Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWUID20240034617
Project Title	CovidVision : Advanced COVID-19 Detection for 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	CovidVision aims to develop a deep learning model to detect COVID-19 from chest X-ray images. The project focuses on creating a convolutional neural network (CNN) that can accurately differentiate between COVID-19, other pneumonias, and healthy lungs. Key objectives include achieving high accuracy, sensitivity, and specificity, as well as ensuring the model is robust and generalizable across diverse datasets. This can provide a faster, resource-efficient alternative to traditional diagnostic methods..
Data Collection Plan	1.Skill Wallet Platform 2.Kaggle

Raw Data Sources Identified	https://www.kaggle.com/code/rollanmaratov/covid19-detection-using-tensorflow-from-chest-xray/data ,Skill Wallet Platform https://www.kaggle.com/datasets/imdevskp/corona-virus-report , kaggle

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Some X-ray images are missing patient metadata (age, gender).	https://www.kaggle.com/code/rollanmaratov/covid19-detection-using-tensorflow-from-chest-xray/data	JPEG, PNG	444 MB	Public
Dataset 2	Includes data on confirmed cases, deaths, recoveries, and other relevant statistics.	https://www.kaggle.com/datasets/imdevskp/corona-virus-report	CSV	4.26 MB	Public
…	…	…	…	…	…