

A Computer-Aided Detection System for Helicobacter Pylori in Gastric Biopsy Slides

Anonymous Author(s)

ABSTRACT

Helicobacter pylori (*H. pylori*) is a bacterium that causes gastric disease and a group I carcinogen. Early detection is important for treatment and controlling the spread of infection. Histopathology, the most common diagnostic method is time-consuming and labor-intensive, limiting detection efficiency. Although immunohistochemistry (IHC) offers improved accuracy and faster detection, its high cost and equipment requirements restrict its widespread use. This project aims to develop a computer-aided detection system for *H. pylori* in gastric biopsy slides with the goal of improving the efficiency of Histopathology examination. The system uses a combination of computer vision techniques and neural networks to highlight areas of bacterial presence in whole slide images (WSI). We implement 2 approaches here based on past work: Classification and Object Detection. Performance is evaluated using precision and recall, based on intersection with the ground truth WSI regions provided by an expert. We use a train-test split stratified based on the distribution of positive regions to evaluate the entire detection system across 22 biopsy slides. The classifier results show 56% recall and 45% precision on extracted candidate regions. The object detector achieved 52% recall and 58% precision based on DIoU. This study highlights the potential of AI in aiding pathologists to improve the efficiency of histopathology detection.

1 INTRODUCTION

Helicobacter pylori (*H. pylori*) is a bacterium that colonizes the stomach lining and duodenum, causing various diseases, including inflammation, gastroduodenal ulcers, malignancies, and extragastric complications [5]. It has been classified as a Group I carcinogen by the International Agency for Research on Cancer (IARC) [8]. Early detection is crucial for timely treatment, preventing severe complications, and controlling the spread of infection. Current diagnostic methods include both invasive and non-invasive techniques, varying in sensitivity, specificity, and practicality. Histopathology is one of the most common diagnostic approaches. It involves detecting *H. pylori* in stained gastric biopsy slides. However, manual histopathological examination of whole slide images (WSIs) is time-consuming and labor-intensive, limiting detection efficiency. Although immunohistochemistry (IHC) offers improved accuracy and faster detection, its high cost often restricts its widespread use in resource-limited settings [3].

Artificial intelligence (AI) offers promising solutions to these challenges. Convolutional neural networks (CNNs) have demonstrated exceptional capabilities in medical image analysis, enabling automated detection of various objects in histology images. AI-driven approaches have the potential to enhance diagnostic efficiency, alleviating the burden on pathologists and making reliable detection more accessible. This study aims to address the

efficiency limitations of traditional methods by developing an AI-driven computer-aided detection (CAD) tool for *H. pylori* in histopathology slides, to enhance diagnostic efficiency and reduce the strain on healthcare systems. We implement and evaluate two approaches based on past work: Classification and Object Detection.

2 RELATED WORKS

There are multiple relevant past works on the use of AI in histopathology image analysis, some focused on *H. pylori* detection and some on other areas in histology and computer vision. We present here a review of the relevant studies that informed our research.

Several studies have focused on classifying *H. pylori* presence in histopathological images. Ibrahim et al. [7] evaluated five pre-trained CNN models (ResNet-101, DenseNet-201, etc.) using five-fold cross-validation on a small dataset of 204 images, achieving an 89.66% F1 score. However, this study was limited to binary classification of pre-cropped regions and did not address region extraction from WSIs. Similarly, Goncalves et al. [6] did a comparison of 3 pre-trained models on their newly introduced DeepHP dataset. DeepHP is a large-scale collection of histology images from slides labeled for *H. pylori* infection. While useful for identifying gastritis, the labeling of images at the slide level rather than the region level, limits the dataset's applicability for detection. Our work builds upon these studies by addressing automated region extraction from WSIs.

