 

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/I mbalanc e: 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. | Modera te. Data Leakage  :  Emphasi zes the need to ensure no  overlap betwee n 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 |  | sets to |  |
| overfittin | prevent |
| g and | inflated |
| poor | perform |
| generaliz | ance |
| ation, | metrics. |
| especiall |  |
| y for |  |
| minority |  |
| classes. |  |
| … | … | … | … |