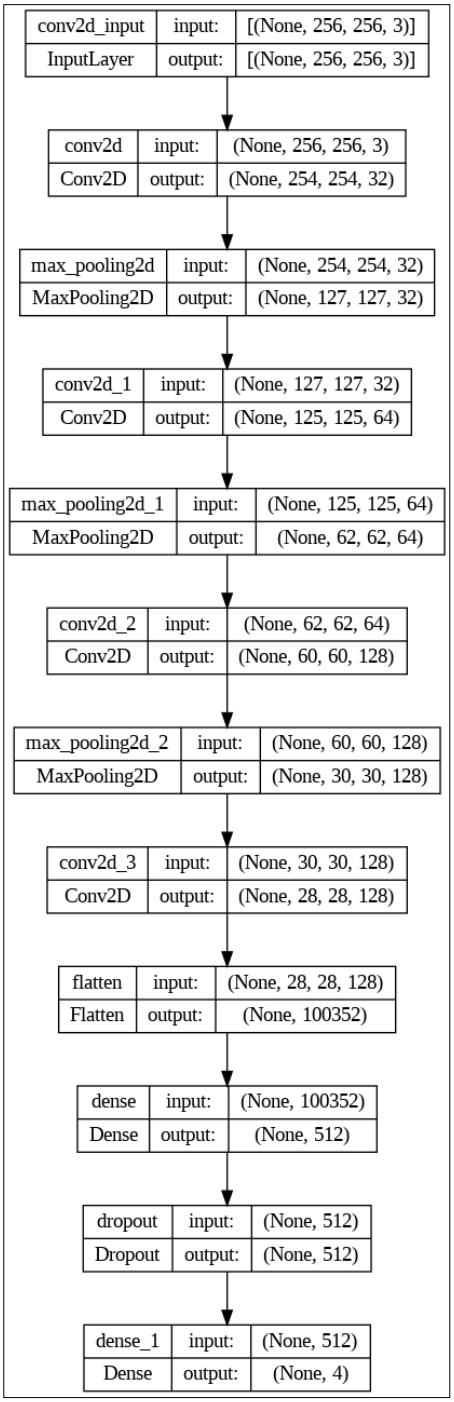
DATA AUGMENTATION:

Use data augmentation techniques to artificially expand the dataset and improve the model's generalization.

Implement random rotations, width and height shifts, shear transformations, zoom, and horizontal flips.

```
datagen = ImageDataGenerator(
    rotation_range=40,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)
datagen.fit(train)
```

MODEL ARCHITECTURE



MODEL PERFORMANCE EVALUATION

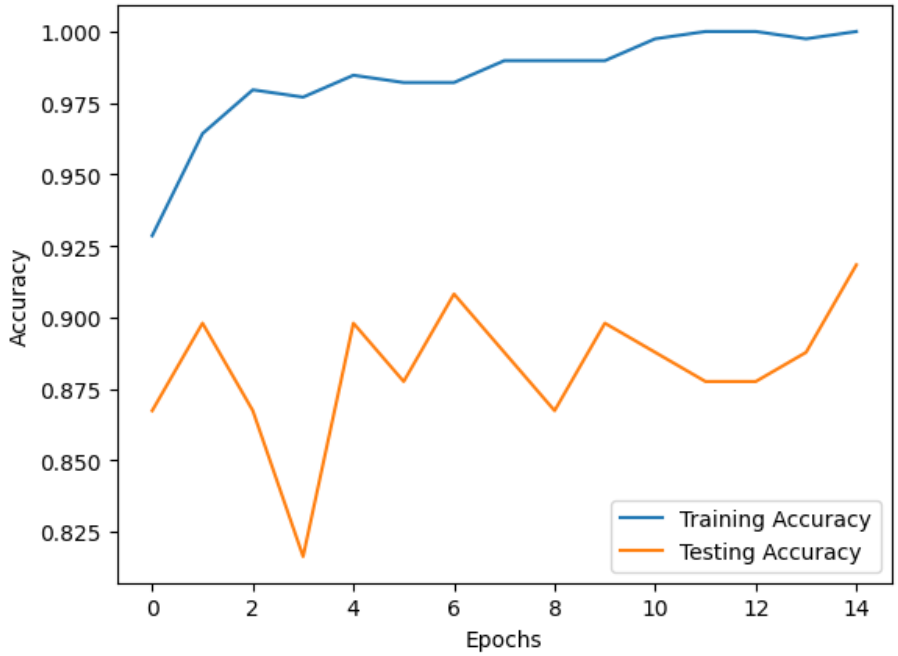
Training and Validation Accuracy:

Training Accuracy: 1.0000 (100%)

Validation Accuracy: 0.9184 (91.84%)

The training accuracy is very high, which indicates that the model has learned the training data very well. However, a perfect training accuracy can sometimes indicate overfitting, especially

when there is a significant gap between training and validation accuracy. In this case, the validation accuracy is also high (91.84%), which suggests that the model generalizes well to unseen data to a certain extent.



