Gastric cancer detection based on Colorectal Cancer transfer learning

Sara Nóbrega[1]
sarapiresnobrega@hotmail.com
Alexandre Neto[1,2]
Alexandre.h.neto@inesctec.pt
Miguel Coimbra[2,3]
mcoimbra@fc.up.pt
António Cunha[1,2]
acunha@utad.pt

[1]Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal
[2]Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Porto, Portugal
[3]Faculdade de Ciências, Universidade do Porto, Portugal

Abstract

Gastric Cancer (GC) and Colorectal Cancer (CRC) are some of the most common cancers in the world. The most common diagnostic methods are upper endoscopy and biopsy. Possible expert distractions can lead to late diagnosis. GC is a less studied malignancy than CRC, leading to scarce public data that difficult the use of AI detection methods, unlike CRC where public data are available. Considering that CRC endoscopic images present some similarities with GC, a CRC Transfer Learning approach could be used to improve AI GC detectors. This paper evaluates a novel Transfer Learning approach for real-time GC detection, using a YOLOv4 model pre-trained on CRC detection. The results achieved are promising since GC detection improved relatively to the traditional Transfer Learning strategy.

1 Introduction

Gastric Cancer (GC) stands as a widespread and lethal malignancy, ranking fifth globally in 2020, while Colorectal Cancer (CRC) claimed the third position [1]. In Portugal, GC constituted 4.90% and CRC 17.40% of new cancer cases. Diagnosis typically relies on endoscopy and biopsy, mirroring the approach for CRC detection via colonoscopy. Timely GC detection is critical due to its typically late-stage symptom onset, resulting in poor long-term prognosis [1]. The elusive early GC manifestations demand meticulous examination, but the complex nature of lesions, coupled with possible expert oversights, can lead to missed diagnoses.

Deep Learning (DL) systems, capable of real-time detection of gastrointestinal endoscopic lesions, offer valuable support to endoscopists, flagging suspicious regions during examinations [2]. These systems also aid in the educational process for young endoscopists by facilitating the understanding of lesion characteristics [1]. While public databases primarily feature colonoscopy-derived polyp images, few include endoscopy images of GC. Visual parallels between GC and CRC endoscopic images, though not clinically significant, serve as valuable training data for DL models (Figure 1).

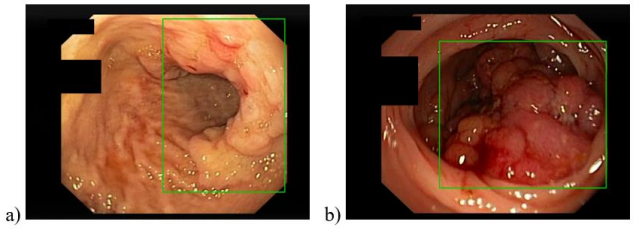

Figure 1: Gastric Cancer a), Colorectal Cancer b). Benchmark dataset [3].

Recognizing the shared features between GC and CRC, transfer learning (TL) approaches could enhance GC pathology detection. However, no existing studies have explored the potential contribution of these shared features in a TL strategy. This paper presents a preliminary study evaluating the effectiveness of a TL approach utilizing a YOLOv4 model for real-time GC detection, leveraging features learned from CRC data.

2 LITERATURE REVIEW

Several studies have tackled GC classification using diverse architectures. Y. Horiuchi et al. [4] employed a GoogLeNet model to differentiate between non-cancerous and cancerous images, achieving 87% accuracy. Sakai et al. [5] developed a GoogLeNet-based model to classify normal and early GC images, achieving an accuracy of 87.60%, sensitivity of 80%, and specificity of 94.80%.

In GC detection, Ikenoyama et al. [6] employed the SSD architecture to process over 13,000 resized endoscopic images, achieving 58.40% sensitivity, 87.30% specificity, 26% Positive Predicted Value (PPV), and 96.50% Negative Predicted Value (NPV). Hirasawa et al. [7] focused on early and advanced GC detection, using a dataset of more than 13,000 images across multiple modalities. Their approach yielded a sensitivity of 92.20% and PPV of 30.60%.

These studies collectively highlight efforts to advance GC pathology classification and detection using a variety of architectural models and datasets. However, the application of transfer learning from CRC data to enhance GC detection remains an underexplored area.

The potential for transfer learning to bolster GC detection by leveraging features learned from CRC data holds promise. As highlighted in the existing literature, classification and detection models have demonstrated noteworthy accuracies and sensitivities. Investigating the transferability of features between these closely related cancers presents an exciting avenue for advancing real-time GC detection capabilities.

