Data Collection and Preprocessing Phase

Date	20 June 2025
Team ID	SWTID1749821186
Project Title	Enhancing Product Reliability: Leveraging Transfer Learning for Fault Detection
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification

Data Collection Plan Template

Section	Description
Project Overview	This project aims to automate the detection of casting defects in manufacturing using a deep learning model (VGG16) integrated into a Flask web app. The objective is to improve quality inspection by replacing manual processes with an AI-powered solution.
Data Collection Plan	The dataset used for this project is sourced from publicly available industrial datasets on Kaggle. The primary dataset contains labeled images of casting products categorized as defective or good. Additional data may be collected through synthetic augmentation.
Raw Data Sources Identified	The main dataset is a real-life industrial dataset of casting products. It includes grayscale images of front-facing casting components, labeled as either defective or non-defective.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Real-life industrial dataset of casting products. Contains grayscale images labeled as defective or good.	https://www.kaggle.com/datasets/ravirajsinh45/real-life-industrial-dataset-of-casting-product	Image	~1.2 GB	Public
Dataset 2	Augmented dataset generated from Dataset 1 using flipping, rotation, and zooming techniques.	Local (generated during preprocessing)	Image	~2.5 GB	Private (project-local)