1 Description

Dataset name	The Hyper Kvasir Dataset
Link	Kaggle
License	CC BY 4.0
Medical discipline	Gastroenterology
Medical procedure	Endoscopy
Multi-class problem	✓
Multi-label problem	✗

This is the code repository for the Hyper-Kvasir dataset which is the largest publicly released gastrointestinal tract image dataset. In total, the dataset contains 110,079 images and 373 videos where it captures anatomical landmarks and pathological and normal findings. The results is more than 1,1 million images and video frames all together. The paper describing the data can be accessed here.

This summary describes the identification of anatomical landmarks in the *upper GI tract*. Classes are:

- pylorus
- retroflex-stomach
- z-line

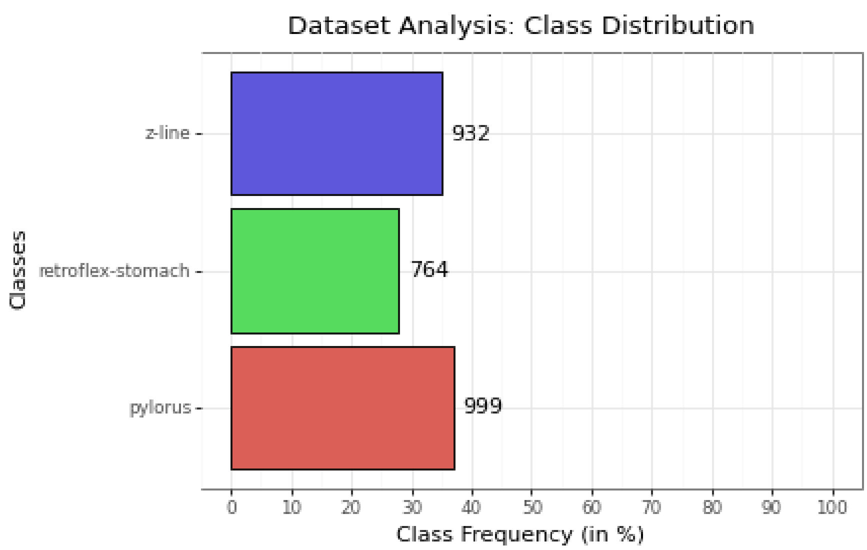


Figure 1: Class distribution: The Hyper Kvasir Dataset(Upper GI tract)

2 Pre-processing

For the purpose at hand, a sub-set of this dataset was used, to tackle the anatomical landmark classification. This dataset provides labeled images for both the lower and upper GI tract. They are treated as two separate classification tasks. In this summary, the classification of the landmarks in the *upper GI tract* is described.

3 Training

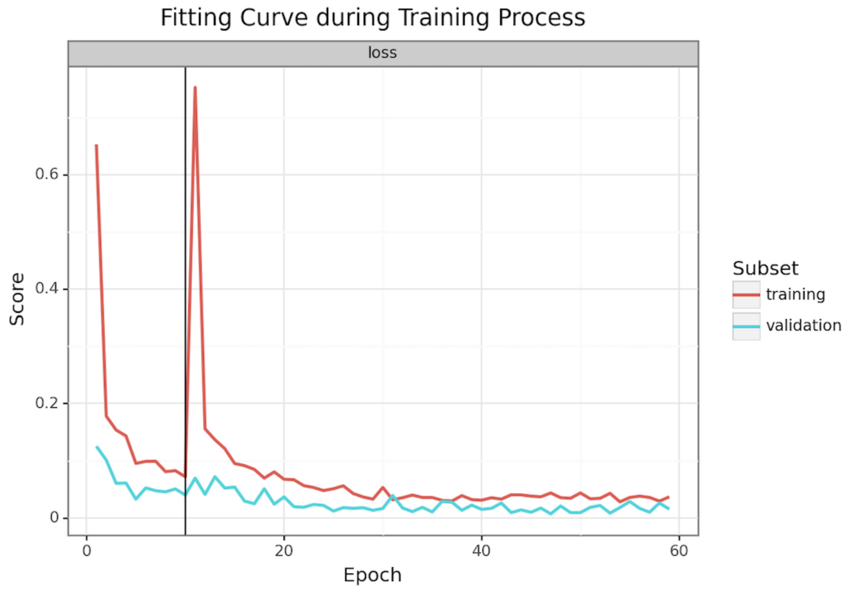


Figure 2: Training: The Hyper Kvasir Dataset(Upper GI tract)

Due to transfer learning, the training loss is peaking as soon as the model's weights are unfrozen for training. Other than that, the training progress shows a smooth reduction of training and validation loss. There is no indicator for overfitting.

4 Results

Figure 3: Metric overview: The Hyper Kvasir Dataset(Upper GI tract)

All metrics in figure 3 and ROC curves in figure 4 underline our model's exceptional classification performance for this task.

