Data Collection and Preprocessing Phase

Date	June 26, 2025
Team ID	SWTID1750058607
Project Title	Early-Stage Disease Diagnosis System Using Human Nail Image
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Data Scarcity/Imbalance: Explicitly mentions the dataset size (655 training, 183 test images across 17 classes) and how this can	Image Quality/Variability: Notes potential issues with inconsistent lighting, resolution, and background noise, which are common in real-world image datasets. Labeling Accuracy: Highlights the importance of accurate labeling for 17 distinct disease classes and the risk of misclassification if labels are ambiguous or incorrect.	Moderate. Data Leakage: Emphasizes the need to ensure no overlap between training and testing	Resolution Plans: Provides technical solutions for each issue, including data augmentation, advanced preprocessing (segmentation, color constancy), expert review for labeling, and strict data splitting protocols.

lead to overfitting and poor generalization, especially for minority classes.		sets to prevent inflated performance metrics.	
…	…	…	…