

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720418653
Project Title	Crystal Clear Vision: Revolutionizing Cataract Prediction through Transfer Learning Mastery
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	Create a highly accurate Deep Learning model with Transfer Learning to identify cataracts based on fundus images. This would allow for early diagnosis and better patient outcomes from prompt and efficient treatment.
Data Collection Plan	We have taken data from our mentor provided data on Skill Wallet platform. For testing purpose we have taken a dataset from Kaggle.
Raw Data Sources Identified	<p>Kaggle Dataset:</p> <p>https://www.kaggle.com/datasets/kershruta/cataract</p> <p>This dataset comprises 672 annotated images categorized into two classes: cataract and normal. Each image has undergone preprocessing and augmentation to enhance its suitability for machine learning tasks</p>

	<p>Skill Wallet Dataset:</p> <p>https://www.kaggle.com/datasets/hemooredaoo/cataract</p> <p>This dataset comprises 1100 annotated images categorized into train and test directories each directory has both cataract and normal images.</p>
--	---

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	This dataset comprises 672 annotated images categorized into two classes: cataract and normal. Each image has undergone preprocessing and augmentation to enhance its suitability for machine learning tasks	https://www.kaggle.com/datasets/kershruta/cataract	Image	200 MB	Public

Dataset 2	<p>This dataset comprises 1100 annotated images categorized into train and test directories each directory has both cataract and normal images.</p>	https://www.kaggle.com/datasets/heemooredao/catara	Image	570 MB	Public
-----------	---	---	-------	--------	--------