Klein et al. [10] proposed a hybrid approach that integrates image processing and a classification model for efficient *H. pylori* screening. Their method first isolates candidate regions (hot spots) in WSIs using image processing, then it crops these regions into non overlapping patches of 224 by 224 pixels, and uses a CNN model to classify them as positive or negative. They then apply SmoothGrad on the classifier to highlight bacteria within a classified positive hot spot for interpretability. They achieved 100% sensitivity with modest specificity of 66.2% over 88 slides using a threshold of 2 positive tiles. Our classification approach is based on Klein et al.'s work with some differences.

Sirinukunwattana et al. [11] worked on improving gland segmentation in colon tissue to make cancer diagnosis more consistent. They published their dataset of segmentation labeled colon glands for the GlaS challenge. We adapt this in our Object Detection approach for detecting gastric glands, where *H. pylori* colonization typically occurs. By combining segmentation with object detection, we efficiently extract candidate regions.

Prior studies have demonstrated the potential of deep learning in histopathology, but challenges remain in automated region extraction and computational efficiency.

3 DATA GATHERING

Our dataset consists of 22 H&E-stained veterinary gastric biopsy slides at 40x magnification provided by LABOKLIN. The slides

were initially annotated by an expert, who circled regions where *H. pylori* bacteria were present. However, this annotation process was limited. Because of the large size of WSIs, it is impractical for a human expert to exhaustively identify all bacteria, resulting in many *H. pylori* positive regions being missed.

To catch more of the positive regions and improve our dataset quality in a time-efficient manner, we used an active learning approach:

- (1) Candidate patches (256 by 256 pixels) were extracted using template matching as described in the Classification approach.
- (2) Patches were labeled as positive if their intersection with the original expert annotations exceeded **30%** of the annotation area
- (3) A CNN classifier was trained on this dataset.
- (4) The classifier's highest-confidence false positives were manually reviewed by the expert, revealing some of them were in fact true positives.
- (5) These mislabeled examples were relabeled and reintegrated into the training set, improving the patch dataset quality.
- (6) The process was repeated multiple times. The number of positive patches in the dataset grew from around 500 to over 1000 patches.

The dataset likely still contains many mislabeled negatives that are actually positives, but their number has decreased a lot after applying active learning.

4 DATA PREPROCESSING

4.1 Classification

Given the large size of WSIs, processing the entire slide with AI is impractical. Downscaling would work for larger objects but *H. pylori* is a very small bacterium. Instead we focused on extracting regions likely to contain *H. pylori* through an image processing step.

H. pylori primarily infects gastric glands[4], which in biopsy slides appear as white regions surrounded by a purple lining of epithelial cells. To extract these gastric glands, we employed template matching in grayscale using two gastric gland templates at various rotations. The match threshold was set at **0.5**. For computational efficiency, we performed this step on our slides at level 2 (16 times downscaled). The templates used were cropped regions around gastric glands from the training set, downsampled to 16 by 16 images. We then applied non-maximum suppression (NMS) to filter out overlapping matches, and extracted out full-scale patches of $256 \times 256 \times 3$ from the original slide for classification. We filtered out background patches (patches with less than 1% tissue) using thresholds on the saturation channel and variance. This patch size was chosen to be small enough so that highlighting a positive patch is enough for a pathologist to easily determine the location of *H. pylori* within it. Patches were labeled as positive if their intersection with the ground truth labels exceeded **30%** of the annotation area. This included both the original circled annotations and the bounding boxes around expert-classified positive patches.

To ensure compatibility with our classifier, we normalized the extracted patches using ImageNet statistics (mean = [0.485, 0.456,

0.406], std = [0.229, 0.224, 0.225]), as our model was built on pre-trained ImageNet architectures. No data augmentation or color normalization was applied in this approach.

4.2 Object Detection

To efficiently extract relevant tissue patches while minimizing background data, we employed the Sliding Window Patch Inference algorithm. This method enables inference on WSIs by segmenting them into smaller patches using a sliding window approach, eliminating the need to load the entire image into memory. Additionally, this process removed a significant portion of background content, such as empty white regions with little or no tissue, thereby reducing the number of patches requiring further processing.