3 Methods and materials

This study introduces a TL approach from CRC to GC detection. The YOLOv4 real-time object detection framework is utilized, structured across three key stages. Firstly, the CSPDarknet-53 classifier is trained and assessed for CRC detection using pre-processed colon images (DS1). The trained CSPDarknet-53 serves as the backbone for the YOLOv4 CRC detector. Subsequently, the YOLOv4 CRC detector is trained and evaluated on CRC and healthy colon images from dataset DS2. Lastly, the YOLOv4 CRC detector's backbone and neck are fine-tuned for GC detection using a five-fold cross-validation strategy on pre-processed images from dataset DS3.

The study employs images from the publicly available Benchmark dataset for digestive tract diagnostics support systems [3]. This dataset contains gastrointestinal endoscopy images, including CRC and GC, collected from various patient examinations. Additionally, four videos from the HyperKvasir dataset [8] were used for real-time GC detection assessment. Three videos represented CRC, while one depicted GC.

	Dataset	Split	Cancer	Non-cancer
CRC *classification*	DS1	Train	7,546 (20*)	7,244 (49*)
		Validation	2,895 (7*)	2,895 (4*)
		Test	1,804 (35*)	1,892 (8*)
CRC *detection*	DS2	Train	1,649 (54*)	1,639 (9*)
		Validation	499 (3*)	500 (1*)
		Test	215 (15*)	118 (2*)
	Real-time evaluation: 3 videos (33s, 120s, and 34s).			
GC *detection*	DS3	Fold 1	56 (6*)	56 (17*)
		Fold 2	56 (6*)	56 (20*)
		Fold 3	56 (7*)	56 (28*)
		Fold 4	56 (7*)	56 (18*)
		Fold 5	56 (7*)	52 (23*)
		Test set	81 (1*)	81 (31)
	Real-time evaluation: 1 video (85s).			

Table 1: Dataset splits.

Images were resized to 416x416 pixels to conserve computational resources. Overlapping alphanumeric characters in colon and gastric images were obscured to prevent model prediction bias. Images were divided into three datasets (DS1, DS2, DS3) for training and evaluation. A five-fold cross-validation strategy was employed for GC detection due to limited GC images (Table 1). The folds were balanced with other

pathologies. Two YOLOv4 GC detectors were trained and compared: one using proposed TL and the other using traditional TL from the MS COCO dataset. The proposed TL involved training YOLOv4 first for CRC detection, followed by fine-tuning for GC using DS3. The GC detector used the backbone and neck of the CRC detector as initial weights.

Evaluation metrics followed standard conventions. For object detection, accurate detection required an Intersection Over the Union (IoU) of ≥30% between predicted and ground-truth bounding boxes.

4 Experiments and Results

CSPDarknet-53 exhibited promising results across both analyses, displaying high metrics. In the per-image examination, it accurately identified 1,588 out of 1,804 cancer images and 1,858 normal colonic images from the DS1 test set. The classifier discerned CRC presence in examinations and correctly identified 89% of healthy colonic mucosa images in a representative healthy colon examination (233 images). With an AUROC of 0.98, it effectively distinguishes classes. Unlike Yang et al. [9] and D. Zhou et al. [10], who focused on neoplastic/non-neoplastic and CRC/non-CRC binary classification, our model excelled in specificity and PPV.

In per-image analysis, YOLOv4 CRC detector achieved 84/96 correct predictions with IoU >30%. Per examination, it detected CRC in 14/15 test cases, directing attention to lesion areas. Despite lower per-image sensitivity (40.93%), it exhibited higher per-examination sensitivity (93.33%). On HyperKvasir videos, YOLOv4 CRC detector identified potential lesion areas at >40 FPS, surpassing real-time requirements. In comparison to [7] and [11], per-image sensitivity was lower, while per-examination results resembled Poon et al.'s [11] polyp-based analysis.

In Figure 2, it is possible to observe the images where all the cancer detectors achieved the highest IoU during the test. Green bounding boxes represent the ground truth, and red bounding boxes represent the model's prediction.

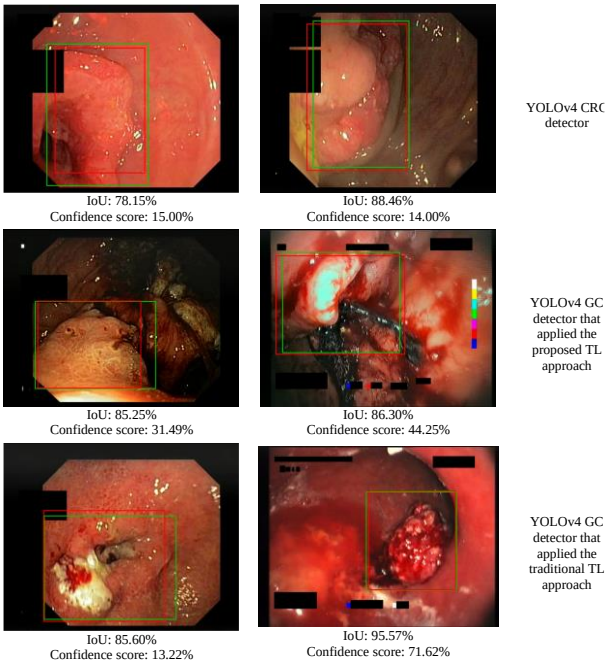

Figure 2: Images where all the cancer detectors achieved the highest IoU.

