Data Collection and Preprocessing Phase

Date	6 JULY 2024
Team ID	SWTID1720110768
Project Title	CovidVision: Advanced COVID-19 Detection from Lung X-rays with Deep Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report

Gathering, validating, and storing all necessary data systematically to ensure accuracy and completeness for predicting covid 19 using lung X-ray.

Data Collection Plan

Section	Description
Project Overview	• This deep learning project aims to predict covid 19,just by examining the lung lung X-rays. • The objective is to improve accuracy in virus detection and reduce time taken for test results.
Data Collection Plan	• Search for datasets which have images of lung X-rays • Prioritize datasets with diverse demographic information.

Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, different images of X-rays.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The data set contains customer details X-ray of lungs of different people,some may have positive ,some may have negative.	https://www.kaggle.com/datasets/sid321axn/covid-cxr-image-dataset-research	CSV	600MB	Public