The labeled data provided in xml format contains one or more "Region" elements, each containing vertices that define a region, for each of these elements we calculate a bounding box, which were then used to locate the regions containing *Helicobacter pylori* (*H. pylori*). We extracted patches of size **640 × 640 pixels**, ensuring that the bacteria remained centered in the patch.

In total, **591 patches** were generated. While some patches contained multiple bounding boxes, most included only a single instance of *H. pylori*. These patches were then subjected to **data augmentation** using a combination of classical image transformation techniques, with both the occurrence and intensity of transformations being randomized:

- **Gaussian Blur:** Applied with a randomly selected intensity to introduce varying levels of focus variation.
- **ISO Noise:** Randomly adjusted noise levels to simulate different illumination conditions.
- **Hue-Saturation-Value (HSV) Adjustments:** Randomized shifts in hue, saturation, and brightness changes to account for potential color variations.
- **Geometric Transformations:** Flipping, rotation, and shifting were applied with **randomly determined parameters**, such as rotation angles and flip directions.

To maximize dataset diversity, each transformation was **randomly applied with varying strengths**. Each original patch was augmented into **three additional versions**, with the specific transformations and their parameters (e.g., rotation angle, flip type, blur intensity) being randomly selected. This augmentation process expanded the dataset to **1,617 patches**, which were then split into **80% training and 20% validation**.

5 METHODS

We followed two different approaches in this project. A classification approach inspired by Klein et al., and an object detection approach.

5.1 Classification

After applying template-matching to extract candidate patches as described in the Data Preprocessing section 4.1, we trained a deep learning classifier to distinguish *H. pylori*-positive patches from negative ones. We used transfer learning with ResNet-18, ResNet-50, and ResNet-101 as feature extractors. These models are pretrained on ImageNet, and we kept their pretrained weights frozen and replaced their fully connected layers with a binary classification

233 MLP. We did not perform any fine-tuning of pretrained model
 234 weights. Figure 1. shows the entire classification pipeline.

235 Our training patch dataset consisted of **72,236** $256 \times 256 \times 3$ im-
 236 ages extracted from the training slides, with **702 positive** and
 237 **71,534 negative** patches. Although the extraction step reduced
 238 the presence of irrelevant patches, class imbalance remained an
 239 issue, as positive samples were inherently less frequent. To address
 240 this, we did not perform any additional resampling but instead
 241 used focal loss, which down-weights easy negative examples and
 242 focuses learning on harder-to-classify positive regions. This helped
 243 mitigate bias toward the majority class and improved sensitivity
 244 to H. pylori-positive samples. Since the models were pretrained on
 245 ImageNet, input patches were normalized using ImageNet statistics,
 246 but no additional augmentations were applied.

247 For the classification head, we performed cross validation grid
 248 search to optimize the MLP architecture and hyperparameters. We
 249 explored hidden layer configurations with one to three layers, units
 250 per layer in 512, 1024 and 2048, dropout with probabilities of 0.1, 0.2,
 251 and 0.3, focal loss gamma values ranging from 2.5 to 5.5. Among the
 252 evaluated architectures, the best-performing model was a single-
 253 hidden-layer MLP with 2048 units, a dropout rate of 0.1, and focal
 254 loss parameters $\alpha = 0.75$, $\gamma = 2.5$. We trained the model using
 255 the Adam optimizer, with a start learning rate of 5e-5. Training
 256 was conducted for 20 epochs and a batch size of 4096 on the pre-
 257 extracted features. We used cosine annealing with 10 warm-up
 258 steps and a min learning rate of 1e-5.

