GASTROINTESTINAL IMAGE CLASSIFICATION

Introduction

Deep learning enables automatic analysis of medical images. In gastrointestinal endoscopy, accurate classification leads to early disease detection and better patient outcomes.

This project uses FastAI and a pretrained ResNet50 model to classify endoscopic images from the Kvasir v2 dataset into 8 different gastrointestinal conditions.

The goal was to build a robust model using data augmentation and suitable metrics for imbalanced classes.

Dataset – Kvasir v2

The Kvasir v2 dataset is a public collection of gastrointestinal endoscopic images provided by Simula Research Laboratory and Telemark Hospital in Norway.

It includes 8 categories:

- Dyed-lifted-polyps
- Esophagitis
- Normal-cecum
- Normal-pylorus
- Normal-z-line
- Polyp
- Ulcerative-colitis
- Barrett’s esophagus

Data Augmentation

To improve generalization and reduce overfitting, several image augmentations were applied using the Albumentations library:

- CLAHE (contrast enhancement)
- Random brightness and contrast
- Gaussian blur
- Horizontal flip and rotation
- Resize to 224×224
- Normalization with ImageNet statistics

These transformations were integrated with FastAI using a custom AlbumentationsTransform class.

Model Training

The training process had two phases:

- Phase 1 – Frozen: 10 epochs with learning rate 1e-3
- Phase 2 – Fine-tuning: 20 additional epochs with discriminative learning rate (slice(1e-6, 1e-4))
- A SaveModelCallback was used to save the model with the lowest validation loss.

Metrics Used

- Accuracy
- Macro Precision
- Macro Recall
- Macro F1-score

Macro averaging gives equal importance to all classes. This is crucial in medical diagnosis, where even minority classes may represent critical conditions.

Results

The final model achieved a Macro F1-score of 0.9119, indicating balanced performance across all classes.

Most predictions were accurate, with only a few consistent misclassifications:

