GastroCAD

Assisting doctors during Esophagogastroduodenoscopy (EGD)



Mudasir Rasheed   Mannan Sharma   Devansh Bartwal

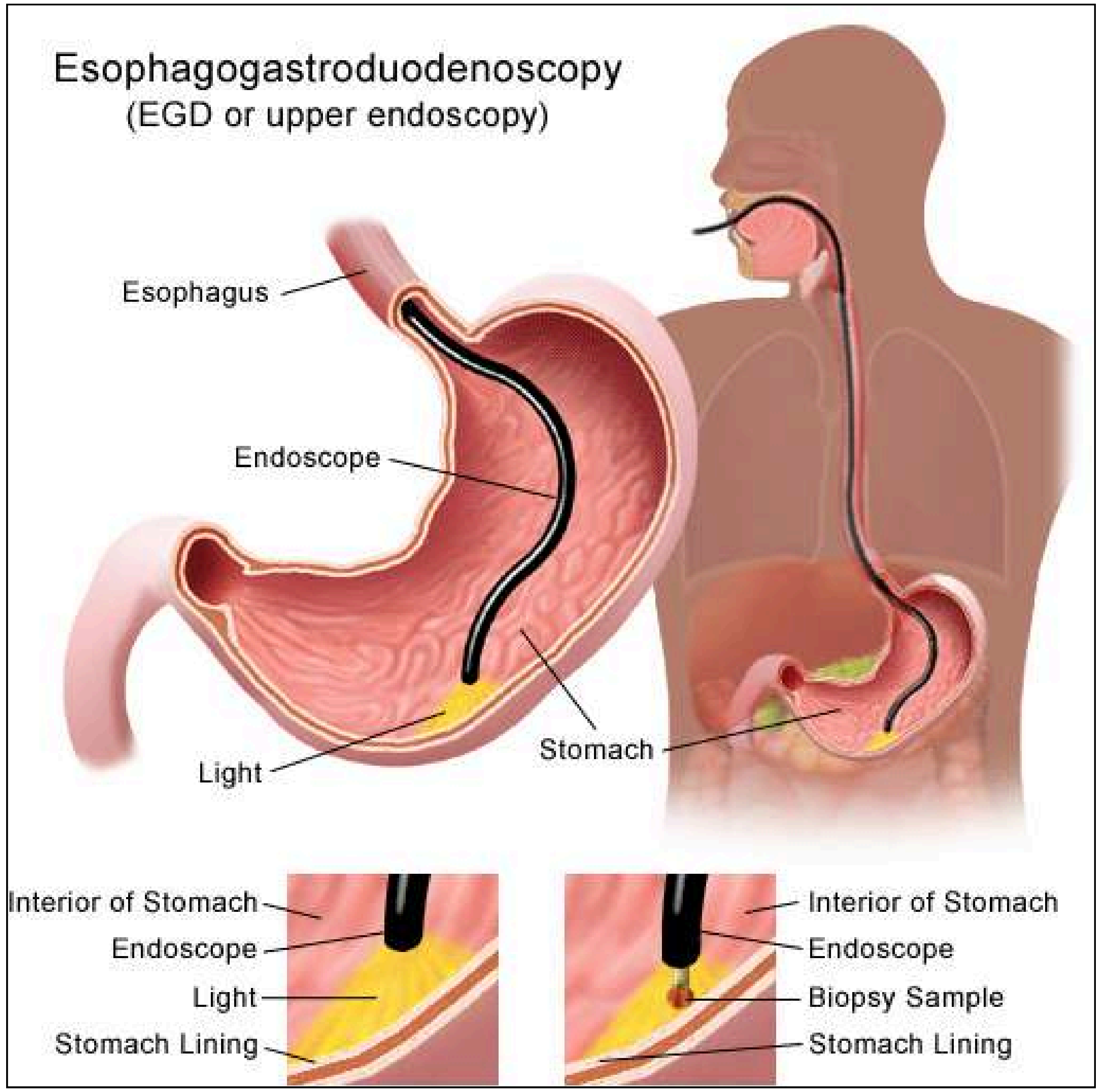
What are we doing?

We are developing a tool to assist doctors in the procedure of EGD.

EGD means Esophagogastroduodenoscopy

What is EGD ?

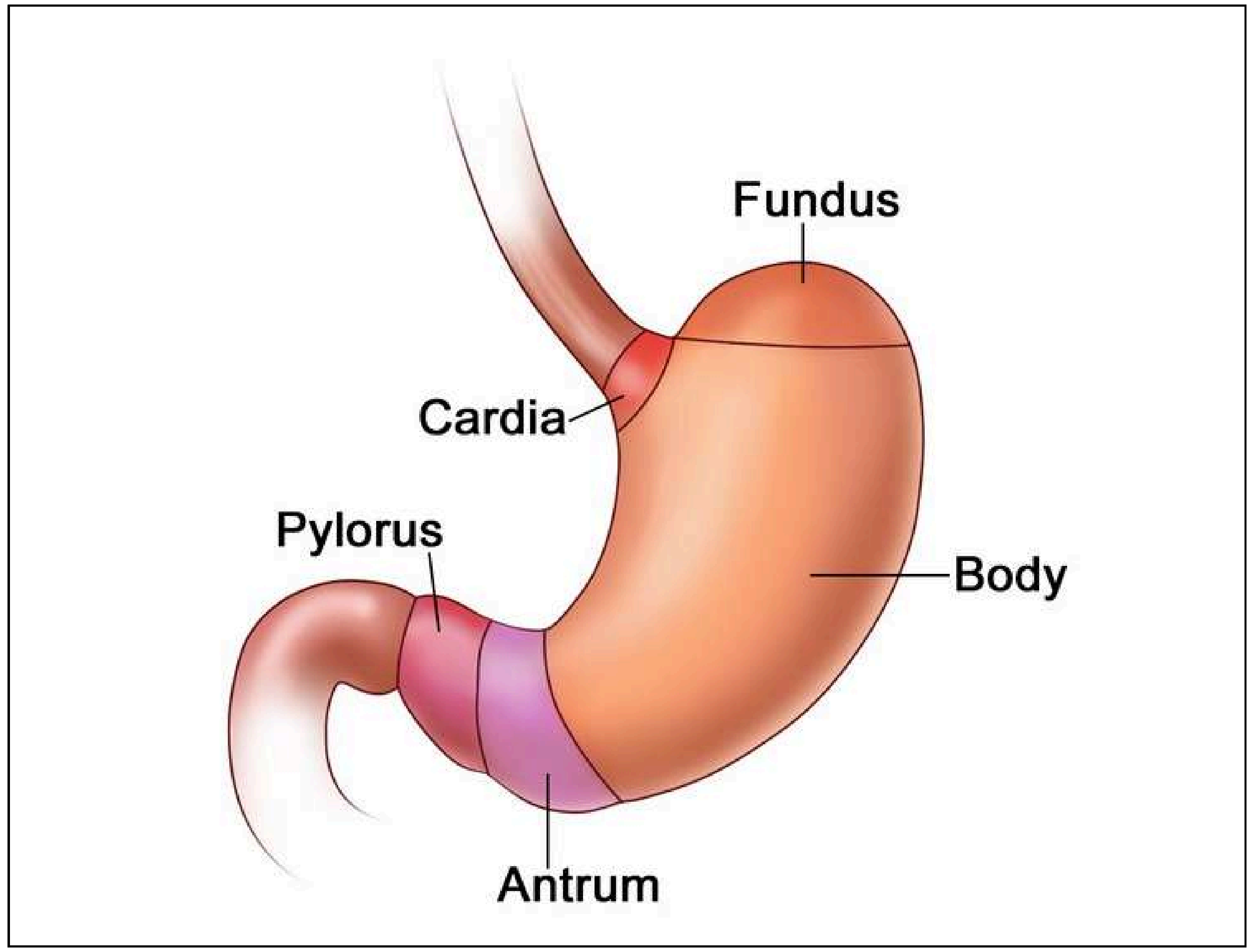
It's a diagnostic procedure where a doctor uses a flexible tube with a camera (called an endoscope) to look at the upper part of the gastrointestinal (GI) tract.



https://emedicine.medscape.com/article/1851864-overview?form=fpf

Photodocumentation Protocols in EGD

- Protocols guide the endoscopist to systematically capture key anatomical landmarks.
- Prevents missed areas or “blind spots,” especially in the stomach, where the organ’s complex shape can cause certain regions to be easily overlooked, particularly by less experienced endoscopists.
- Helps standardise what a "complete" EGD looks like.