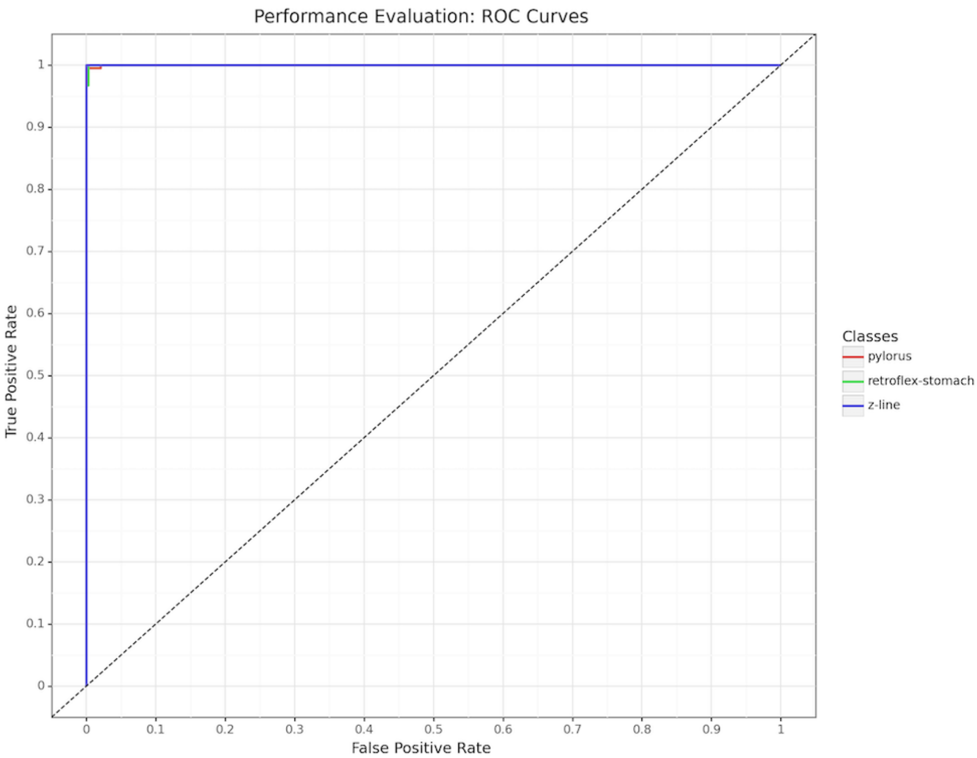


Figure 4: ROC curve: The Hyper Kvasir Dataset(Upper GI tract)

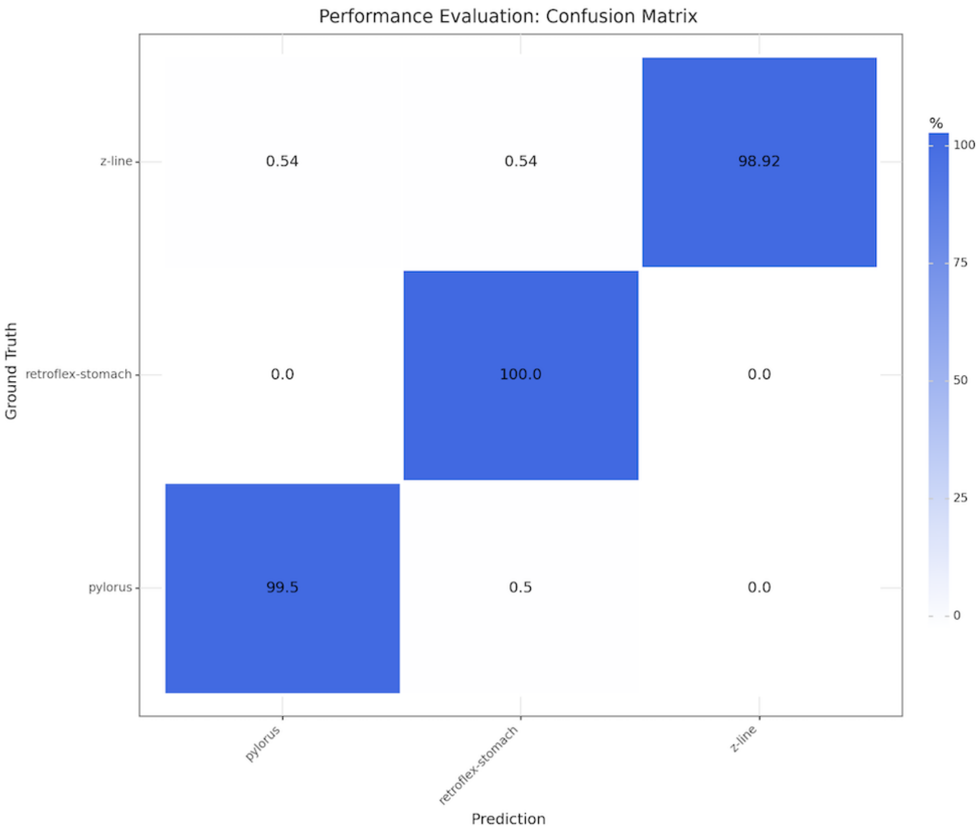


Figure 5: Confusion matrix: The Hyper Kvasir Dataset(Upper GI tract)