260 261 5.2 Object Detection

262 The overall workflow of our object detection system follows a two-
 263 step process for detecting *Helicobacter pylori* (*H. pylori*) in histology
 264 slides. Initially, we employed a Sliding Window Patch Inference
 265 approach to divide the whole tissue region into smaller patches.
 266 Subsequently, a **segmentation model** was used to further refine
 267 the extraction of regions of interest (ROIs), followed by the appli-
 268 cation of a YOLO11x [9] object detection model for final inference.
 269 The fine-tuning process for this model involved adjusting hyperpa-
 270 rameters, loss functions, learning rates, confidence threshold and
 271 various others to better suit the detection of microscopic features
 272 in our images. By doing so, we were able to enhance the model’s accu-
 273 racy and robustness for our specific use case. For processing each
 274 slide, we first partitioned the tissue into **640 × 640 pixel** patches.
 275 However, to further enhance the identification of relevant regions,
 276 we introduced an additional segmentation model for inference. This
 277 segmentation model operated on 1024×1024 pixel patches or even
 278 larger, allowing for more efficient processing while maintaining
 279 accuracy.

280 The segmentation model used in our system was NuClick [1], an
 281 interactive annotation model from the MONAI [2] project. While
 282 NuClick [1] is originally designed for annotation tasks in pathology,
 283 we adapted it for segmentation by training it on a custom segmenta-
 284 tion dataset. The dataset consisted of approximately **75–100 manually**
 285 **labeled regions**, where segmentation masks were generated
 286 around **gastric glands**, a region where *H. pylori* is prominently
 287 located. To improve model generalization and robustness, we addi-
 288 tionally incorporated data from the publicly available GLAS dataset
 289 [11], which provides glandular segmentation masks.

290 Once trained, the NuClick model was used to infer regions of
 291 interest from the sliding window output. This step refined the ex-
 292 tracted **640 × 640 pixel** patches, ensuring that only relevant tissue
 293 regions were retained for final processing. The filtered patches were
 294 then passed into the YOLO11x [9] model for detection of *H. pylori*.

295 By integrating an additional segmentation model before object
 296 detection, our methodology showcases enhanced extraction of ROIs,
 297 leading to improved localization and computational efficiency.

300 6 EVALUATION

301 Our dataset consists of 22 WSIs of gastric biopsy samples, annotated
 302 for the presence of *H. pylori*. To ensure a balanced distribution of
 303 positive regions across the training and testing sets, we performed
 304 a stratified split of slides based on the quartiles of the number of
 305 ground truth positive regions per slide. This ensures that both the
 306 training (15 slides) and testing (7 slides) subsets contain a similar
 307 distribution of slides with low (Q1), medium (Q2), high (Q3), and
 308 very high (Q4) positive region counts.

309 310 6.1 Classification

311 We evaluated our classification approach at 2 levels:

312 **6.1.1 Patch Extractor.** The candidate region extraction process,
 313 based on template matching, was designed to reduce the number
 314 of regions analyzed by the classifier while ensuring that most pos-
 315 itive regions remained detectable. Of the 437 ground truth positive
 316 regions, 352 of them were covered by at least one candidate region.
 317 That means 19.46% of positive regions were missed at the candidate
 318 extraction stage, making them impossible for the classifier to iden-
 319 tify. At the same time, out of 678,302 possible tiles the WSI could
 320 be split into, 1,4346 candidates were extracted. In other words the
 321 region extraction reduced number of regions passed to the classifier
 322 by 97.88%, thus greatly reducing computational costs. The coverage
 323 rate could be increased by better extraction parameters for example
 324 reducing the template matching threshold, at the cost of more irre-
 325 levant negatives that would slow down the classifier and increase
 326 the data imbalance further. This trade-off between efficiency and
 327 recall is a limitation of the extraction approach. However a better
 328 region extraction method, could be able to balance both.

