Final Project Report

CovidVision: Advanced COVID-19 Detection from Lung X-rays with Deep Learning

Team ID : SWTID1720351492

Team Leader : Ishan Singh

Team member : Garvit Pathak

Team member : Dhaanya Yelleti

Team member : Anushka Deshmukh

1. Introduction

1.1. The COVID-19 pandemic has highlighted the need for rapid and accurate diagnostic tools to identify infected individuals. Chest X-rays are a widely available and cost-effective imaging modality that can provide valuable information for COVID-19 diagnosis.

1.2. This project aims to develop a deep learning model to analyze chest X-rays and classify them as either COVID-19 positive or normal. The model will be trained on a large dataset of chest X-rays from patients with confirmed COVID-19 and normal controls.

1.3. The model is deployed at HuggingFace Spaces. It can be accessed here

2. Project Overview

Objectives:

- To develop a deep learning model for the classification of COVID-19 from chest X-rays.
- To evaluate the performance of the model on a large dataset of chest X-rays.
- To make the model available to healthcare providers as a tool for COVID-19 diagnosis.

3. Project Initialization and Planning Phase

3.1. Define Problem Statement

Customer Problem Statement:

- Healthcare providers are facing challenges in diagnosing COVID-19 due to limited resources, high patient volumes, and a lack of expert radiologists.
- Public health authorities need accurate data to monitor COVID-19 trends and allocate resources effectively.

3.2. Project Proposal (Proposed Solution)

- Develop a deep learning model to analyze chest X-rays and classify them as either COVID-19 positive or normal.
- Integrate the model with hospital systems for expedited diagnosis.
- Deploy the model in rural clinics for enhanced screening.
- Utilize the model for real-time data analysis for public health monitoring and resource allocation.

3.3. Initial Project Planning

- **Sprint 1:** Registration, login, and data collection.
- **Sprint 2:** Image preprocessing, model development, and validation.
- **Sprint 3:** Model optimization and tuning.
- **Sprint 4:** Deployment, documentation, and user training.

4. Data Collection and Preprocessing Phase

4.1. Data Collection Plan and Raw Data Sources Identified

- **Data Source:** COVID-19 Radiography Database
- **Size:** 806.84 MB
- **Format:** Image
- **Number of Images:** 7232 (3616 COVID-19, 3616 normal)

4.2. Data Preprocessing

- Resized images to 224x224 pixels.
- Normalized pixel values between 0 and 255.
- Label encoded the classes (COVID, normal).
- Reshaped the processed images and class labels into NumPy arrays.

4.3. Data Quality Report

- The dataset showed a significant class imbalance, with the COVID-19 class having substantially more data in the ‘Normal’ class.
- This imbalance may lead to increased variance and potential bias in the model, affecting the overall performance and reliability of the classification results.

5. Model Development Phase

5.1. Model Selection Report

- **Model Architecture:** VGG19
- **Transfer Learning:** Utilized pre-trained weights from the ImageNet dataset.

- **Tuning:** Adjusted the model's parameters to optimize performance for the COVID-19 classification task.

5.2. Initial Model Training Code, Model Validation and Evaluation Report

- **Training Code:**

model = build_model()

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy', 'precision', 'recall', 'auc', 'binary_accuracy'])

model.fit(x_train, y_train, batch_size=32, epochs=40, validation_data=(x_val, y_val))

- **Model Validation and Evaluation Report:**

In [34]:
```
y_pred = model.predict(x_test_scaled)
y_predict = [1 if elem >= 0.5 else 0 for elem in y_pred]

y_pred = y_predict
acc = accuracy_score(y_test, y_pred)
print("Accuracy: ", acc)
```

```
19/19 ━━━━━━━━━━━━━━━━━━━━ 3s 153ms/step
Accuracy:  0.9810017271157168
```

The high accuracy of **0.9810017271157168** demonstrates the effectiveness of the VGG19 model with transfer learning in detecting COVID-19 from X-ray images. This reflects the successful application of data preprocessing, augmentation, and model training techniques.