Both traditional TL and proposed TL YOLOv4 GC detectors showed similar per-image/examination results. Encouragingly, a high number of GC images were detected, achieving PPV >55% in both analyses. Both detectors exhibited improved true negative classification of other pathologies (Specificity >70%). The proposed TL approach outperformed traditional TL, detecting GC in over 18.57% of images during cross-validation. In an independent test set, similar per-image results were observed, with the novel TL approach outperforming MS COCO-based weights. Notably, our study's datasets are more extensive and diverse compared to prior literature, offering improved sensitivity, PPV, and NPV than certain established models [6][7]. False positives primarily correlated with negative class pathologies.

5 Conclusions

This study assesses the impact of a new transfer learning method on YOLOv4 for real-time Gastric Cancer detection. Training and evaluation were performed on the Benchmark dataset and HyperKvasir public dataset. Encouragingly, the transfer learning model detected more Gastric Cancer images compared to the non-transfer learning model. Future endeavors include validating this approach on a broader and more diverse dataset.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project PTDC/EEI-EEE/5557/2020. Co-funded by the European Union (grant number 101095359) and supported by the UK Research and Innovation (grant number 10058099). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

References

[1] J. Arribas et al., «Standalone performance of artificial intelligence for upper GI neoplasia: a meta-analysis», Gut, vol. 70, n. 8, pp. 1458–1468, ago. 2021, doi: 10.1136/gutjnl-2020-321922.

[2] F. Renna et al., «Artificial Intelligence for Upper Gastrointestinal Endoscopy: A Roadmap from Technology Development to Clinical Practice», Diagnostics, vol. 12, n. 5, p. 1278, mai. 2022, doi: 10.3390/diagnostics12051278.

[3] Benchmark dataset for a digestive tract diagnostics support systems. Retrieved May 03, 2021, from https://cvlab.eti.pg.gda.pl/publications/ endoscopy-dataset

[4] Y. Horiuchi, T. Hirasawa, N. Ishizuka, Y. Tokai, K. Namikawa, S. Yoshimizu, A. Ishiyama, T. Yoshio, T. Tsuchida, J. Fujisaki, and T. Tada, "Performance of a computer-aided diagnosis system in diagnosing early gastric cancer using magnifying endoscopy vídeos with narrow-band imaging (with videos)," Gastrointestinal Endoscopy, vol. 92, no. 4, pp. 856–865.e1, oct 2020.

[5] Y. Sakai et al. , S. Takemoto, K. Hori, M. Nishimura, H. Ikematsu, T. Yano, and H. Yokota, "Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, jul 2018.

[6] Y. Ikenoyama, T. Hirasawa, M. Ishioka, K. Namikawa, S. Yoshimizu, Y. Horiuchi, A. Ishiyama, T. Yoshio, T. Tsuchida, Y. Takeuchi, S. Shichijo, N. Katayama, J. Fujisaki, and T. Tada, "Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists," Digestive Endoscopy, vol. 33, no. 1, pp. 141–150, jun 2020.

[7] T. Hirasawa et al., 'Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images', Gastric Cancer, vol. 21, no. 4, pp. 653–660, Jul. 2018, doi: 10.1007/s10120-018-0793-2.

[8] Borgli, Hanna, Thambawita, Vajira, Smedsrud, Pia H., Hicks, Steven, Jha, Debesh, Eskeland, Sigrun L., de Lange, Thomas. (2020). HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Scientific Data, 7(1), 283.

[9] Yang, Young Joo, Cho, Bum-Joo, Lee, Myung-Je, Kim, Ju Han, Lim, Hyun, Bang, Chang Seok, Baik, Gwang Ho. (2020). Automated classification of colorectal neoplasms in white-light colonoscopy images via deep learning. Journal of clinical medicine, 9(5), 1593.

[10] Zhou, Dejun, Tian, Fei, Tian, Xiangdong, Sun, Lin, Huang, Xianghui, Zhao, Feng, Yang, Meng. (2020). Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. Nature communications, 11(1), 1-9.

[11] Poon, Carmen CY, Jiang, Yuqi, Zhang, Ruikai, Lo, Winnie WY, Cheung, Maggie SH, Yu, Ruoxi, Wong, Sunny H. (2020). AI-doscopist: a real-time deep-learning-based algorithm for localising polyps in colonoscopy videos with edge computing devices. NPJ Digital Medicine, 3(1), 1-8.