Data Collection and Preprocessing Phase

Date	6th July 2024
Team ID	SWTID1720195938
Project Title	**CovidVision: Advanced COVID-19 Detection from Lung X-Rays with Deep Learning**
Maximum Marks	2 Marks

Data Collection Plan Template

Section	Description
Project Overview	COVID-19 Detection from Lung X-rays" uses deep learning algorithms to examine lung X-ray images for signs of COVID-19 infection. This research intends to deliver accurate and speedy diagnosis using large datasets and image recognition technologies, assisting in early detection and viral containment.
Data Collection Plan	Data will be collected from publicly accessible datasets, online sources, and published papers. These include Kaggle and GitHub repositories.
Raw Data Sources Identified	1. COVID-19 CHEST X-RAY DATABASE: A collection of chest X-ray images for COVID-19 positive cases, normal lung images, lung opacity images, and viral pneumonia images. The data has been collected from multiple sources including Padchest, Germany medical schools, SIRM, RSNA, and Kaggle.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	COVID-19 CHEST X-RAY DATABASE: Contains images for COVID-19 positive cases, normal lung images, lung opacity images, and viral pneumonia images.	Link of Dataset	PNG	**778 MB**	Public
Dataset 2	Collection of 2473 CXR images from Padchest dataset.	Link of Dataset	PNG	153 GB	Public
Dataset 3	Collection of 183 CXR images from a Germany medical school.	Link to Dataset	PNG	…	Public
Dataset 4	Collection of 400 CXR images from	Link to Dataset	PNG		Public

	another GitHub source.				
Dataset 5	Collection of 1345 viral pneumonia images from Chest X-Ray Images (pneumonia) database.	Link to Dataset	PNG	2 GB	Public