Raw Data Sources and Data Quality Report

Date	11-03-2025
Team ID	740037
Project Title	Early Stage Disease Diagnosis System Using Human Nail Image Processing Using Deep Learning.
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	This project aims to create an AI-based system for early-stage disease diagnosis using human nail image processing. By leveraging deep learning models, specifically convolutional neural networks (CNNs), the system will analyze nail images to detect visual indicators of health conditions such as anemia, liver disease, and fungal infections. The solution is designed to be accessible, non-invasive, and supportive of early healthcare intervention, especially in regions lacking advanced medical infrastructure.
Data Collection Plan	□ **Gathering** high-quality images of fingernails under consistent lighting conditions □ **Labeling** each image based on the associated medical condition (e.g., healthy, anemic, jaundiced, fungal infection, etc.) □ **Augmenting** the dataset using image transformations (e.g., rotation, zoom, brightness adjustments) to improve model robustness

	□ **Splitting** the data into training, validation, and test sets to ensure accurate evaluation
Raw Data Sources Identified	□ **Public medical image databases** (e.g., Kaggle, NIH datasets, open dermatology image repositories) □ **Healthcare partnerships** (e.g., collaborating with local clinics or hospitals for anonymized nail images, if possible) □ **Online sources and research papers** (e.g., scraping or requesting access to datasets from published studies) □ **Synthetic image generation** using GANs (Generative Adversarial Networks) for rare disease cases where data is limited

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises of various damaged car images.	Kaggle	CSV	4 MB	Public