5 XAI

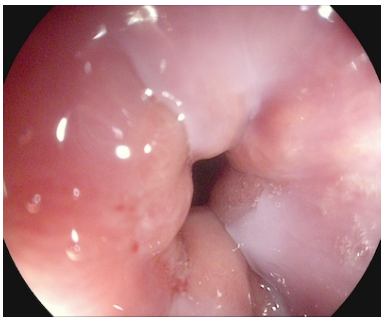
For a endoscopy it is common that medical images display not only tissue, but also medical instrumentation. To build a reliable classificator, we need to understand, if our model is basing its decision on the seen tissue or on the visible instruments. Our example **d000a896-1e01-4ade-bf33-69a8a659bdc3** indicates, that both tissue and the instrument are relevant for the model's decision. Since the instrument itself is black, its relevance cannot be determined definitively.

Furthermore, we only generated Grad-CAM images for one sample. To find out, if our model is basing all of its classification decisions on the appearance of a medical instrument, more Grad-CAM images should be considered.

Image: 0a168064-208f-4d4d-bfdd-ee3e552f9c47

Class: z-line

Classified as: z-line (99.3 %)



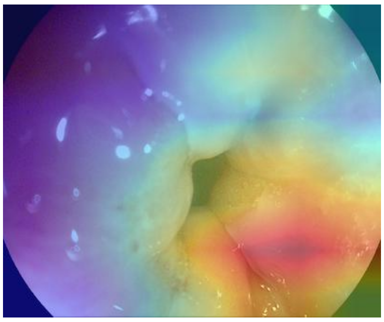
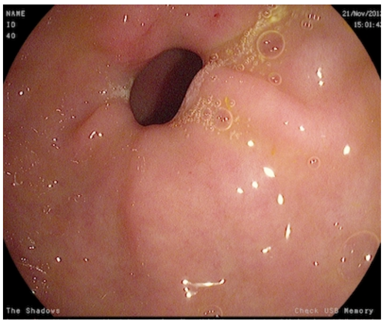


Image: 8c410fb4-1060-49fc-93b0-249ff09d95e6

Class: pylorus

Classified as: pylorus (100.0 %)



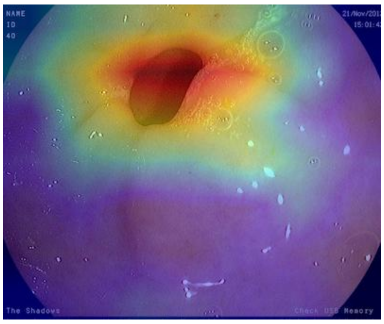
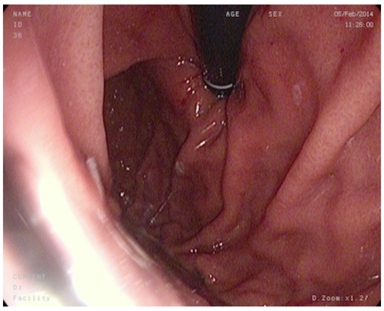
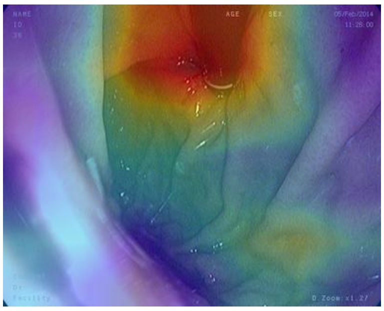


Image: d000a896-1e01-4ade-bf33-69a8a659bdc3

Class: retroflex-stomach

Classified as: retroflex-stomach (100.0 %)





6 Summary

In comparison to the other anatomical landmark classification task for the lower GI tract, this example works on a more balanced dataset (see figure 1). That results in remarkable classification scores.

In this instance, the model seems to partly base its decision upon the appearance of a medical instrument.That is a good example to demonstrate one major challenge in medical image classification: A machine potentially uses all available information. Obviously, that can result in features, a human would never use as foundation for a medically relevant decision. Before accepting a machine-made decision, it should always be considered, how the decision is made.