329 **6.1.2 Classifier.** The CNN classifier was assessed on the test set,
 330 which consisted of candidate regions extracted from testing slides
 331 by the region extractor. The best model used Resnet101 for feature
 332 extraction and achieved a recall of 56.45% and precision of 45.63%.
 333 PR AUC of 0.468 highlights the model had some trouble distinguish-
 334 ing positive regions from negative ones. A Matthews correlation
 335 coefficient (MCC) of 0.496 suggests only moderate agreement be-
 336 tween predictions and ground truth. The precision-recall curve at
 337 figure 3 shows the classifier’s performance across different thresh-
 338 olds. The classifier’s results can be explained by the quality and
 339 quantity of training data. Even with the active learning approach,
 340 which incrementally increased the number of correctly identified
 341 positive regions, labeling errors persisted. With each iteration of
 342 correcting mislabeled samples, we observed a gradual improve-
 343 ment in performance, though the dataset’s quality still constrained
 344 the model’s performance. In the final batch of expert labeling, ap-
 345 proximately 10% of candidate regions were incorrectly marked as

negative when they were actually positive. Better labels would lead to better training and more reliable metrics. However, additional steps such as data augmentation could have mitigated some of these issues by artificially increasing the diversity of training examples, potentially leading to better results.

6.2 Object Detection

Our evaluation focused on two key aspects:

- **Region extraction efficacy.**
- **Overall detection performance.**

For detection performance evaluation, we employed Distance IoU (DIoU) as the primary metric. DIoU [12] was preferred over traditional IoU as it accounts for spatial relationships by incorporating both overlap and center-point distance. Given the nature of *H. pylori* detection—where precise boundary delineation is less critical than correct localization—we adopted a lenient DIoU [12] threshold of 0.15.

The DIoU [12] metric is computed as follows:

$$\text{DIoU} = \frac{\text{IoU} - \rho^2(b, b_{gt})}{\max(\text{IoU}, \rho^2(b, b_{gt}))}$$

Where:

- IoU is the Intersection over Union (IoU) between the predicted bounding box b and the ground truth bounding box b_{gt} ,
- $\rho(b, b_{gt})$ is the Euclidean distance between the centers of the predicted and ground truth boxes.

At this threshold, we computed the following evaluation metrics:

- **Precision:** 58% – proportion of correctly predicted *H. pylori* regions.
- **Recall:** 52% – proportion of actual *H. pylori* regions successfully detected.
- **F1 Score:** 54% – harmonic mean of precision and recall.

To further clarify the detection system's performance, the following image shows a comparison between the ground truth and predicted regions, with all relevant details annotated for better understanding.

In terms of model performance, the system processed approximately 242 patches. The average inference time for each patch was 7.25 seconds, leading to a total inference time of approximately 33 seconds for the entire set of test patches.

7 LIMITATIONS

Despite the promising results achieved, several limitations remain that affect both the dataset and the evaluation metrics(for object detection).

First, the data labeling for the object detection task is not entirely precise. The expert annotations provided for the *H. pylori* regions were limited and did not encompass all possible bacteria locations within the large whole slide images (WSIs). This incomplete labeling creates challenges in training robust detection models, as some true positive regions may be underrepresented or missed entirely. Furthermore, since the expert annotations were primarily drawn at a region level rather than pixel level, the detection system may

struggle to accurately localize smaller instances of *H. pylori*, which is a major drawback in fine-grained detection tasks.

Another limitation stems from the relatively low number of instances of *H. pylori* within the dataset. The small number of positive patches compounded with the large size of the WSIs, created difficulties in preprocessing. Extracting meaningful patches from the WSIs at an appropriate scale while maintaining computational efficiency was challenging. Despite the active learning approach used to improve dataset quality, the class imbalance remained an issue, which could have impacted the model's sensitivity to rare positive instances.

In terms of evaluation, while the Distance Intersection over Union (DIoU) metric used for Object Detection method was employed to assess detection performance, it does not fully account for the inherent challenges posed by the imprecise ground truth annotations. The DIoU metric relies on bounding box overlap and the spatial relationship between predicted and ground truth boxes. However, because the ground truth regions were not annotated in precision, the DIoU metric may not fully capture the discrepancies between the predicted and true locations of *H. pylori*. We need a metric which does better than this.