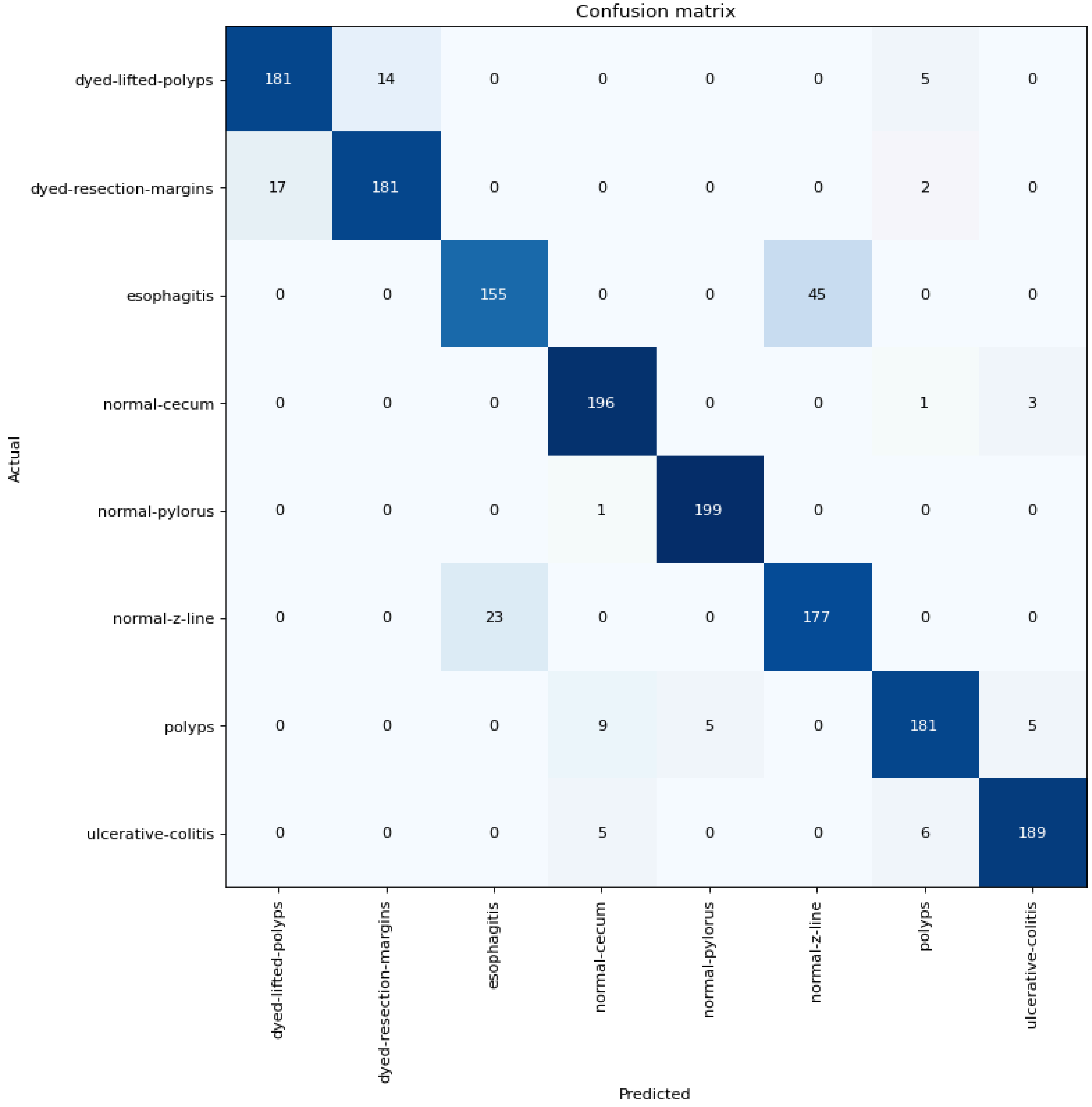
- Esophagitis → Normal-z-line (45 times)
- Ulcerative colitis → Polyp (11 times)

Confusion Matrix

The confusion matrix shows strong performance, with most values concentrated along the diagonal.

Notable issues:

- Esophagitis was frequently misclassified as Normal-z-line
- Normal-z-line had lower precision (0.7973), often confused with Esophagitis