https://doi.org/10.1055/s-0042-113128
https://doi.org/10.7704/kjhugr.2020.0013

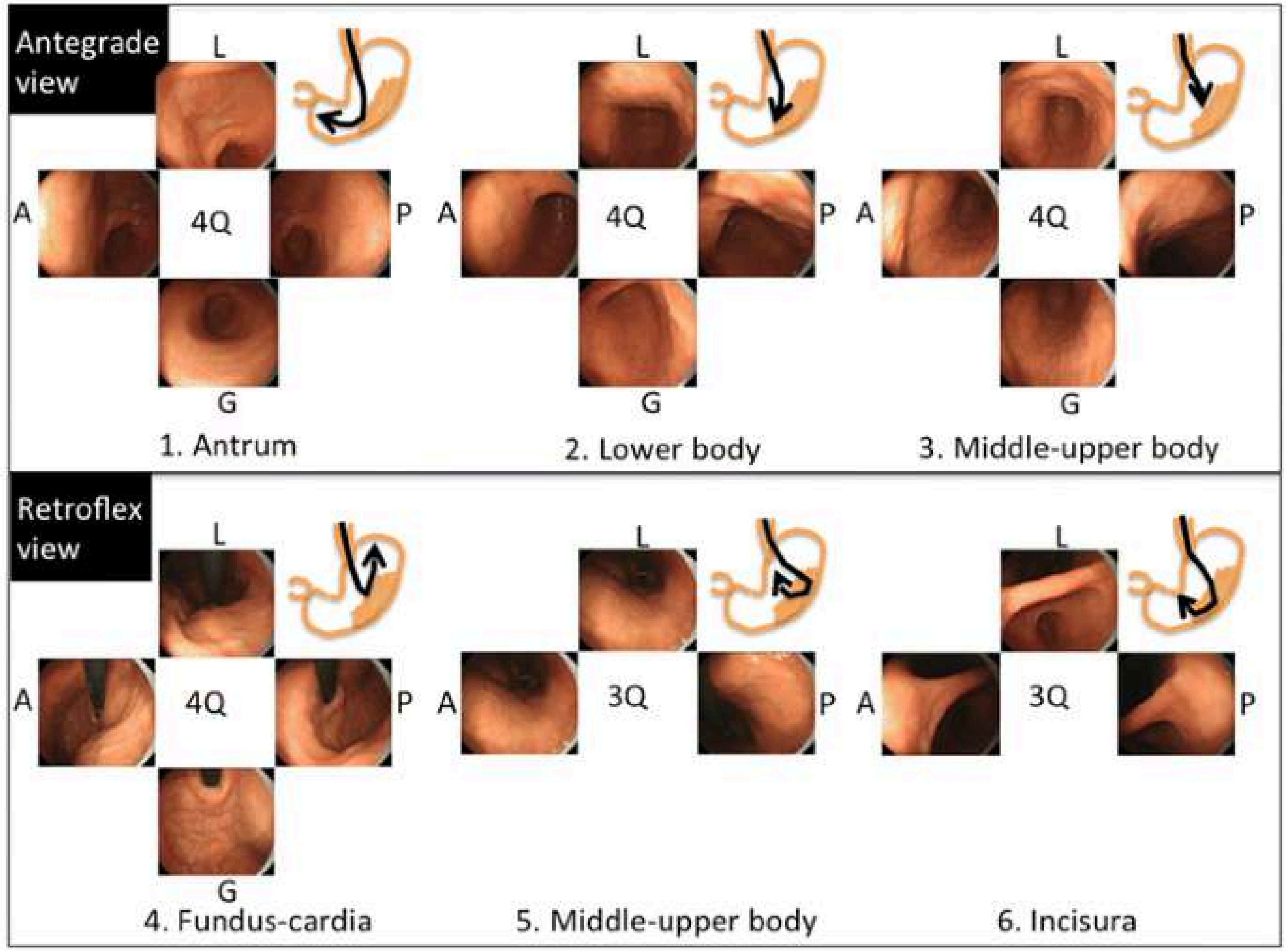
Photodocumentation Protocols in EGD

- **ESGE(European Society of Gastrointestinal Endoscopy) protocol**: Recommends examination of 10 distinct regions of stomach for complete examination.
- **KSGE(Korean Society of Gastrointestinal Endoscopy) protocol**: Divides stomach into 8 distinct regions for examination during EGD.
- **(SSS kehshi Yahoo – Systematic Screening of the Stomach)**: Uses a rigorous 22-image protocol to comprehensively screen the stomach and minimise blind spots for early cancer detection.

https://doi.org/10.1055/s-0042-113128
https://doi.org/10.7704/kjhugr.2020.0013
https://doi.org/10.1055/s-2001-42537

The SSS-Kenshi Yao Protocol

- Developed by Dr. Kenshi Yao, a Japanese endoscopist renowned for early gastric cancer detection.
- Covers 22 standard images to eliminate blind spots in the stomach and ensure complete mucosal visualisation.

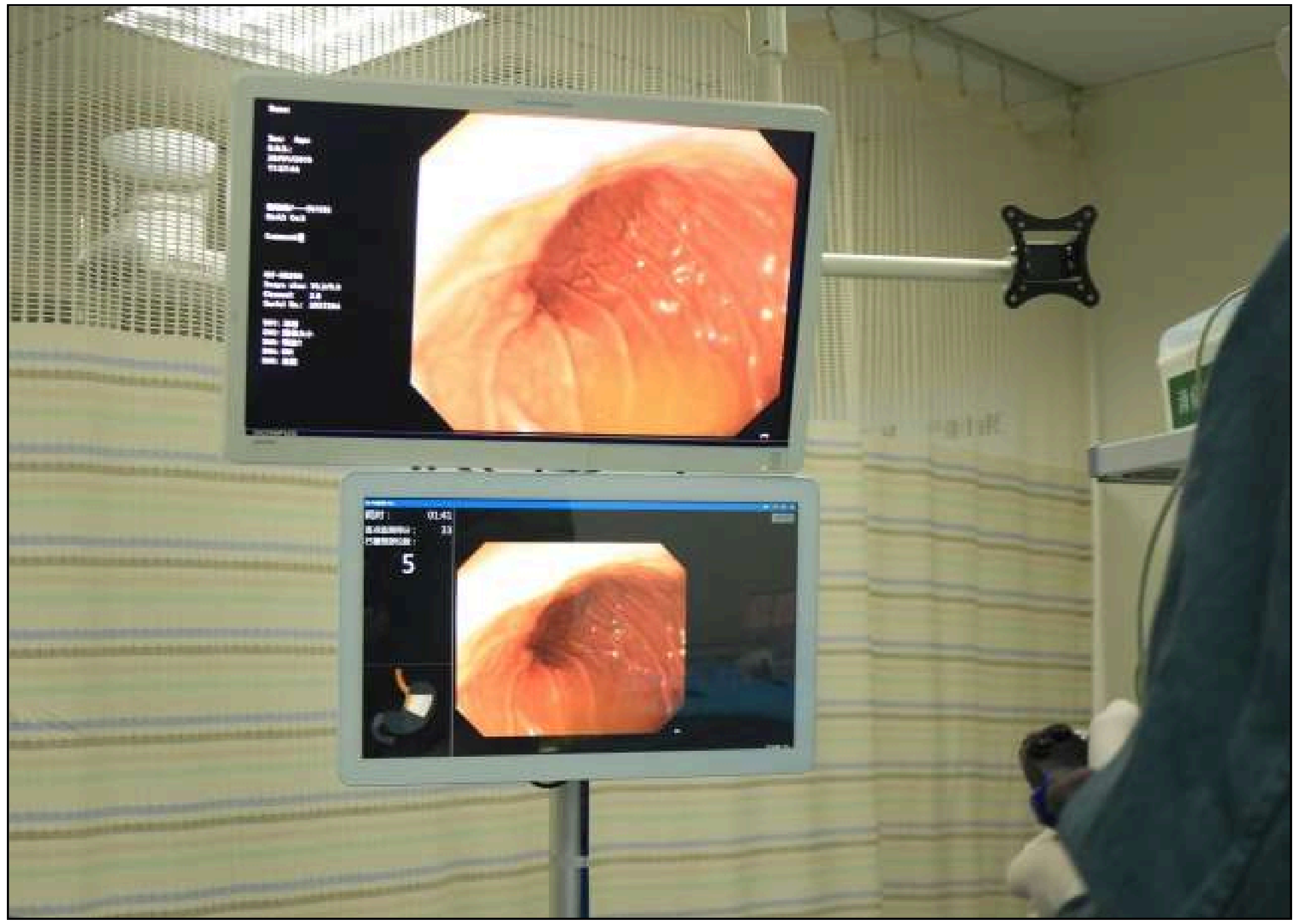


http://www.annalsgastro.gr/index.php/annalsgastro/article/download/1184/1014?inline=1#:~:text=A%20basic%20technique%20for%20avoiding,3)%20mapping%20the%20entire%20stomach

The Debate among gastroenterologist

Although the **SSS (Kenshi Yao)** protocol in gastroenterology is more detailed and ensures a thorough examination of the stomach, it is not commonly used in daily practice due to the longer procedure time required.

The **SSS (Kenshi Yao)** protocol requires more time, which can affect patient comfort as they must tolerate the endoscope throughout the extended examination.



https://doi.org/10.1053/j.gastro.2017.05.009
https://doi.org/10.1016/j.cgh.2014.07.059
https://doi.org/10.7704/kjhugr.2020.0013

Gap in the Literature

Inexperienced gastroendoscopists often miss certain regions during the EDG procedure

↓

Different photodocumentation protocols are developed to minimise it

↓

SSS yahoo protocol is most granular screening protocol

↓

Takes more time as photoducemenattion needs more images

↓

Our model ensures no spot is left unexamined in EGD, adhering to SSS protocol, and EGD takes less time.

Existing ML Work

Existing literature has predominantly focused on disease classification and lesion detection during **esophagogastroduodenoscopy** (EGD), while limited research has addressed the standardisation of EGD procedures and the improvement of overall EGD procedure itself.

https://doi.org/10.5230/jgc.2024.24.e28

Existing ML Work

ENAD CAD-G: Based on the ESGE protocol, EfficientNetB0 was trained to classify endoscopic images into 10 anatomical gastric locations. accuracy, 82.47% confidence interval [CI], 76.46%–88.47%.

ENDOANGEL-ME: Based on M-IEEE Protocol (Magnifying Image-Enhanced Endoscopy)
Based on convolutional neural network
M-IEE uses optical zoom and Narrow-band imaging (NBI).
sensitivity 82.03%

https://doi.org/10.5230/jgc.2024.24.e28
https://doi.org/10.1016/j.gie.2021.11.040

Our Dataset

Publication Date: 17 January 2025

Published By / Authors: Diego Bravo, Juan Frias, Felipe Vera, Juan Trejos, Carlos Martínez, Martín Gómez, Fabio González, and Eduardo Romero

Published In: Scientific Data, Volume 12, Article number: 102 (2025)

Total Patients: 387 high-definition esophagogastroduodenoscopy (EGD) cases

https://doi.org/10.1038/s41597-025-04401-5

Our Dataset

Images:

- 8,834 images
- Each image represents a selected frame from one of the 22 gastric anatomical landmarks or categorised as "NA" (unqualified images)

Video Sequences:

- 4,729 labelled sequences from 223 videoendoscopies
- Each sequence corresponds to one anatomical location

Annotations:

