Data Collection and Preprocessing Phase

Date	12 July 2024
Team ID	SWTID1720351492
Project Name	CovidVision: Advanced COVID-19 Detection from Lung X-Rays with Deep Learning
Maximum Marks	3 Marks

Data Collection Plan & Raw Data Sources Identification

For our project, we began by preprocessing images from the COVID-19 Radiography Database, a compilation by researchers from Qatar University, Doha, in collaboration with Dhaka University, Pakistan, and Malaysia. Initially, the database included 219 COVID-19, 1341 normal, and 1345 viral pneumonia images, later expanding to 3616 COVID-19 cases, 10,192 normal, 6012 lung opacity, and 1345 viral pneumonia images. We applied transfer learning algorithms to these images, studied how deep neural networks detect diseases, assessed model accuracy, and developed web applications using the Flask framework. The database is publicly accessible on Kaggle (https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database), with a size of 806.84 MB.

Data Collection Plan

Section	Description
Project Overview	What we did involves preprocessing images, applying transfer learning algorithms, understanding how deep neural networks detect diseases, evaluating model accuracy, and building web applications using the Flask framework.
Data Collection Plan	Kaggle Dataset

Raw Data Sources Identified	COVID-19 Radiography Database

Raw Data Sources

Source Name:

COVID-19 Radiography Database

Description:

Researchers from Qatar University, Doha, in collaboration with Dhaka University and partners in Pakistan and Malaysia, have compiled a COVID-19 chest X-ray database. Initially containing 219 COVID-19, 1341 normal, and 1345 viral pneumonia images, the database grew to 3616 COVID-19 cases, 10,192 normal, 6012 lung opacity, and 1345 viral pneumonia images over subsequent updates.

Location/URL:

https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

Format:

Image

Size:

806.84 MB

Access Permissions:

public (available on Kaggle)