Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 20-06-2025 |
| Team ID | SWDTID1749906902 |
| Project Title | Early Stage Disease Diagnosis System Using Human Nail Image Processing |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Report:**

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:**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | The machine learning project aims to predict early-stage diseases based on human nail image analysis. Using a dataset with features extracted from nail images, the objective is to build a model that accurately classifies the presence or absence of specific diseases, facilitating efficient and informed decision-making in medical diagnosis. |
| Data Collection Plan | * Search for datasets related to human nail images, dermatological conditions, and medical image analysis. Prioritize datasets with diverse image characteristics and confirmed diagnostic labels. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle & UCI, popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as image data, and associated medical details for machine learning analysis |

**Raw Data Sources Report:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access**  **Permissions** |
| Kaggle Dataset | The dataset comprises human nail images and associated diagnostic labels for various diseases (e.g., fungal infections, anemia, psoriasis). | https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs | CSV | 48.9 MB | Public |