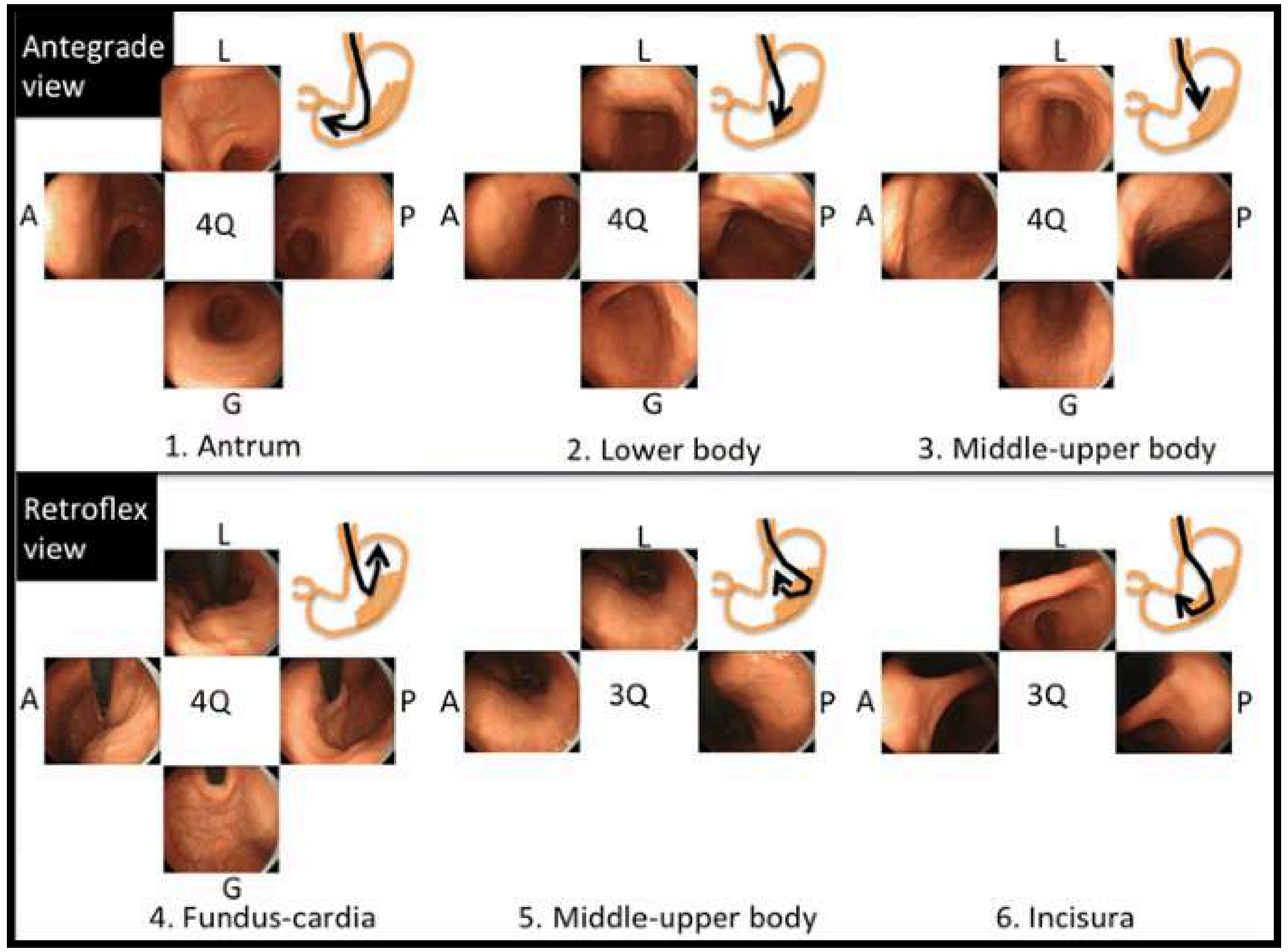
- Labelled by four medical experts
- Two recent medical graduates and two final-year gastroenterology fellows
- 23 categories total: 22 anatomical landmarks + 1 unqualified category

Protocol Used:

- Systematic Screening Protocol for the Stomach (SSS)
- Developed by Kenshi Yao (Japan)

https://doi.org/10.1038/s41597-025-04401-5

Our Dataset



Code	Region
A1	Anterior wall - Antrum
A2	Anterior wall - Lower body
A3	Anterior wall - Middle upper body
P1	Posterior wall - Antrum
P2	Posterior wall - Lower body
P3	Posterior wall - Middle upper body
L1	Lesser curvature - Antrum
L2	Lesser curvature - Lower body
L3	Lesser curvature - Middle upper bod
G1	Greater curvature - Antrum
G2	Greater curvature - Lower body
G3	Greater curvature - Middle upper bo
A4	Anterior wall - Fundus cardia
A5	Anterior wall - Middle upper body
A6	Anterior wall - Incisura
P4	Posterior wall - Fundus cardia
P5	Posterior wall - Middle upper body
P6	Posterior wall - Incisura
L4	Lesser curvature - Fundus cardia
L5	Lesser curvature - Middle upper bod
L6	Lesser curvature - Incisura
G4	Greater curvature - Fundus cardia
NA	Not applicable (unqualified images)

https://doi.org/10.1038/s41597-025-04401-5

Data Cleaning and Exploratory Data Analysis

GastroHUN(Images)

Train Set Size: 6108

Validation Set Size : 1306

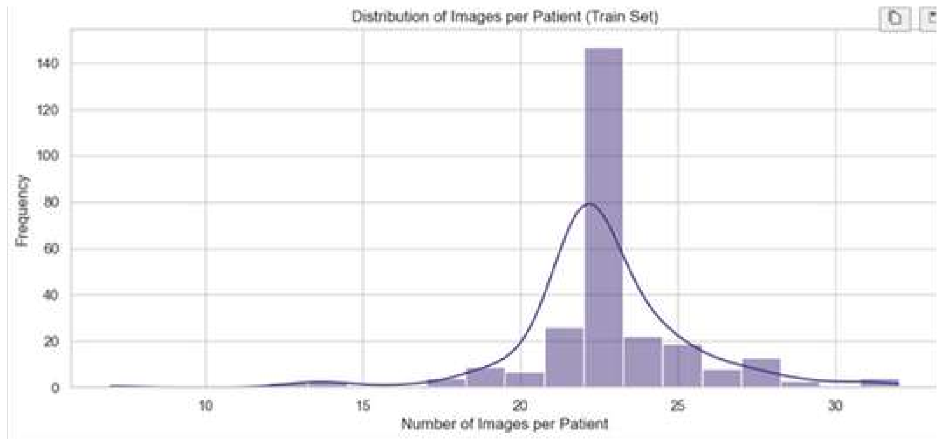
Test Set Size : 1345

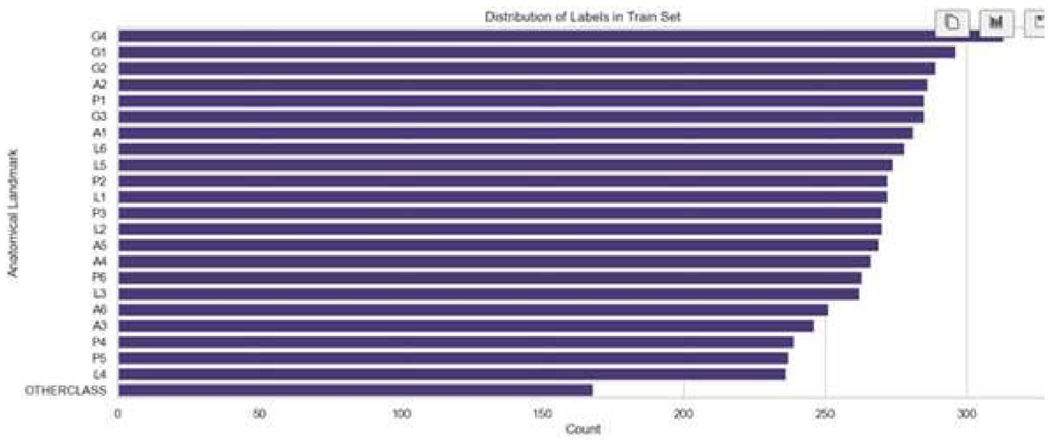
No. of labels : 22+1

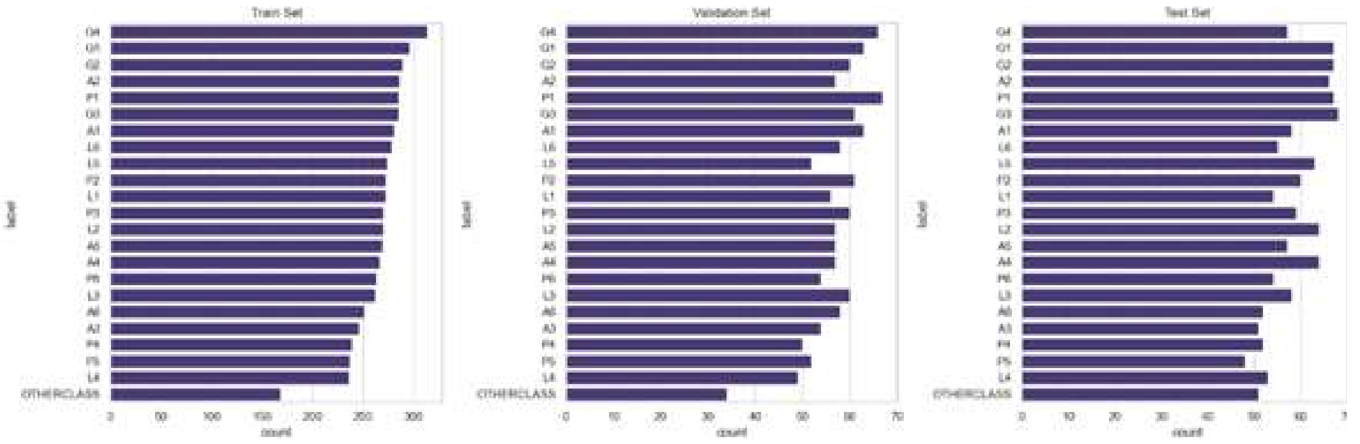
No of rows in dataset =No. of images = 8759

Total Unique Patients=387

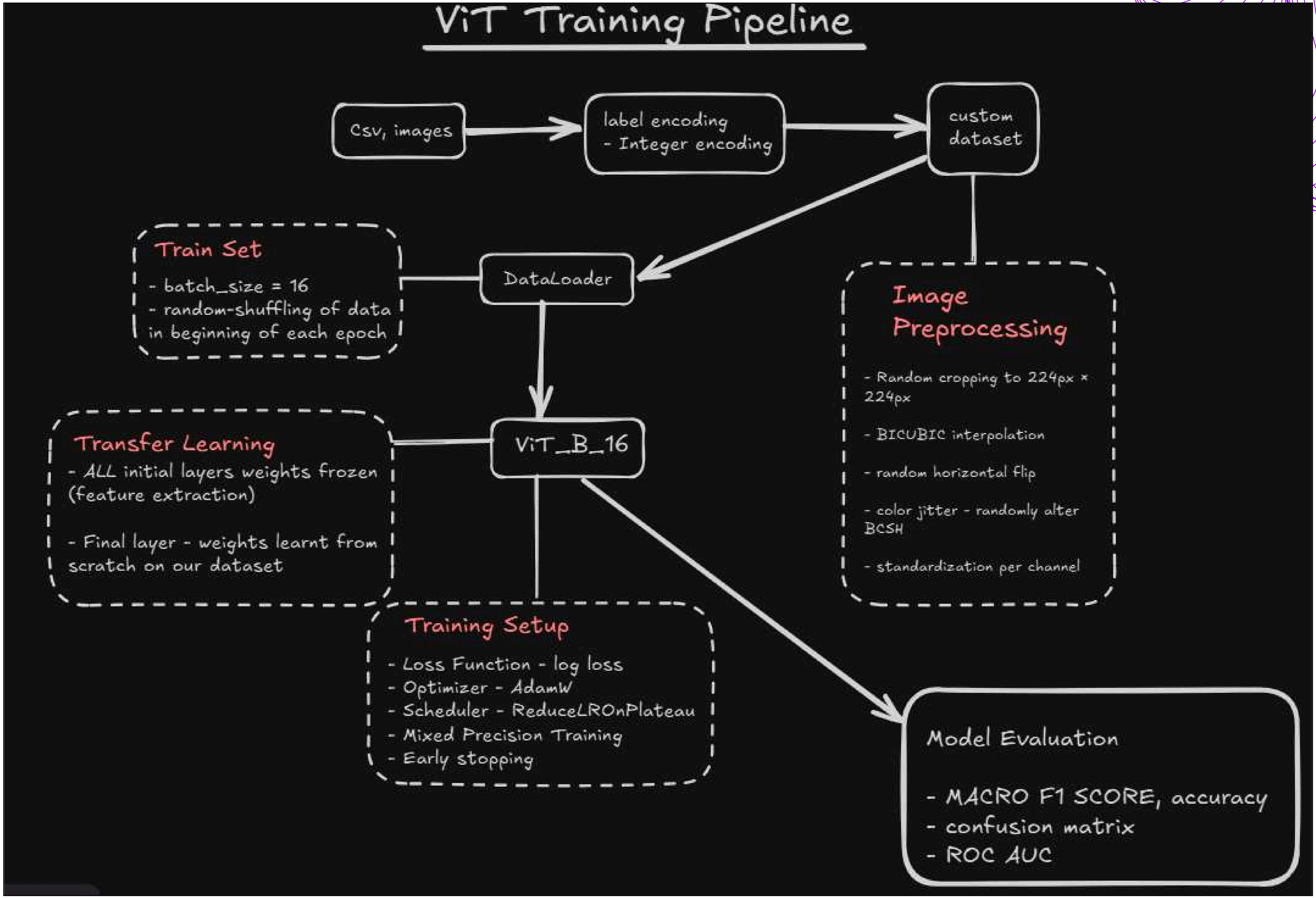
In each of the sets, no class is imbalanced compared to others