8 CONCLUSION

In conclusion, this project explored the potential of AI-driven computer-aided detection systems for *H. pylori* in gastric biopsy slides, addressing the limitations of traditional histopathological methods. Our study implemented and evaluated two distinct approaches: classification and object detection. The classification approach, leveraging template matching for region extraction and a CNN classifier, achieved a recall of 56% and precision of 45% on candidate regions. The object detection approach, incorporating sliding window patch inference, segmentation with the NuClick model, and detection with YOLOv1x, demonstrated a recall of 52% and precision of 58 % based on DIoU. While both methods showcased the feasibility of AI in *H. pylori* detection, challenges remain in dataset quality, labeling precision, and computational efficiency. Future research should focus on refining region extraction techniques, expanding and improving datasets with more comprehensive and precise annotations, exploring advanced deep learning architectures, and optimizing computational workflows for real-world deployment. Despite these challenges, this study highlights the significant potential of AI to assist pathologists, improve diagnostic accuracy, and ultimately enhance the efficiency of histopathology examination in the detection of *H. pylori*.

REFERENCES

- [1] Navid Alemi Koohbanani, Mostafa Jahanifar, Neda Zamani Tajadin, and Nasir Rajpoot. 2020. NuClick: A deep learning framework for interactive segmentation of microscopic images. *Medical Image Analysis* 65 (2020), 101771. <https://doi.org/10.1016/j.media.2020.101771>
- [2] M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murray, Andriy Myronenko, Can Zhao, Dong Yang, Vishwesh Nath, Yufan He, Ziyue Xu, Ali Hatamizadeh, Andriy Myronenko, Wentao Zhu, Yun Liu, Mingxin Zheng, Yucheng Tang, Isaac Yang, Michael Zephyr, Behrooz Hashemian, Sachidanand Alle, Mohammad Zalbagi Darestani, Charlie Budd, Marc Modat, Tom Vercauteren, Guotai Wang, Yiwen Li, Yipeng Hu, Yunguan Fu, Benjamin Gorman, Hans Johnson, Brad Genereaux, Barbaros S. Erdal, Vikash Gupta, Andres Diaz-Pinto, Andre Dourson, Lena Maier-Hein, Paul F. Jaeger, Michael Baumgartner, Jayashree Kalpathy-Cramer, Mona Flores, Justin Kirby, Lee A. D. Cooper, Holger R. Roth, Daguang Xu, David Bericat, Ralf Floca, S. Kevin

