Universidad Panamericana

UNIVERSIDAD
PANAMERICANA

Intelligent Agents

"Multi-class classification of polyps"

Pablo Ivan Perez Jauregui

02/07/25

Introduction

In this project, we set out to build a reliable and accurate classifier for detecting three important gastrointestinal conditions — esophagitis, polyps, and ulcerative colitis — from endoscopic images. Using transfer learning with MobileNetV2, we aimed not just for good accuracy, but also to experiment with a custom preprocessing pipeline to enhance image quality and reduce noise/artifacts.

We approached this with the mindset of exploring multiple preprocessing techniques and fine-tuning strategies to squeeze the best performance out of a relatively small but diverse dataset.

Dataset Description

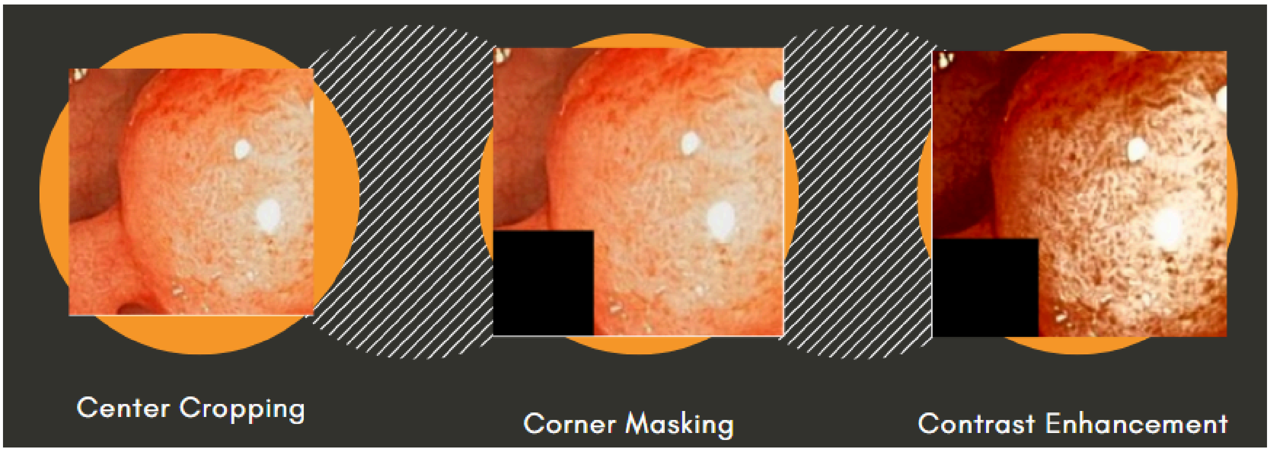
I used the Kvasir dataset v2, which contains labeled images from various gastrointestinal disorders. For this study, we focused on three categories with 1000 images each — a balanced dataset of 3000 images in total. The images were split 80/20 for training and validation respectively.

Given the variability in image conditions (lighting, angle, noise), we knew preprocessing would be key for stable performance.

Method Details

To tackle the challenges in image quality, I designed a custom preprocessing pipeline including:

- Center cropping to standardize input sizes and focus on the region of interest.
- Removal of the bottom-left corner pixels, which we noticed often contained distracting artifacts.
- Histogram equalization in YCrCb color space to enhance contrast and highlight subtle features.



On the modeling side, we used MobileNetV2 pretrained on ImageNet as a backbone, freezing it initially to leverage learned general visual features. The classification head was custom built with dropout to reduce overfitting.

We trained the model in two stages: first freezing the base to stabilize training, then fine-tuning the last 30 layers with a lower learning rate to adapt to domain-specific nuances.

Throughout training, we monitored validation loss and accuracy closely, employing early stopping and model checkpointing to avoid overfitting and ensure we saved the best performing weights.

Results

The experiments paid off — the final model achieved a strong 94.0% accuracy on the validation set.

Looking deeper, the confusion matrix and classification report showed:

- Esophagitis class had stellar precision and recall around 97.5%, indicating the model was excellent at correctly identifying these cases with minimal false alarms.
- Polyps showed solid performance too, with a slight drop in recall (90%), suggesting some cases were trickier to catch.
- Ulcerative colitis was classified with a balanced precision/recall over 90%, demonstrating robust model generalization.

The detailed metrics reinforced that the model was well-calibrated, not just overall but per class — a crucial factor for medical applications.

```
📊 Matriz de confusión:
[[195   1   4]
 [  5 180  15]
 [  0  11 189]]
```

```
📝 Reporte de clasificación:
                   precision    recall  f1-score   support

     esophagitis     0.9750    0.9750    0.9750       200
          polyps     0.9375    0.9000    0.9184       200
ulcerative-colitis   0.9087    0.9450    0.9265       200

        accuracy                         0.9400       600
       macro avg     0.9404    0.9400    0.9399       600
    weighted avg     0.9404    0.9400    0.9399       600
```

```
Clase: esophagitis
TP: 195, FP: 5, FN: 5, TN: 395
Precisión: 0.9750, Recall: 0.9750, Specificity: 0.9875
Accuracy: 0.9833, F1-score: 0.9750

Clase: polyps
TP: 180, FP: 12, FN: 20, TN: 388
Precisión: 0.9375, Recall: 0.9000, Specificity: 0.9700
Accuracy: 0.9467, F1-score: 0.9184

Clase: ulcerative-colitis
TP: 189, FP: 19, FN: 11, TN: 381
Precisión: 0.9087, Recall: 0.9450, Specificity: 0.9525
Accuracy: 0.9500, F1-score: 0.9265
```

Conclusion

In summary, this project was a rewarding journey of experimenting with preprocessing tricks and transfer learning fine-tuning. The custom image enhancements clearly helped the model focus on the right features and boosted performance.

While the dataset size limited us from pushing even higher accuracies, the model's performance is promising for real-world use cases. Future directions might explore additional data augmentation, ensemble learning, or even newer architectures like EfficientNet or vision transformers to push boundaries further.

Overall, this work highlights the power of combining domain-informed preprocessing with modern deep learning to tackle challenging medical image classification tasks.