https://github.com/MannanSharma01/GastroCAD


ViT Training Pipeline
Csv, images
label encoding
- Integer encoding
custom dataset
Train Set
- batch_size = 16
- random-shuffling of data in beginning of each epoch
DataLoader
Image Preprocessing
- Random cropping to 224px × 224px
- BICUBIC interpolation
- random horizontal flip
- color jitter - randomly alter BCSH
- standardization per channel
Transfer Learning
- ALL initial layers weights frozen (feature extraction)
- Final layer - weights learnt from scratch on our dataset
ViT_B_16
Training Setup
- Loss Function - log loss
- Optimizer - AdamW
- Scheduler - ReduceLROnPlateau
- Mixed Precision Training
- Early stopping
Model Evaluation
- MACRO F1 SCORE, accuracy
- confusion matrix
- ROC AUC

ViT

- **Model Used : vit_b_16** from torchvision.models
- **Optimiser**:-
  - **AdamW** - combine the benefits of **adaptive learning rate** and **L2-regularisation**
  - More suitable for transformer-based architectures than classic ADAM
  - Initial LR = 1e-4 , weight decay (similar to L2 penalty) = 1e-4

- **ReduceLROnPlateau:** reduce LR on constant val-loss across **3** epochs.
- **Early Stopping** (on val loss): **patience = 5 epochs.**

- **Dropout:** Built-in, ViT includes dropout after linear layers, before attention layers, etc. In **vit_b_16**, **dropout_probability = 0.1**

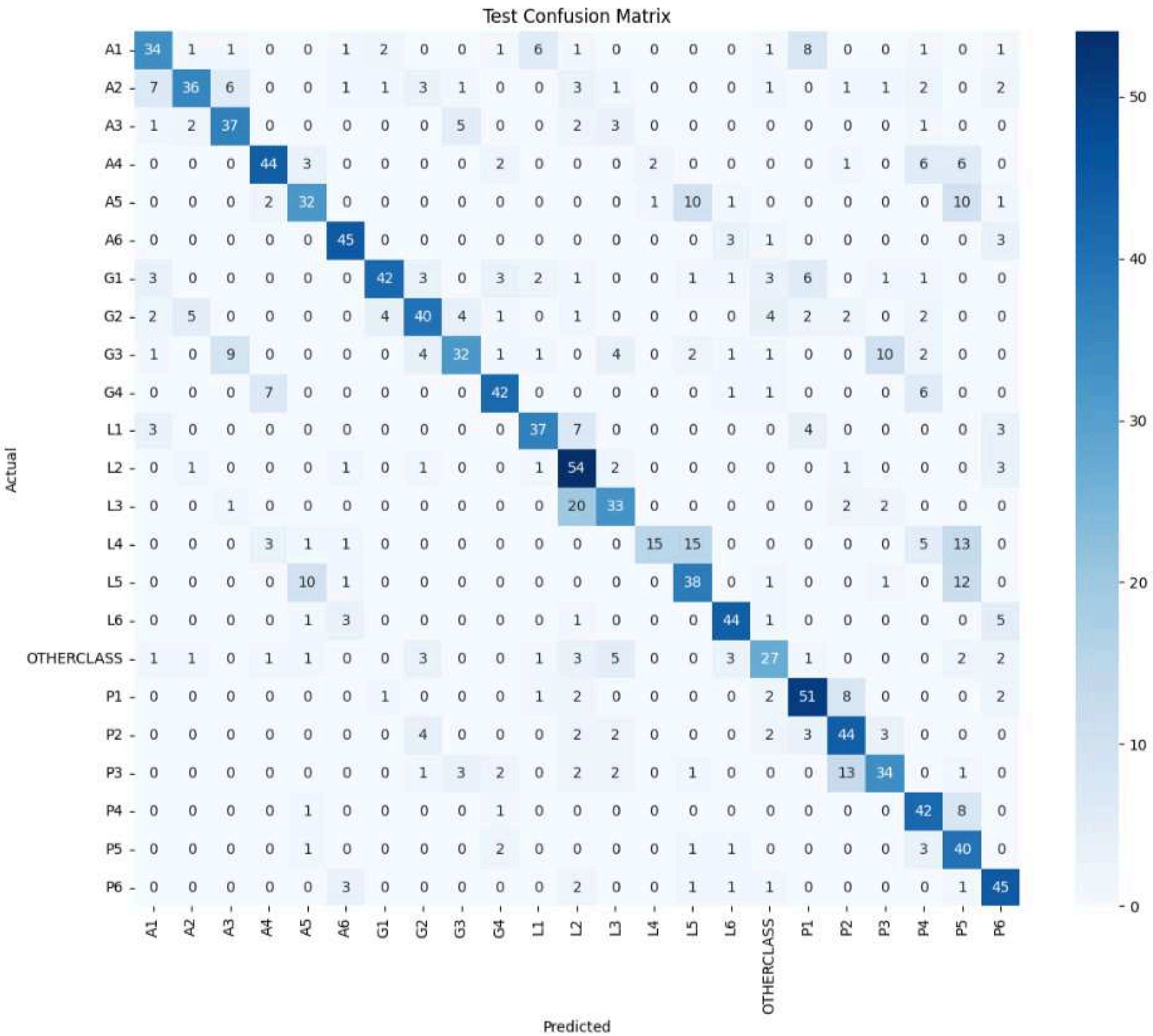
- **Transfer Learning** - Using **pre-trained vit_b_16 weights** and biases – freezing the early layer, and fine-tuning on the FC layer.

HyperParameters

- **Batch Size** = 16 (balanced for performance and available compute power)
- **Patch Size** = 16 (vit_base expects this)
- **Starting learning rate** = 1e-4: commonly used in medical images with AdamW optimiser.
- **Weight_decay** = 1e-4 (regularisation to prevent overfitting)
- **Scheduler patience** = 3 epochs: Reduce LR if no improvement,
- **Early stopping patience** = 5

Pre-defined hyperparameters (Characteristics of vit_b_16)

- **Dropout Probability** = 0.1
- **Depth** (No. of transformer layers) = 12
- Each transformer layer uses 12

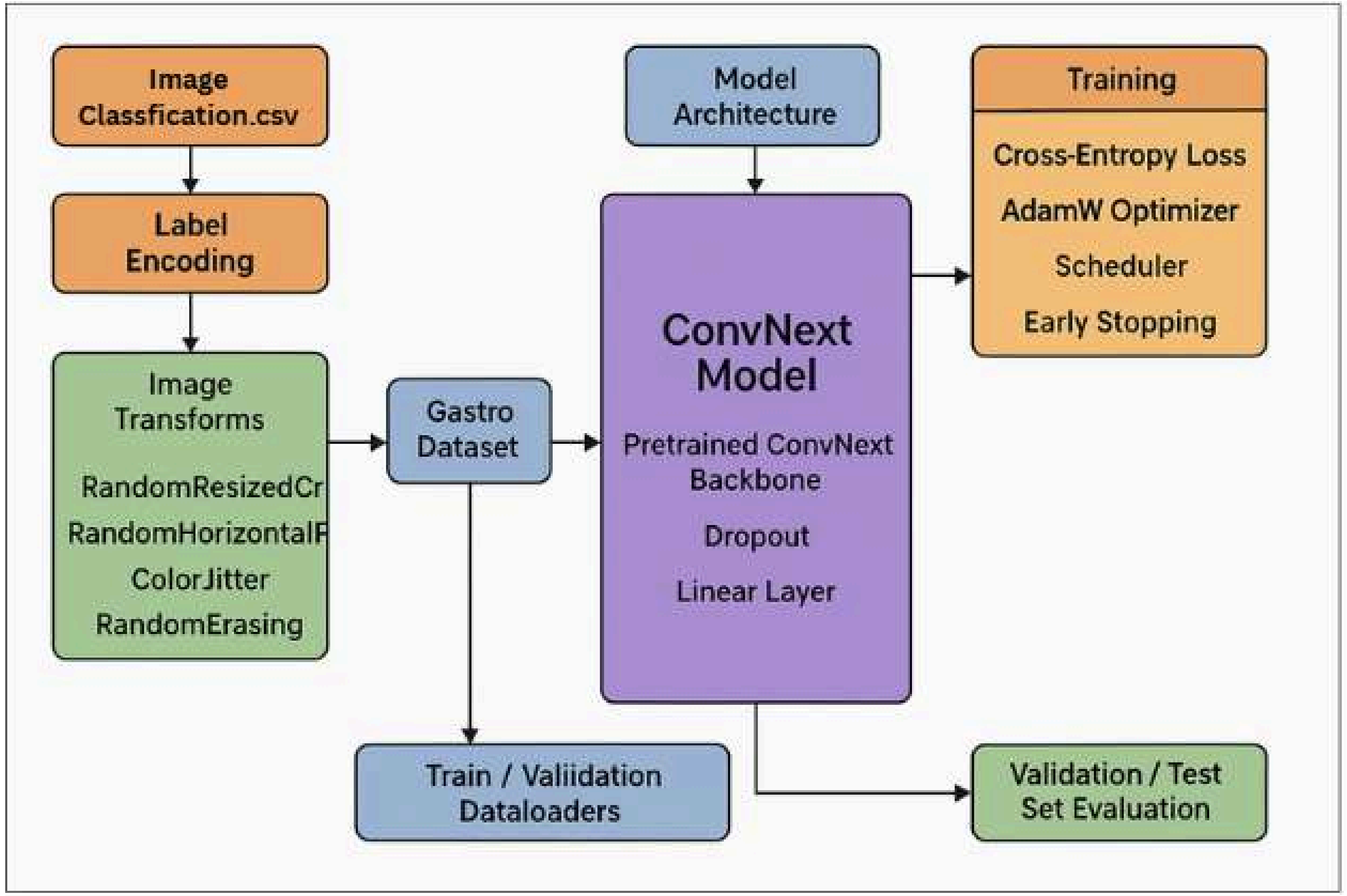

Test Confusion Matrix
Actual
Predicted
A1
A2
A3
A4
A5
A6
G1
G2
G3
G4
L1
L2
L3
L4
L5
L6
OTHERCLASS
P1
P2
P3
P4
P5
P6
0
10
20
30
40
50

Macro F1: 68%

ConvNeXt Large

Our Architecture

- **ConvNeXt_base** as a backbone with no layers frozen.
- Aggressive **augmentations** (random crops, flips, rotations, colour jitter, erasing) to increase **dataset diversity**.
- **Normalization**: Using ImageNet's mean and std_devn ensures the input distribution matches what the pretrained ConvNeXt model expects.

