

Data Collection and Preprocessing Phase

Date	05 July 2025
Team ID	SWTID1749835721
Project Title	HematoVision - Blood Cell Classification using Transfer Learning
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 model development phase.

Data Collection Plan:

Section	Description
Project Overview	The machine learning project aims to classify blood cell types based on microscopic image data. The objective is to build a convolutional neural network (CNN) that can accurately distinguish between four major classes of white blood cells: NEUTROPHIL , LYMPHOCYTE , MONOCYTE , and EOSINOPHIL . By leveraging transfer learning techniques, the model intends to support efficient and accurate diagnosis in medical imaging.
Data Collection Plan	<ul style="list-style-type: none">● Search for blood cell image datasets suitable for classification tasks● Prioritize datasets with labeled subfolders or metadata for supervised learning● Ensure high-resolution, color images compatible with CNN architectures● Validate images for consistency, corruption, and labeling errors before modeling

Raw Data Sources Report	<p>This dataset forms the backbone of the image classification pipeline. All images were structured into respective class folders and verified for consistency. Preprocessing steps, including resizing, normalization, and label encoding, were performed to ensure data readiness for training.</p> <p>.</p>
-------------------------	--

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle data set	Contains labeled images of NEUTROPHIL, EOSINOPHIL, MONOCYTE, and LYMPHOCYTE	https://www.kaggle.com/datasets/paultimothymooney/blood-cells/data	CSV	15 kB	Public
Google drive folder	ZIP file uploaded from Google Drive into Google Colab for data loading and extraction	https://drive.google.com/drive/folders/1Ca3JzJ-QZDatAGnst0JMVozAgymc5_nH?usp=sharing	CSV	13.6 kB	Public

