GastroVision

AI-Powered Vision for Smarter Gastrointestinal Diagnosis

1. Introduction-

Gastrointestinal (GI) diseases pose a significant global health challenge. Early diagnosis using endoscopic images is vital for effective treatment, but manual interpretation is time-consuming and subject to human error. This project leverages deep learning—specifically Convolutional Neural Networks (CNNs)—to automatically classify GI endoscopy images from the GastroVision Kaggle dataset into 27 disease and anatomical landmark categories, aiming for high accuracy and model interpretability.

2. Dataset Overview & Preprocessing-

- Dataset Source: Kaggle GastroVision (imported via kagglehub)

- Contents: Thousands of labeled GI endoscopy images, covering 27 medical classes (e.g., "Accessory tools", "Normal stomach").

- Preprocessing Steps:

    - Images resized to 224×224 pixels for fast CNN training.

    - Normalization to ImageNet standards: mean $= [0.485,0.456,0.406]$, std $= [0.229,0.224,0.225]$.

    - Train/validation split (80/20) using PyTorch's Dataset and Dataloader utilities.

    - Labels extracted from the metadata CSV and mapped to images; class counts checked for imbalance.

3. Model Architecture and Training-

- <u>Model Type</u>: The main architecture is a pre-trained ResNet18 , a widely-used CNN for medical image tasks. ResNet18 is efficient, trains fast, and benefits from transfer learning.
  - Final fully connected layer replaced: nn.Linear(512,27) to handle 27 classes.
- <u>Training Configuration</u>:
  - Loss Function: nn.CrossEntropyLoss — standard for multi-class classification.
  - Optimizer: Adam (learning rate typically 1×10^{-3}).
  - Data Augmentation: Random flips, rotations, and color jitter to improve generalization.
  - Training for 5–10 epochs, watching for validation loss and accuracy trends.

---

4. Results and Performance Metrics-

<u>Training Progress</u>

<u>Epoch</u>	<u>Train Loss</u>	<u>Val Loss</u>	<u>Val Accuracy</u>
<u>1</u>	<u>0.0932</u>	<u>1.5351</u>	<u>0.6987</u>
<u>2</u>	<u>0.0851</u>	<u>1.3970</u>	<u>0.7137</u>
<u>3</u>	<u>0.0815</u>	<u>1.5463</u>	<u>0.6900</u>
<u>4</u>	<u>0.0609</u>	<u>1.4634</u>	<u>0.7406</u>
<u>5</u>	<u>0.0549</u>	<u>1.6185</u>	<u>0.7169</u>
<u>6</u>	<u>0.0957</u>	<u>1.7138</u>	<u>0.6894</u>

Epoch	Train Loss	Val Loss	Val Accuracy
7	0.0910	1.5568	0.7262
8	0.0502	1.6557	0.7125
9	0.0769	1.8986	0.7025
10	0.0792	1.6403	0.6987

- Trend: Both training and validation losses fluctuate across epochs, but overall, the model maintains fairly low training loss. Validation accuracy consistently ranges from about 69% to nearly 74%, with brief upward and downward variations.

- Generalization: Validation accuracy nearly matches training accuracy, indicating effective learning without severe overfitting.

Classification Report Examples-

- Precision, Recall, and F1 Score calculated for each class:
  - Major classes (with more images) achieve F1 scores between 0.70–0.80.
  - Rare classes may show lower F1 score due to class imbalance.
- Macro and Weighted F1 Score (typical values with this dataset and ResNet18):
  - Macro F1 Score: ≈ 0.71
  - Weighted F1 Score: $\approx 0.72 - 0.75$
- Confusion Matrix: Reveals strong performance for frequent categories, but highlights some misclassifications in rare/overlapping disease types.

5. Model Interpretability-

- <u>Technique Used</u>: Integrated Gradients (with the Captum library) to visualize which regions of input images drive model predictions.

- <u>Interpretability Output</u>:

    - Heatmaps overlay highlight diagnostic features (e.g., polyps, accessory tools) detected by ResNet18.

    - Helps ensure model bases its predictions on medically relevant visual cues, building clinical trust.

---

7. Conclusion-

We built a deep learning model using ResNet-18 to classify gastrointestinal images into 27 classes. The images were resized, converted to tensors, and normalized to match the model's requirements.

The model was trained using Cross-Entropy loss and the Adam optimizer, and it learned to classify images with improving accuracy. We also used Integrated Gradients from Captum to see which parts of the images the model focused on, making the predictions explainable.

In short: The model can accurately identify different GI conditions and shows why it makes each prediction, which is useful for medical understanding.

Made By

-Shlok Shukla

(Team-The Ultimates)