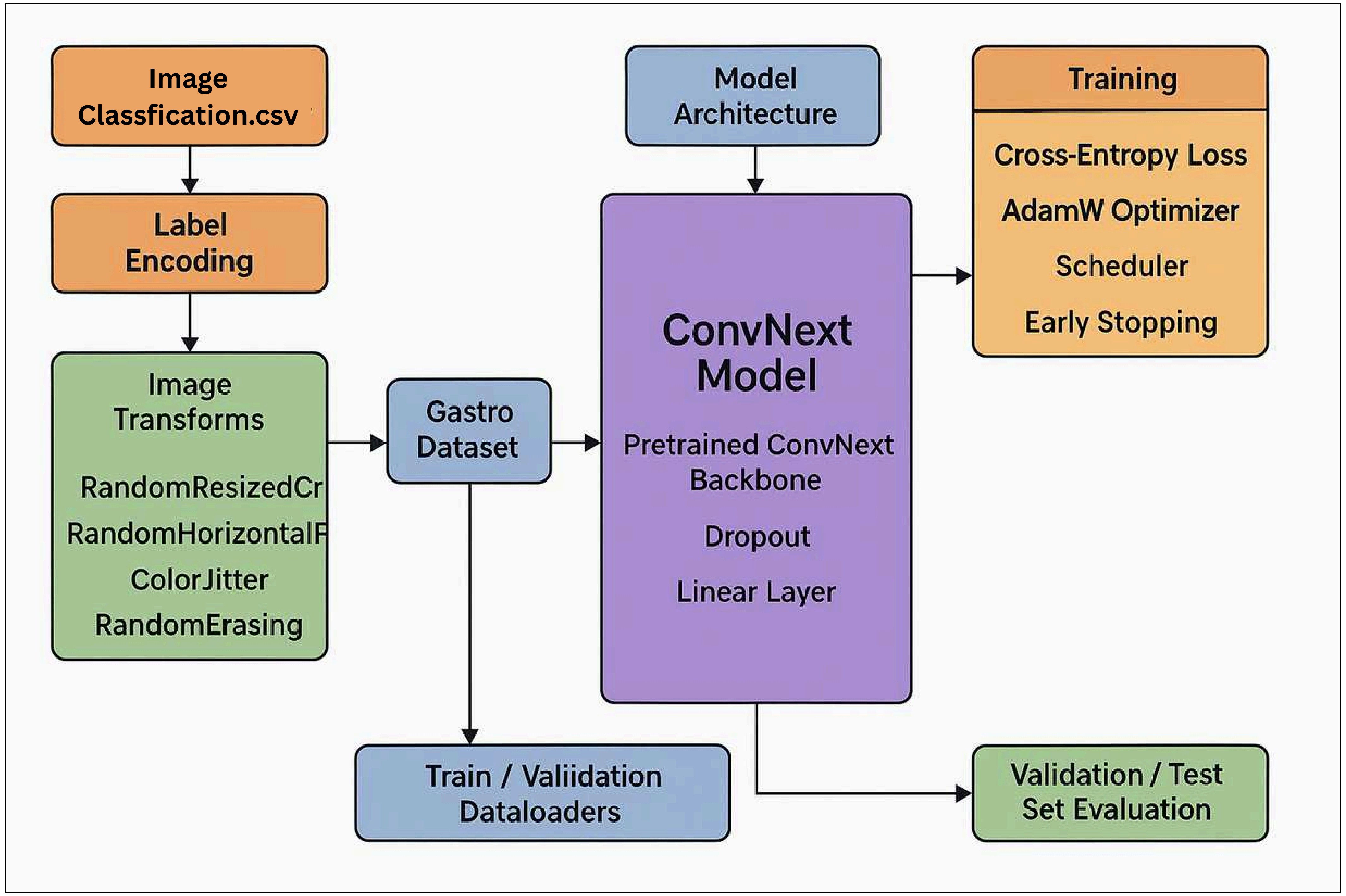


ConvNeXt Large

Our Architecture

- Calculate class weights based on the **frequency of labels** in the **training set** and use them in CrossEntropyLoss. (**Weighted crossEntropyLoss**)

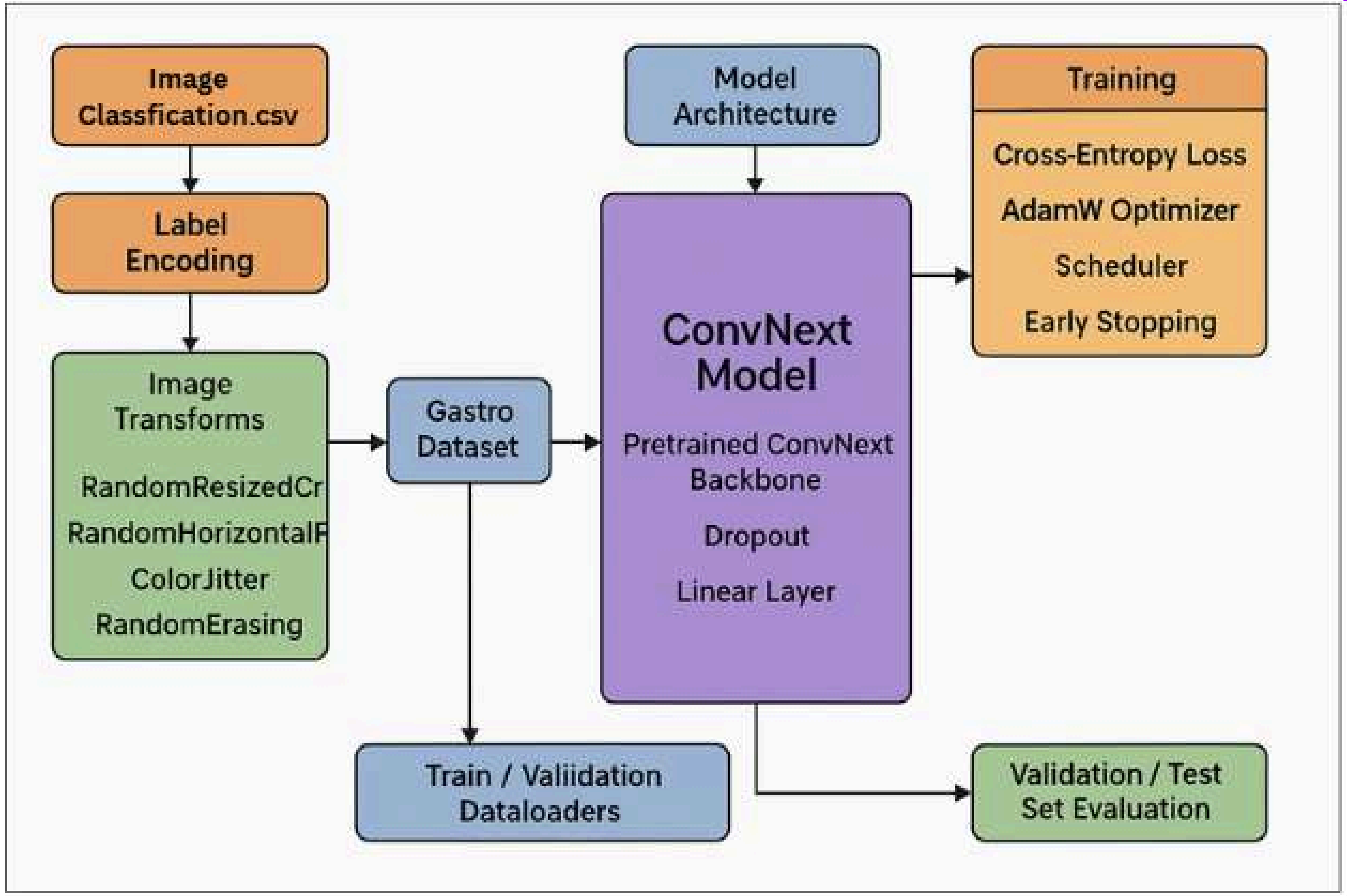
- **Dropout Layer** for Regularisation
- **AdamW** with a 1e-3
- ReduceLROnPlateau scheduler
- **Early stopping**, patience = 10 epoches



ConvNeXt Large

Our Architecture

- Batch_size = 16
- Learning_rate = 0.001
- Num_epochs = 50
- Num_classes = 22
- Image_size = 224 * 224