Classification Report:

```
3/3 [==============================] - 3s 768ms/step
                         precision    recall  f1-score   support

squamous.cell.carcinoma       0.68      0.87      0.76        15
                 normal       1.00      0.92      0.96        13
   large.cell.carcinoma       0.89      0.76      0.82        21
         adenocarcinoma       0.74      0.74      0.74        23

               accuracy                           0.81        72
              macro avg       0.83      0.82      0.82        72
           weighted avg       0.82      0.81      0.81        72

[[13  0  0  2]
 [ 0 12  1  0]
 [ 1  0 16  4]
 [ 5  0  1 17]]
```

Overall Accuracy: 0.81 (81%)

Class-wise Precision, Recall, and F1-score:

Squamous.cell.carcinoma: Precision 0.68, Recall 0.87, F1-score 0.76

Normal: Precision 1.00, Recall 0.92, F1-score 0.96

Large.cell.carcinoma: Precision 0.89, Recall 0.76, F1-score 0.82

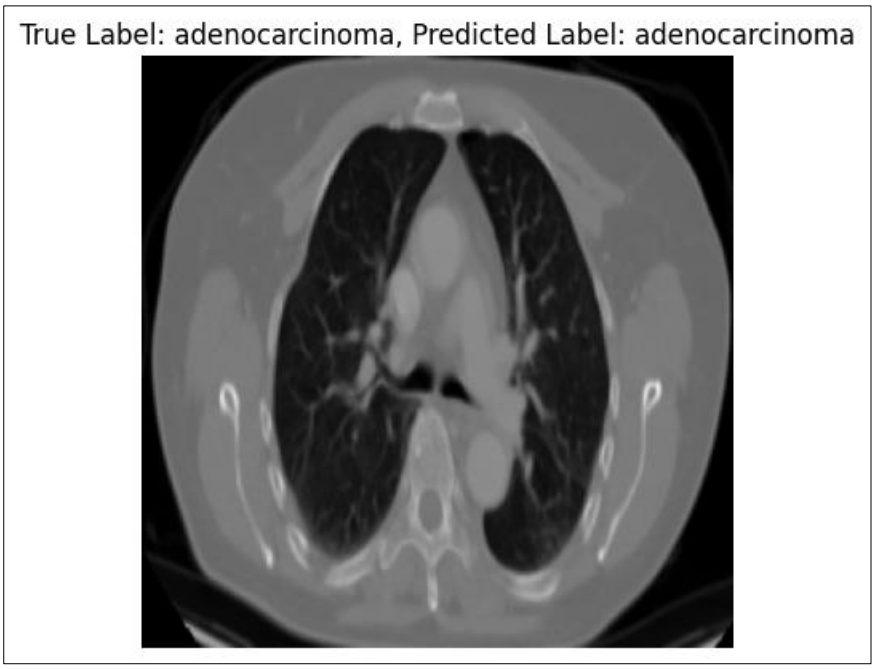
Adenocarcinoma: Precision 0.74, Recall 0.74, F1-score 0.74

The overall performance is solid.

Confusion Matrix:

The confusion matrix shows where the model is making mistakes. The diagonal elements represent correct predictions, and off-diagonal elements represent misclassifications. This model has a reasonable number of correct predictions across all classes.

MODEL PREDICTION



This sample image shows that the model performs well and predict the classes correctly.

Strengths:

High Validation Accuracy: A validation accuracy of 91.84% is quite good.

Balanced Performance: The model performs reasonably well across different classes.

Detailed Evaluation: The inclusion of precision, recall, and F1-score for each class demonstrates a thorough evaluation.

Weaknesses:

Overfitting Potential: The perfect training accuracy suggests potential overfitting, although the high validation accuracy mitigates this concern to some extent.

Class Imbalance: There are some issues with class imbalance that might need addressing to improve minority class performance further.

CONCLUSION:

The results showcase a successful implementation of a CNN for medical image classification, highlighting the potential of deep learning in healthcare diagnostics.

Project Code:

https://github.com/Atchayah-S/LungCancerDetection_Adaovi/blob/main/ATCHAYAH_S_AdaoviProject%20Submission.ipynb

Project Repository:

https://github.com/Atchayah-S/LungCancerDetection_Adaovi