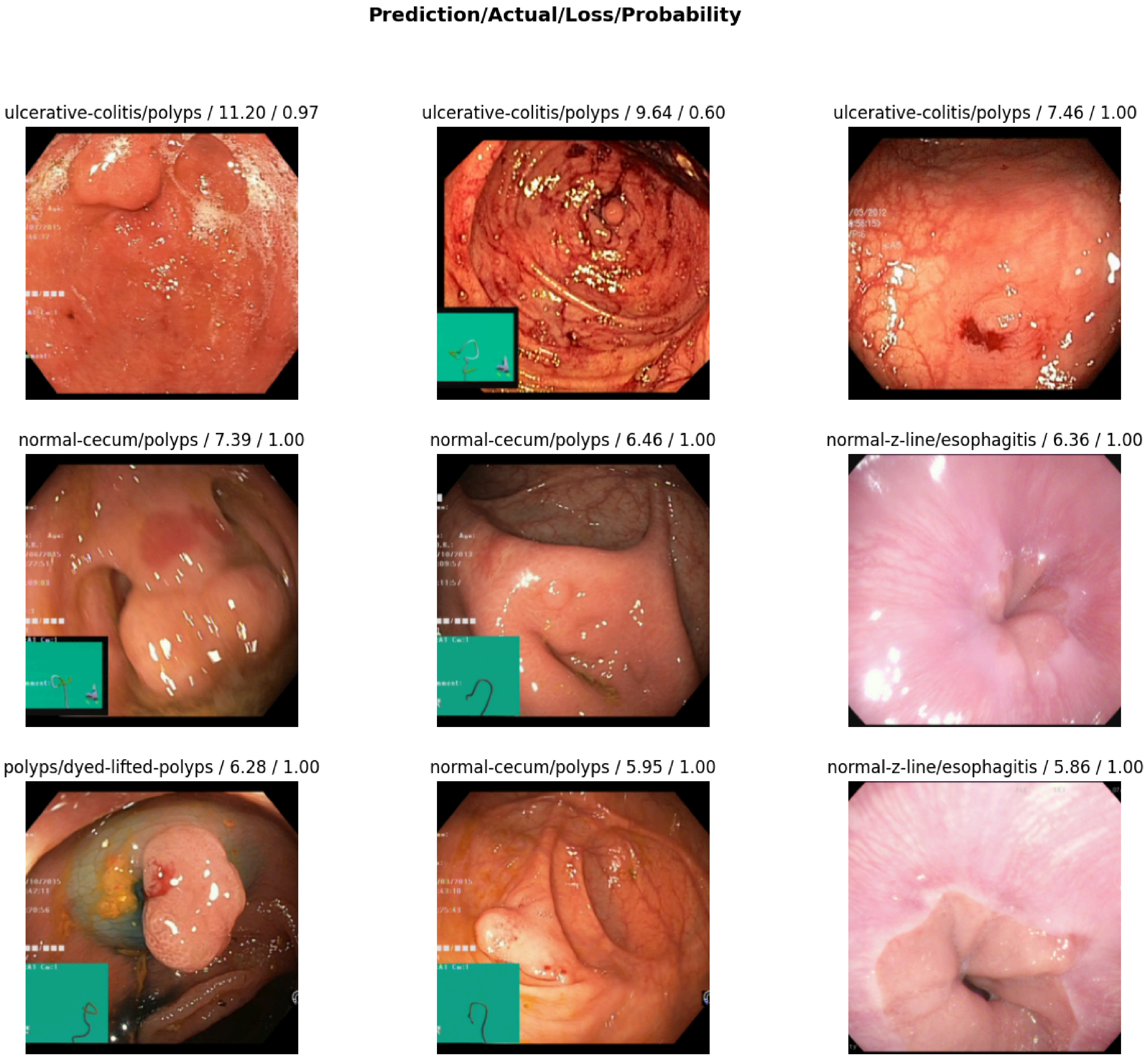
Top Losses

Top-loss examples showed:

- Highly ambiguous cases
- High-confidence incorrect predictions

Suggestions for improvement:

- Use label smoothing or focal loss to reduce overconfidence in predictions



Per-Class Metrics

Per-class evaluation highlighted overall balance, with certain categories showing slightly lower performance:

- Esophagitis and Normal-z-line showed more confusion
- Macro averaging ensured fair evaluation across all classes.

class	precision	recall	f1_score	accuracy	TP	TN	FP	FN
dyed-lifted-polyps	0.9141	0.905	0.9095	0.9119	181	1383	17	19
dyed-resection-margins	0.9282	0.905	0.9165	0.9119	181	1386	14	19
esophagitis	0.8708	0.775	0.8201	0.9119	155	1377	23	45
normal-cecum	0.9289	0.98	0.9538	0.9119	196	1385	15	4
normal-pylorus	0.9755	0.995	0.9851	0.9119	199	1395	5	1
normal-z-line	0.7973	0.885	0.8389	0.9119	177	1355	45	23
polyps	0.9282	0.905	0.9165	0.9119	181	1386	14	19
ulcerative-colitis	0.9594	0.945	0.9521	0.9119	189	1392	8	11

Conclusion

This project demonstrated the effectiveness of deep learning for classifying gastrointestinal endoscopic images.

This system is a promising step toward AI-assisted diagnosis in gastroenterology.

By enabling faster and more consistent image interpretation, such tools can support clinicians in delivering timely and accurate diagnoses.