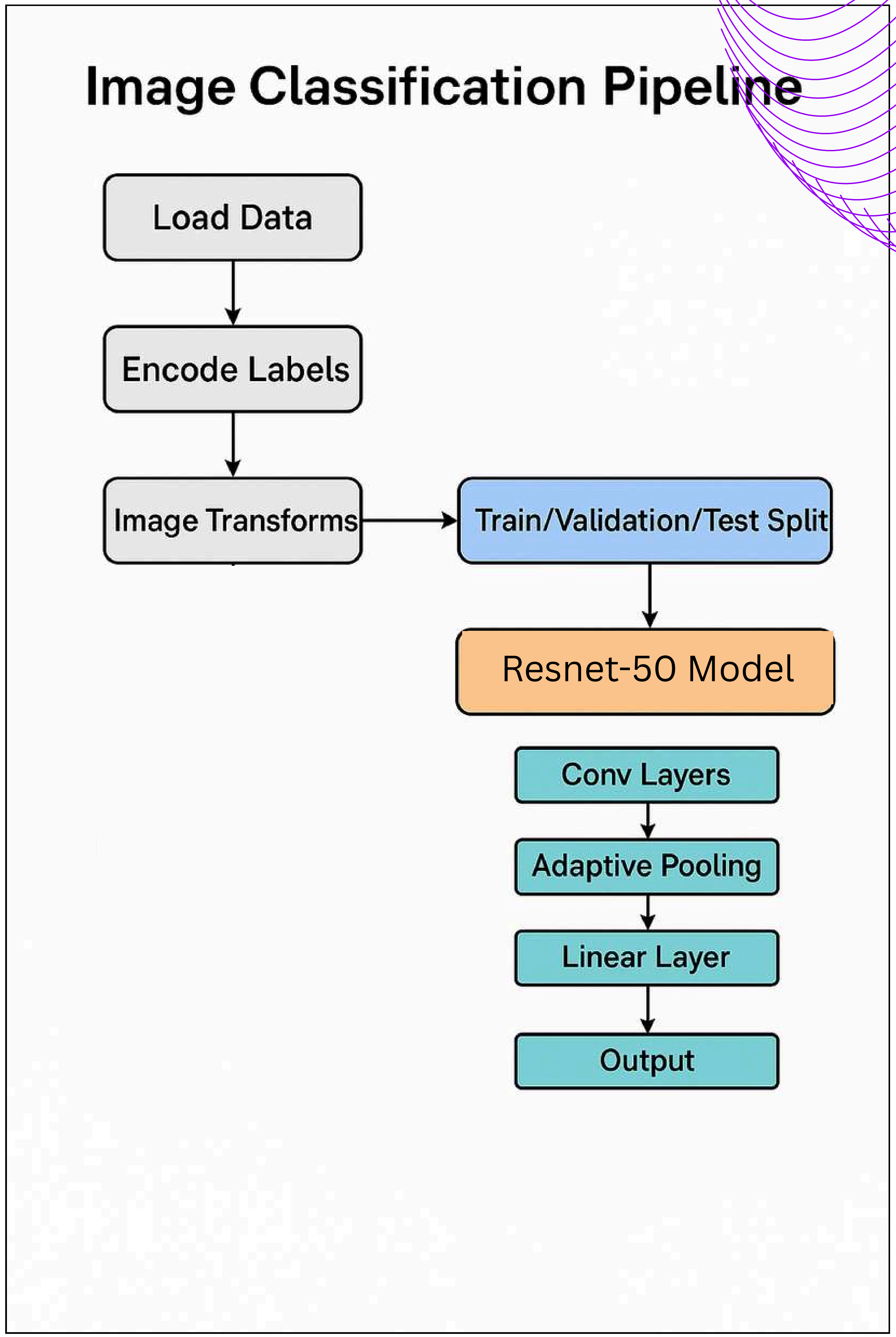
Evaluation

- Test Macro-F1: 61 %

ResNet-50

Our Architecture

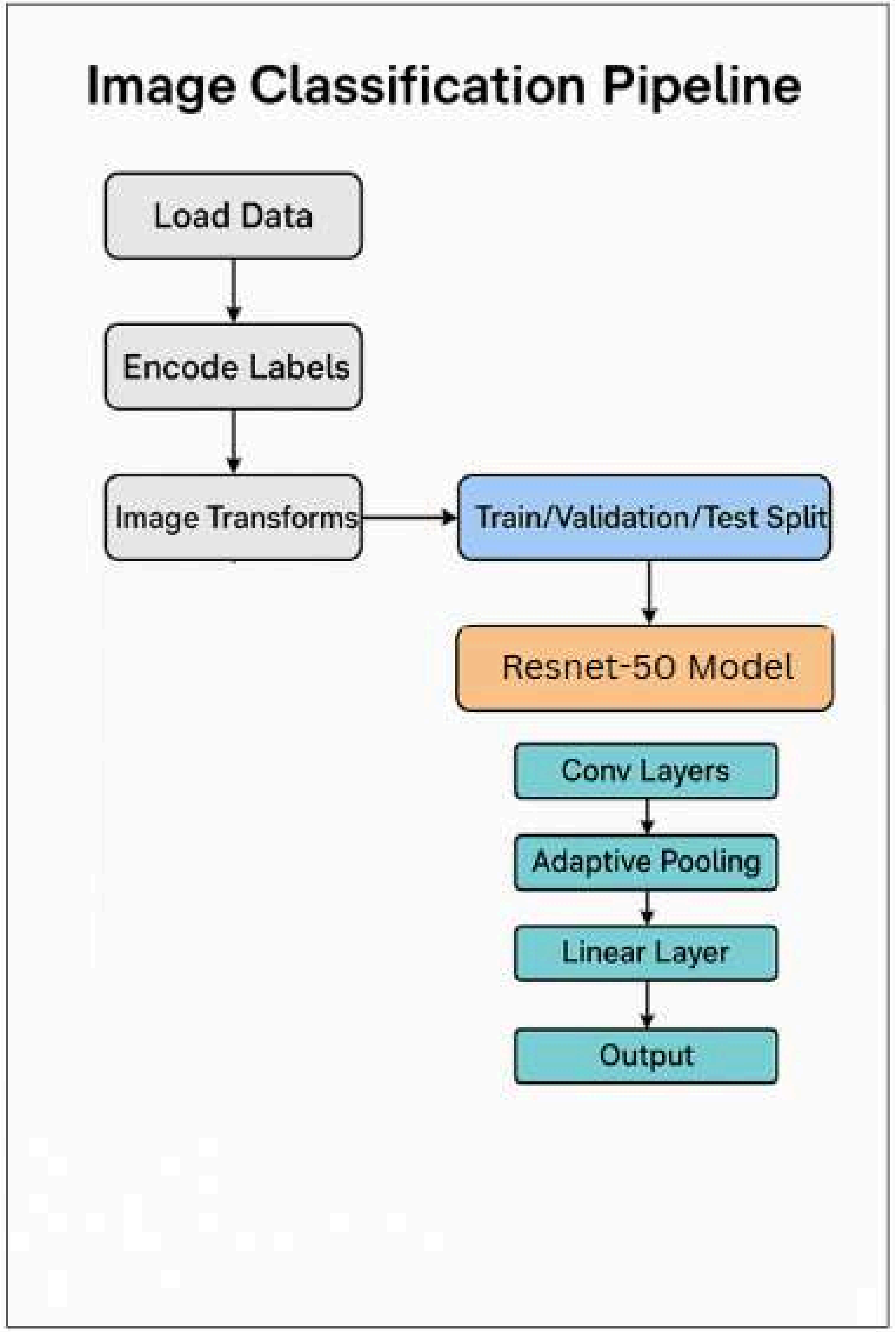
- Pretrained **ResNet-50** Model as the Backbone.
- Applied **data augmentation**: RandomResizedCrop (224x224), RandomHorizontalFlip, ColorJitter, and ImageNet normalisation.
- Addressed **class imbalance** by calculating class weights (inversely proportional to class frequency) and using them in CrossEntropyLoss.



ResNet-50

Our Architecture

- Used AdamW optimiser with lr=0.0001 and weight decay 1e-4. Implemented StepLR scheduler to reduce learning rate by 0.1 every 5 epochs.
- Implemented early stopping with patience=5.



