

## Data Collection and Preprocessing Phase

Date	15 April 2025
Name	Sainath Prakash Nimbalkar
Project Title	NailCare AI
Maximum Marks	2 Marks

### Data Collection Plan

Section	Description
Project Overview	This deep learning project focuses on classifying images of 16 different Nail disease—using Convolutional Neural Networks (CNNs). The objective is to uncover hidden patterns and visual cues that distinguish each type, contributing to better Nail Disease identification in the wild.
Data Collection Plan	The dataset has been sourced from a ZIP file provided by the SmartInternz, which includes categorized images in subdirectories named after each NAIL DISEASE. Additional reference images were accessed from publicly available sources such as Wikimedia.
Raw Data Sources Identified	The raw data includes SmartInternz provided images saved in structured subdirectories, supplemented by publicly available datasets for training and validation purposes.

## Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
SmartInternz Provided Dataset	Curated image dataset provided by SmartInternz, containing Nail disease images in separate subdirectories.	<a href="https://drive.google.com/drive/folders/1AXTYsbjarS1TCAgfj0mancTSrJYYMWMs?usp=sharing">https://drive.google.com/drive/folders/1AXTYsbjarS1TCAgfj0mancTSrJYYMWMs?usp=sharing</a>	ZIP File	~ 175 MB	Public
Field Captured Images	Manually photographed images taken in natural environments, used for supplementing the dataset.	Local Storage	JPG/PNG	~100 MB	Private
Wikimedia	Open-source mushroom images used for visual verification and dataset augmentation.	<a href="https://en.wikipedia.org/wiki/File:Fingernails2.jpg">https://en.wikipedia.org/wiki/File:Fingernails2.jpg</a>	JPG/PNG	~10 MB	Public