Data Collection and Preprocessing Phase

Date	2 Dec 2024
Team ID	TMID739650
Project Title	ADVANCED COVID-19 DETECTION USING LUNG X-RAYS BY DEEP LEARNING
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	The project aims to develop a deep learning model for advanced COVID-19 detection using lung X-rays. It will classify lung X-ray images into categories such as "COVID-19 positive," "non-COVID-19 pneumonia," and "healthy."
Data Collection Plan	The data will be collected from publicly available lung X-ray image datasets such as the COVID-19 X-ray dataset from Kaggle, the Chest X-ray14 dataset, and other reliable medical repositories.
Raw Data Sources Identified	1. **COVID-19 X-ray Dataset (Kaggle)**: Contains labeled X-ray images of COVID-19, pneumonia, and normal lung scans. 2. **Chest X-ray14 Dataset**: A large-scale dataset of chest X-rays, useful for training models for various lung diseases. 3. **Radiological Society of North America (RSNA) Pneumonia Detection Challenge**: Includes labeled pneumonia X-ray images for comparison.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Lung x-ray of a covid affected person.	Covid19 detection using Tensorflow from Chest Xray	Image	36 KB	Public
Dataset 2	Lung x-ray of normal people.	Covid19 detection using Tensorflow from Chest Xray	Image	35 GB	Public
Dataset 3	Lung x-ray of person affected with lung opacity	Covid19 detection using Tensorflow from Chest Xray	Image	35.34 KB	Public