- 465 Zhou, Haris Shuaib, Keyvan Farahani, Klaus H. Maier-Hein, Stephen Aylward,
 466 Prerna Dogra, Sebastien Ourselin, and Andrew Feng. 2022. MONAI: An open-
 467 source framework for deep learning in healthcare. arXiv:2211.02701 [cs.LG]
<https://arxiv.org/abs/2211.02701>
- 468 [3] Hyun Ho Choi. 2024. Strategic Role of Immunohistochemical Staining in De-
 469 tector of Helicobacter pylori. *The Korean Journal of Helicobacter and Upper*
Gastrointestinal Research 24, 1 (mar 2024), 3–4. [Online; accessed 2025-01-22].
- 470 [4] Ayman Elbehiry, Eman Marzouk, Musaad Aldubaib, Adil Abalkhail, Sulaiman
 471 Anagreyyah, Nuhu Anajirih, Abdulaziz Almuzaini, Mohammed Rawway, Ab-
 472 dulmajeed Alfadhel, Abdelmaged Draz, and Akram Abu-Okail. 2023. Helicobacter
 473 pylori Infection: Current Status and Future Prospects on Diagnostic, Therapeutic
 474 and Control. *Antibiotics* 12 (01 2023), 25. <https://doi.org/10.3390/antibiotics12020191>
- 475 [5] Ayman Elbehiry, Eman Marzouk, Musaad Aldubaib, Adil Abalkhail, Sulaiman
 476 Anagreyyah, Nuhu Anajirih, Abdulaziz Almuzaini, Mohammed Rawway, Ab-
 477 dulmajeed Alfadhel, Abdelmaged Draz, and Akram Abu-Okail. 2023. Helicobacter
 478 pylori Infection: Current Status and Future Prospects on Diagnostic, Therapeutic
 479 and Control Challenges. *Antibiotics* 12, 2 (jan 17 2023), 191. [Online;
 480 accessed 2025-01-22].
- 481 [6] Wanderson Gonçalves E. Gonçalves, Marcelo Henrique Paula Dos Santos,
 482 Leonardo Miranda Brito, Helber Gonzales Almeida Palheta, Fábio Manoel França
 483 Lobato, Samia Demachki, Andrea Ribeiro-Dos-Santos, and Gilderlanio Santana de
 484 Araújo. 2022. DeepHP: A New Gastric Mucosa Histopathology Dataset for Helicobacter
 485 pylori Infection Diagnosis. *International journal of molecular sciences* 23, 23 (Nov. 2022). <https://doi.org/10.3390/ijms232314581> Place: Switzerland.
- 486 [7] Abdullah Umar Ibrahim, Fikret Dirilenoglu, Uğuray Payam Hacisalihoglu, Ah-
 487 met İlhan, and Omid Mirzaei. 2024. Classification of H. pylori Infection from
 488 Histopathological Images Using Deep Learning. *Journal of imaging informatics in medicine* 37, 3 (June 2024), 1177–1186. <https://doi.org/10.1007/s10278-024-01021-0> Place: Switzerland.
- 489 [8] International Agency for Research on Cancer. n.d.. Agents Classified by the
 490 IARC Monographs, Volumes 1–132. <https://monographs.iarc.who.int/agents-classified-by-the-iarc> Retrieved January 22, 2025.
- 491 [9] Glenn Jocher and Jing Qiu. 2024. Ultralytics YOLO11. <https://github.com/ultralytics/ultralytics>
- 492 [10] Sebastian Klein, Jacob Gildenblat, Michael Angelika Ihle, Sabine Merkelbach-
 493 Bruse, Ka-Won Noh, Martin Peifer, Alexander Quaas, and Reinhard Büttner. 2020.
 494 Deep learning for sensitive detection of Helicobacter Pylori in gastric biopsies.
 495 *BMC Gastroenterology* 20, 1 (Dec. 2020), 417. <https://doi.org/10.1186/s12876-020-01494-7>
- 496 [11] Korsuk Sirinukunwattana, Josien P. W. Pluim, Hao Chen, Xiaojuan Qi, Pheng-
 497 Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J. Matuszewski, Elia Bruni,
 498 Urko Sanchez, Anton Böhm, Olaf Ronneberger, Bassem Ben Cheikh, Daniel
 499 Racoceanu, Philipp Kainz, Michael Pfeiffer, Martin Urschler, David R. J. Snead,
 500 and Nasir M. Rajpoot. 2017. Gland segmentation in colon histology images:
 501 The glas challenge contest. *Medical Image Analysis* 35 (2017), 489–502. <https://doi.org/10.1016/j.media.2016.08.008>
- 502 [12] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren.
 503 2019. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression.
 504 *CoRR* abs/1911.08287 (2019). arXiv:1911.08287 <http://arxiv.org/abs/1911.08287>