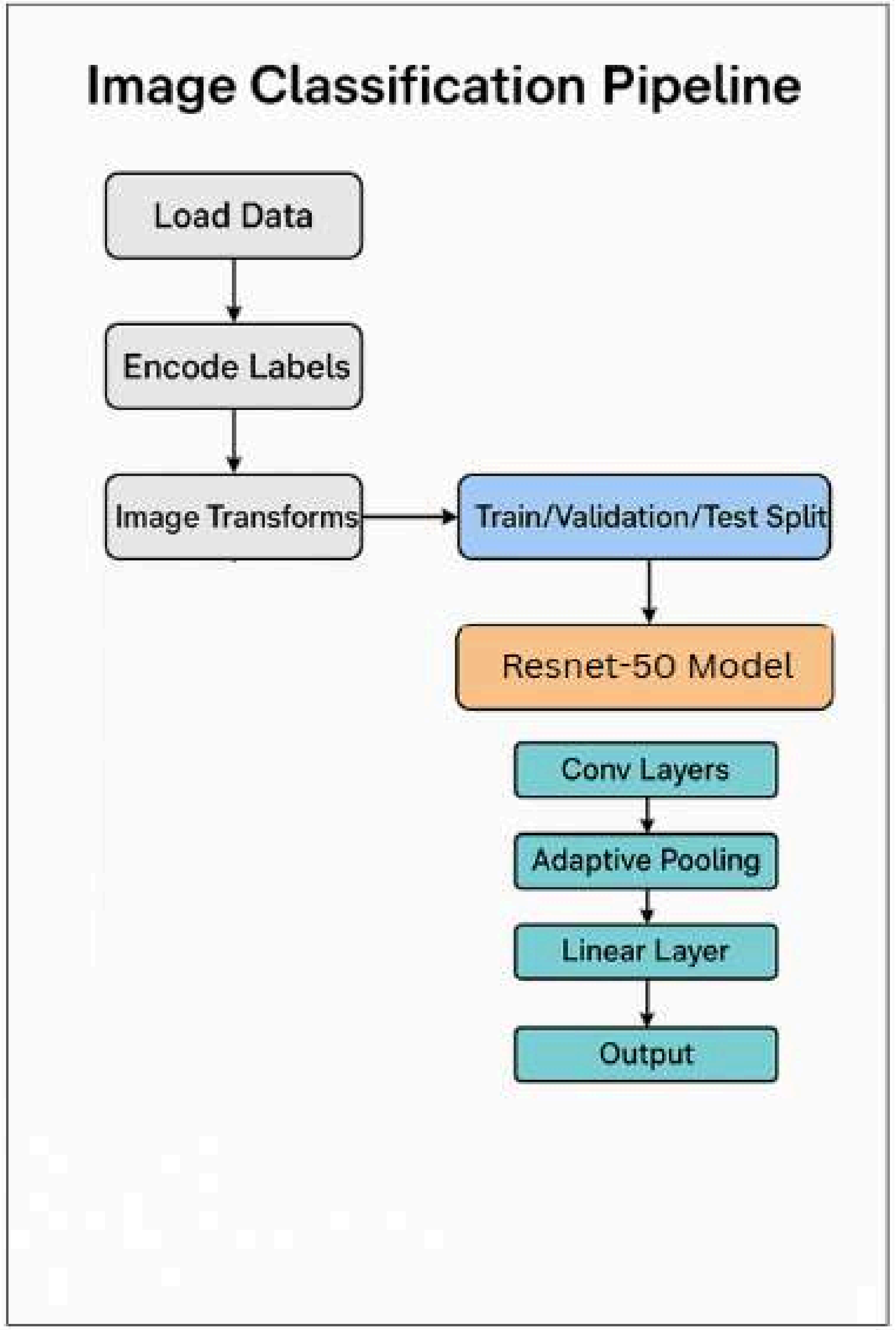
ResNet-50

Our Architecture

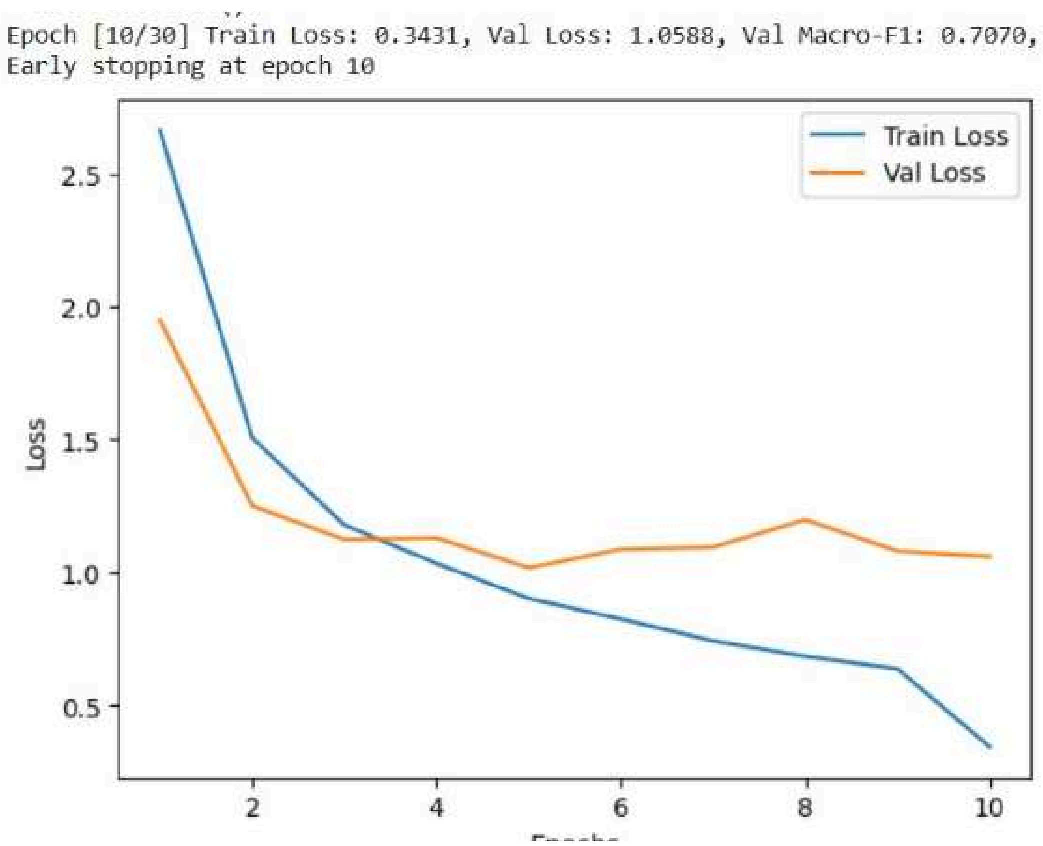
- Batch_Size = 16
- Num_Classes = 22
- Epochs = 30

Evaluation

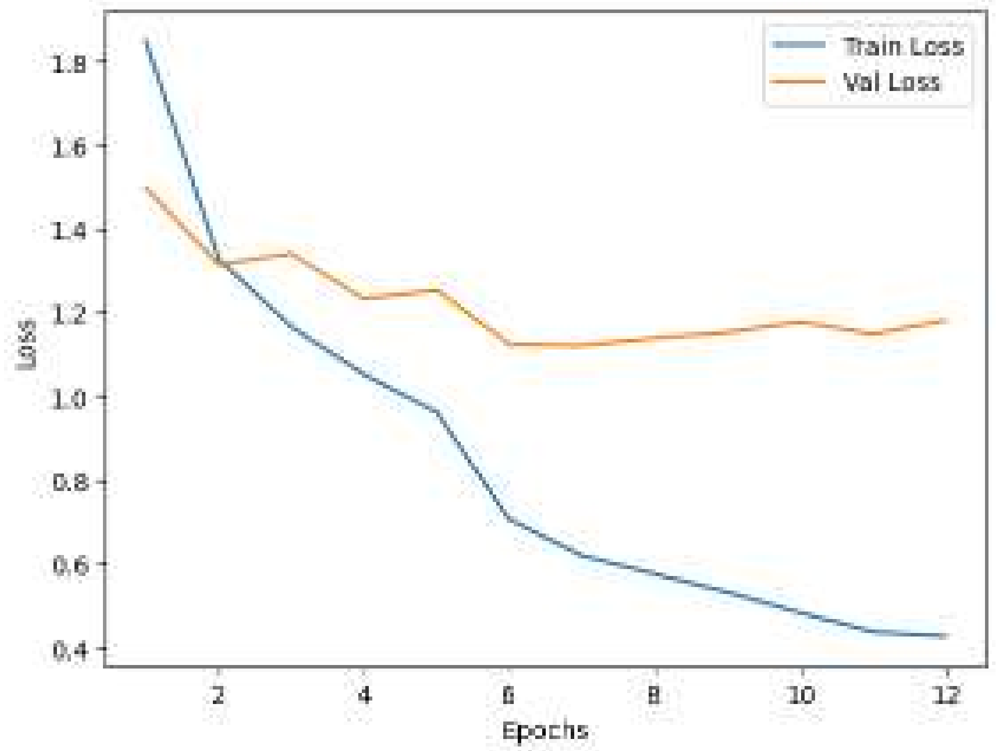
- Test Macro F1 = 65 %
- Accuracy = 66 %



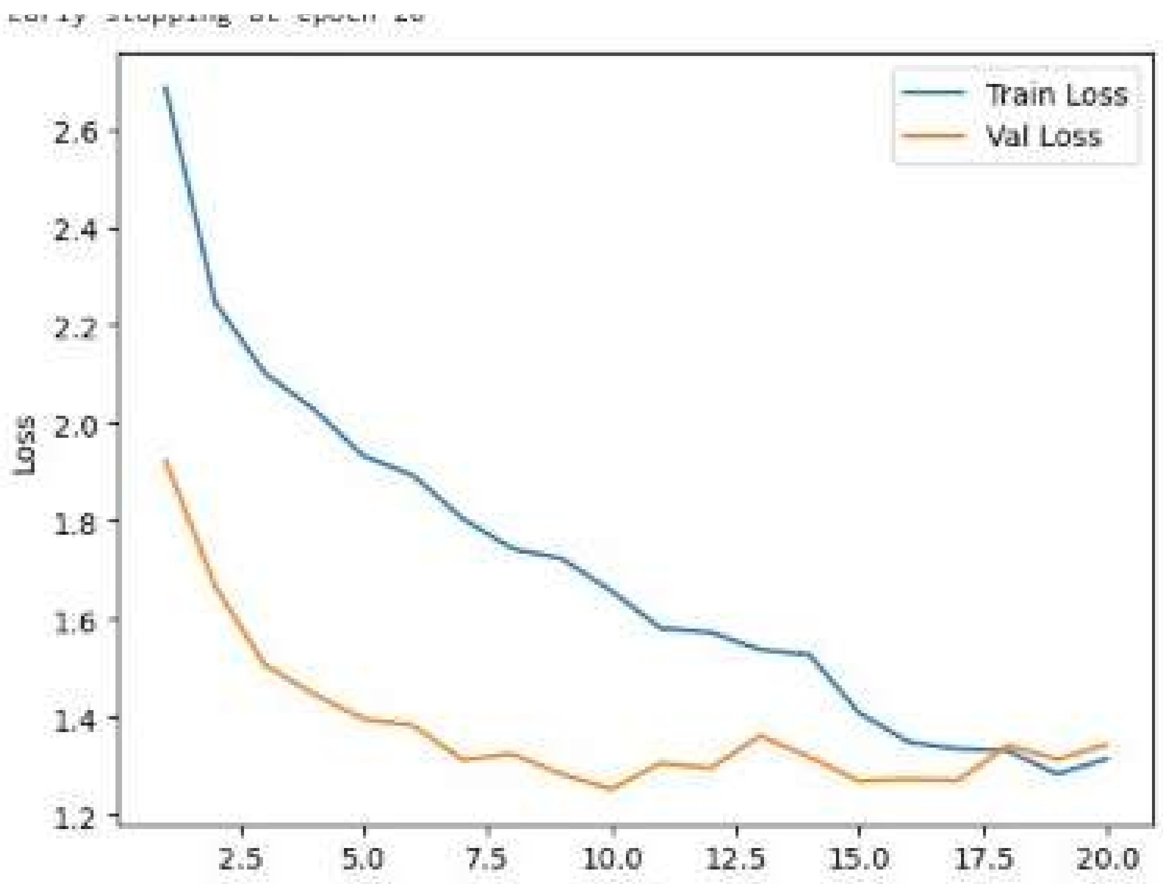
ViT



ConvNeXt Large



ResNet



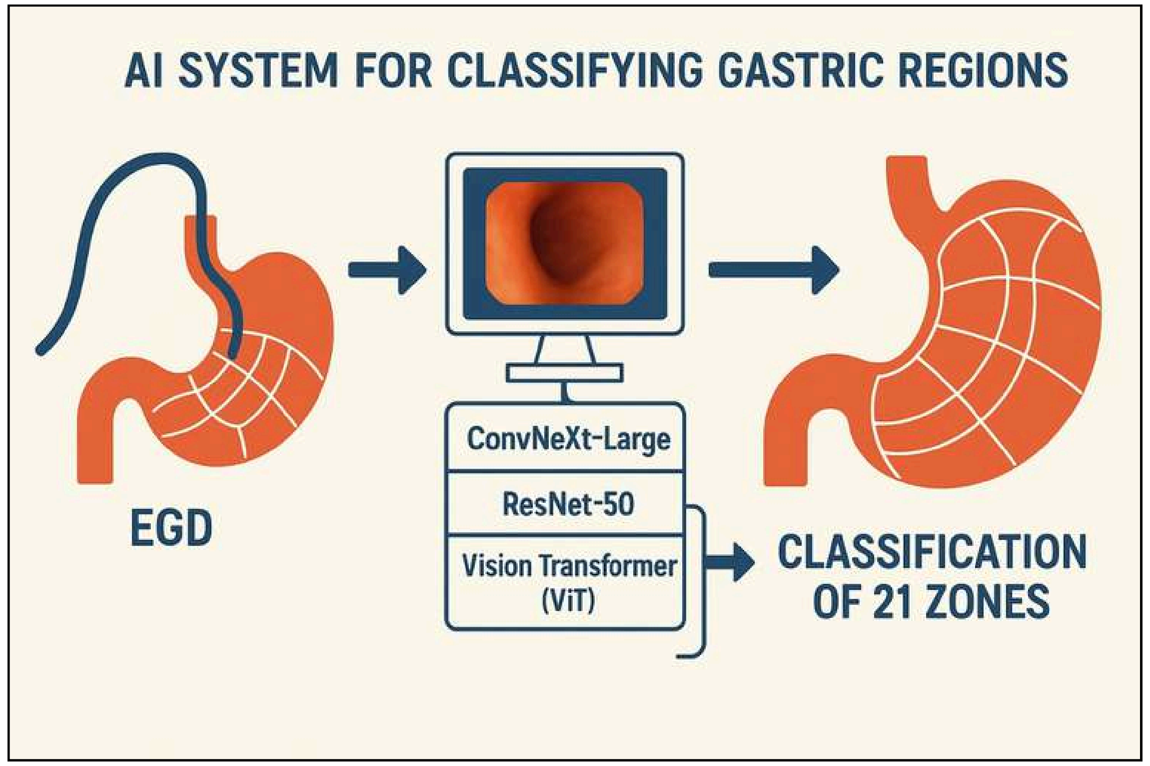
Deployability

- The ultimate goal: a **CAD system** which ensures that the doctor **covers ALL anatomical landmarks**, and no region is left out in accordance with the **SSS-Kenshi-Yao Protocol**.