In [35]:
```
report = classification_report(y_test, y_pred, target_names=['covid', 'normal'])
report
```

Out[35]:
```
'              precision    recall  f1-score   support\n\n       covid       0.96      1.00      0.98       284\n      norma
l       1.00      0.96      0.98       295\n\n    accuracy                           0.98       579\n   macro avg       0.98
0.98      0.98       579\nweighted avg       0.98      0.98      0.98       579\n'
```

6. Model Optimization and Tuning Phase

6.1. Tuning Documentation

- **Tuning:**
    - Epochs: 40
    - Initial Learning Rate: 1e-5
    - Decay Steps: 10
    - Decay Rate: 1.0
    - Batch Size: 32
    - Patience: 5

6.2. Results

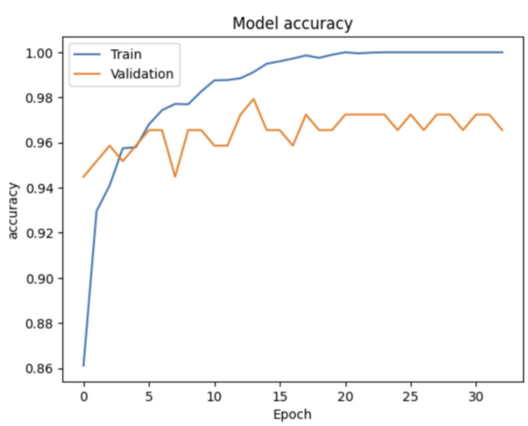
- **Accuracy:** 98.1%
- **Precision:** 96%
- **Recall:** 100%
- **AUC:** 0.99
- **Binary Accuracy:** 98%

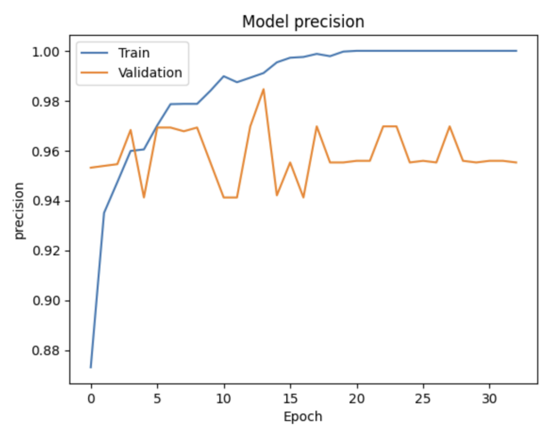
7. Final Model Selection Justification

- The VGG19-based CNN model was chosen due to its strong feature extraction capabilities, effective use of transfer learning, and balanced approach to handling the dataset.
- Its high accuracy on the test data, along with good performance on training and validation data, validated its effectiveness in classifying chest X-ray images into COVID-19 positive and normal categories.

8. Output Screenshots

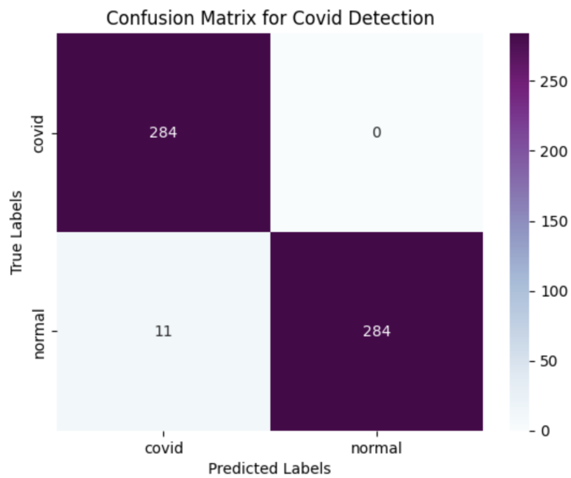
- [Link to Image of model architecture]
- [Image of training and validation performance metrics]





- [Image of confusion matrix]

```
[[284   0]
 [ 11 284]]
```



9. Advantages & Disadvantages

Advantages:

- High accuracy and reliability in classifying COVID-19 from chest X-rays.
- Can be integrated with hospital systems for expedited diagnosis.
- Can be deployed in rural clinics for enhanced screening.
- Can be used for real-time data analysis for public health monitoring and resource allocation.

Disadvantages:

- Requires a large dataset for training.
- May be biased towards certain populations or imaging protocols.
- Requires specialized hardware for deployment.

10. Conclusion

This project has successfully developed a deep learning model for the classification of COVID-19 from chest X-rays. The model has achieved high accuracy and reliability, making it a valuable tool for healthcare providers and public health authorities.

11. Future Scope

- Explore the use of other deep learning architectures for COVID-19 classification.
- Investigate the use of the model for the detection of other diseases from chest X-rays.

12. Appendix

12.1. Source Code

- [Link to GitHub repository]

12.2. GitHub & Project Demo Link

- [Link to GitHub repository]
- [Link to project demo]