

**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 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.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | This project aims to develop an "Early Stage Disease Diagnosis System Using Human Nail Image Processing" by creating a deep learning model that analyzes nail images to detect potential health issues such as discoloration, ridges, and texture changes, offering a non-invasive and accessible tool for early disease screening. |
| Data Collection Plan | Data for this project is collected from various sources, specifically through the use of image datasets. The data consists of human nail images which are organized into  train.zip and test.zip files. These zip files contain images categorized into 17 different disease classes. The images are unzipped into |

|  |  |
| --- | --- |
|  | /content/train and /content/test directories for processing. |
| Raw Data Sources Identified | Dataset 1 (Train Data): This source contains images of human nails used for training the deep learning model. These images are categorized into 17 different disease classes. The original format is likely image files (e.g., PNG), and there are 655 images in the training dataset.  The data is accessed locally after unzipping the train.zip file.  Dataset 2 (Test Data): This source contains images of human nails used for testing the deep learning model.  Similar to the training data, these images are categorized into 17 different disease classes. The original format is also likely image files (e.g., PNG), and there are 183  images in the test dataset. The data is accessed locally after unzipping the  test.zip file. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Train Data | Images of human nails used for training the deep  learning model, categorized into 17 different disease classes. | [https://drive.g](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [oogle.com/driv](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [e/folders/1AXT](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [YsbiarS1TCAgȠ](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [0mancTSrJYYM](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [WMs?usp=sha](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [ring](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) | Image | Imag e (PNG,  impli ed  from usage  ) | Public |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  | 655  image s |  |
| Test Data | Images of human nails used for testing the deep  learning model, categorized into 17 different disease classes. | [https://drive.g](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [oogle.com/driv](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [e/folders/1AXT](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [YsbiarS1TCAgȠ](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [0mancTSrJYYM](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [WMs?usp=sha](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) [ring](https://drive.google.com/drive/folders/1AXTYsbiarS1TCAgfj0mancTSrJYYMWMs?usp=sharing) | Image | Imag e (PNG,  impli ed  from usage  )  183  image s | Private (with access) |
| … | … | … | … | … | … |