- Moreover, integrated with a **disease-detection model** to detect diseases side by side – Already lot of models exist to solve this problem.

Outcomes

- Reduces diagnostic errors.
- Ensures **complete coverage of 21 anatomical landmarks**.
- **Standardises reporting** and reduces **inter-operator variability**.



System Architecture

- Input: Either a live stream (video frames) or buffered input to the CAD
- **Inference:**
  - Run inference every n-frames or m-seconds.
  - Tracks which anatomical landmarks have been seen, and which are yet to be seen. Notification (screen pop-up) in real time if any landmark missed wrt. the SSS protocol.
  - **Disease Detection Model:** runs in parallel, triggers alerts for possible findings (eg. ulcers, tumors, bleeding)

Deployability

Considerations

- **Real-Time Inference:** Run models on edge devices (eg, NVIDIA Jetson Xavier/Orin) co-located with endoscopy hardware.
  - Ensure model decisions don't delay or block visualisation
- **Frame Acquisition & Processing:** Use a **frame grabber card** (eg. Epiphan, Blackmagic) to capture **high-quality video feed**.

.

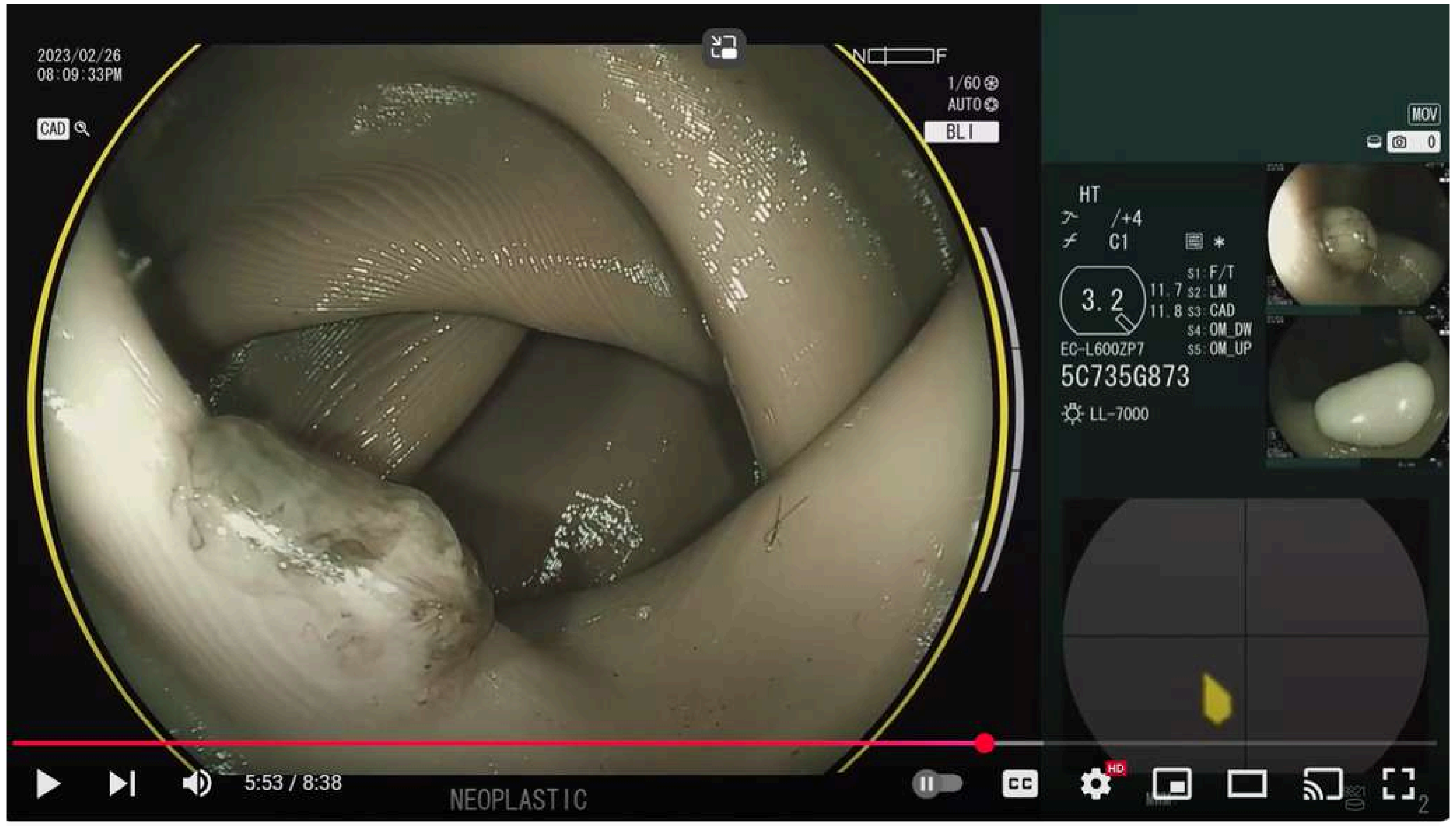
CAD EYE (https://www.youtube.com/watch?v=eUmjOxLVv5s)

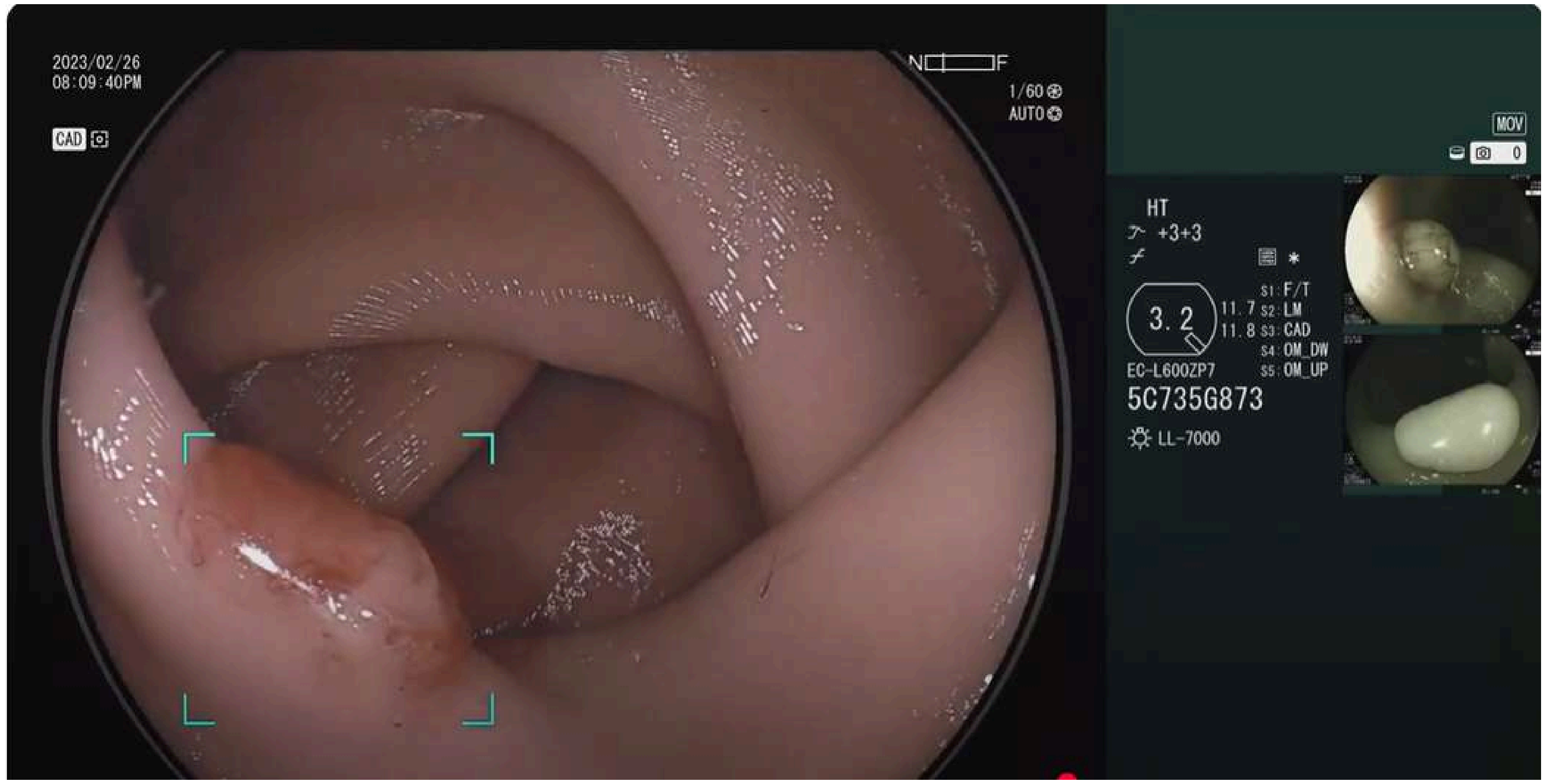
- An AI system to **detect and characterise polyps** during endoscopy/colonoscopy.
- **Helps us get an idea about how any endoscopic CAD system should be.**
- **"The ideal interface for me is the interface that actually is invisible" -** Prof. Dr. Pieter Dewint.

Regulatory Paths

- Prototype → IRB approval → Clinical trial → CE/US FDA clearance as Class II device (if marketed)

Deployability





What a good CAD system looks like - CAD EYE

Acknowledgements

- Dr. Sanjeev Batia - Laparoscopic Surgeon, Global Health Care Clinic Chandigarh Sector 21-C
- Dr. Siddharth - Assistant Professor, Plaksha University
- Pushpinder Singh, Indrani Purkayastha, Sowmya Mallu - Course Teaching Assistants
- Tanmay Rai Nanda , Utkarsh Agarwal - Student Teaching Assistants