A.2 Figures

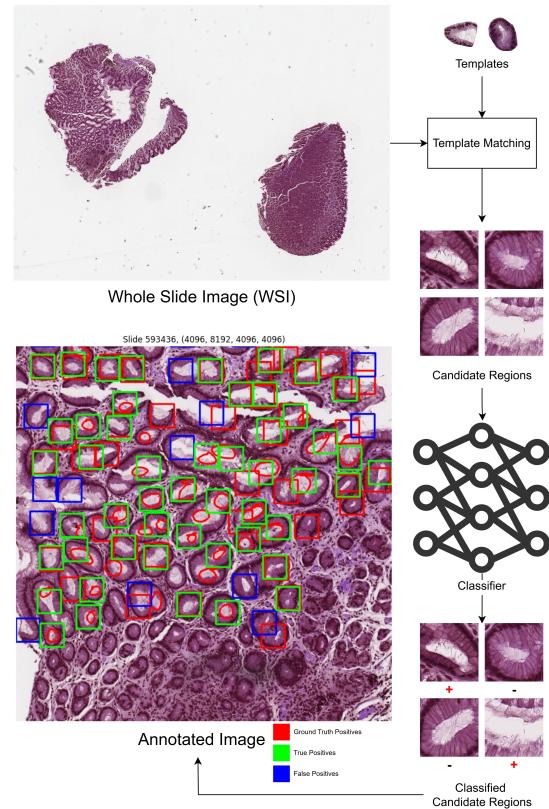


Figure 1: Classification approach

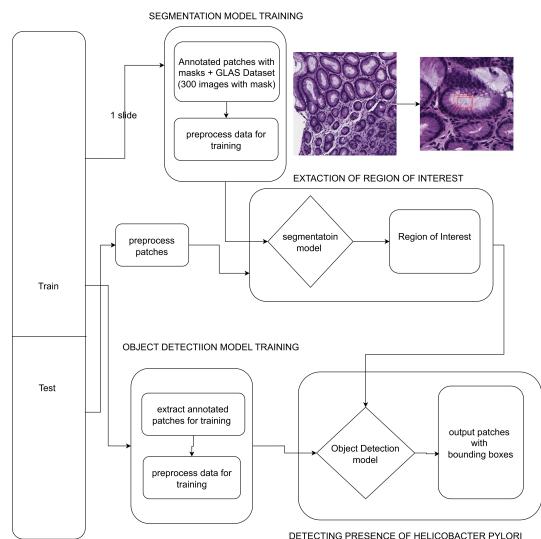


Figure 2: Object Detection approach

A APPENDICES

A.1 Acknowledgments

We sincerely thank LABOKLIN for their invaluable support in providing the dataset and ground truth labeling for this study. Their contributions were instrumental in our research, and this work would not have been possible without their expertise and support.

523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580

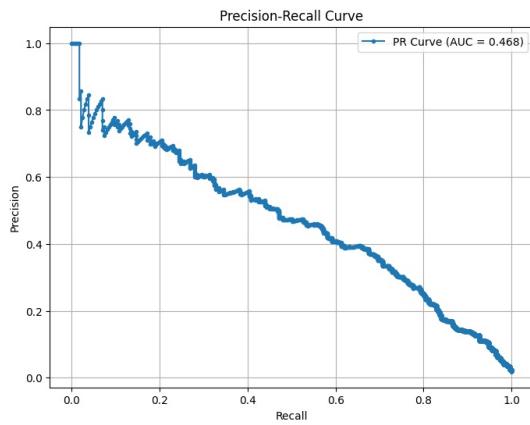


Figure 3: Precision-recall curve for best classifier

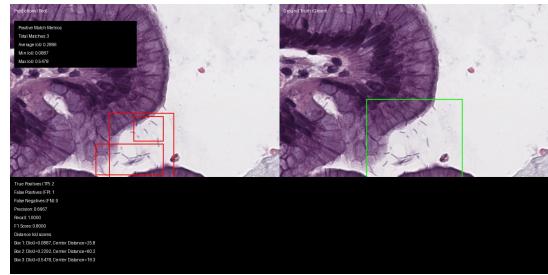


Figure 4: Comparison of ground truth and predicted *H. pylori* regions using Object Detection