Data Collection and Preprocessing Phase

Date	4 July 2024
Team ID	SWTID1720097611
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 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 CovidVision project utilizes deep learning algorithms to analyze lung X-ray images for signs of COVID-19 infection. By leveraging extensive datasets and image recognition technology, the objective is to build a model that accurately detects COVID-19, facilitating rapid and reliable diagnosis. This system aims to aid medical staff in early detection and triage, especially in overwhelmed hospitals and rural areas with limited access to expert radiologists.
Data Collection Plan	Identified and Gathered Data Sources on collecting datasets specifically from Kaggle, which includes labeled lung X-ray images for COVID-19, pneumonia, and healthy cases. Implemented preprocessing steps such as resizing, normalizing, and augmenting images to create a standardized input for the deep learning model.

Raw Data Sources Identified	The dataset from Kaggle includes: • **COVID-19 Chest X-ray Images**: A collection of X-ray images labeled for COVID-19 detection. • **Pneumonia and Normal Images**: X-ray images labeled for pneumonia and normal cases, providing a comparative basis for the model. • **Viral Pneumonia X-ray Images:** A collection of lung X-ray images of people having pneumonia.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	The dataset comprises chest X-ray images categorized into four classes: COVID-19, pneumonia, normal, and lung Opacity. It includes images of patients with COVID-19, various forms of pneumonia, healthy individuals, and those with other lung diseases. This diverse dataset is designed to aid in developing and validating models for accurate COVID-19 detection and comprehensive lung condition analysis.	https://www.kaggle.com/code/rollanmaratov/covid19-detection-using-tensorflow-from-chest-xray/data	PNG	769MB	Public