Kvasir-Capsule, a video capsule endoscopy dataset

Pia H. Smedsrud[1,3,6,+,*], Henrik Gjestang[1,3], Oda Olsen Nedrejord[1,3], Espen Næss[1,3], Vajira Thambawita[1,2,+], Steven A. Hicks[1,2,+], Hanna Borgli[1,3], Debesh Jha[1,7,+], Tor Jan Derek Berstad[6], Sigrun L. Eskeland[4], Mathias Lux[11], Håvard Espeland[6], Andreas Petlund[6], Duc Tien Dang Nguyen[5], Enrique Garcia-Ceja[13], Dag Johansen[7], Peter T. Schmidt[9], Hugo L. Hammer[1,2], Thomas de Lange[4,6,12,@,+], Michael A. Riegler[1,@,+], and Pål Halvorsen[1,2,@,+]

[1]SimulaMet, Norway
[2]Oslo Metropolitan University, Norway
[3]University of Oslo, Norway
[4]Department of Medical Research, Bærum Hospital, Norway
[5]University of Bergen, Norway
[6]Augere Medical AS, Norway
[7]UIT The Arctic University of Norway, Norway
[8]Simula Research Laboratory, Norway
[9]Karolinska University Hospital, Sweden
[10]Cancer Registry of Norway, Norway
[11]Klagenfurt University, Austria
[12]Medical Department, Sahlgrenska University Hospital-Mölndal, Sweden
[13]SINTEF Digital, Norway
[*]Corresponding author: pia@simula.no
[+]these authors contributed equally to this work
[@]joint senior authors

ABSTRACT

Artificial intelligence (AI) is predicted to have profound effects on the future of video capsule endoscopy (VCE) technology. The potential lies in improving anomaly detection while reducing manual labour. However, medical data is often sparse and unavailable to the research community, and qualified medical personnel rarely have time for the tedious labelling work. In this respect, we present *Kvasir-Capsule*, a large VCE dataset collected from examinations at Hospitals in Norway. *Kvasir-Capsule* consists of 118 videos which can be used to extract a total of $4,820,739$ image frames. We have labelled and medically verified $44,228$ frames with a bounding box around detected anomalies from 13 different classes of findings. In addition to these labelled images, there are $4,776,479$ unlabelled frames included in the dataset. Initial work demonstrates the potential benefits of AI-based computer-assisted diagnosis systems for VCE. However, they also show that there is great potential for improvements, and the *Kvasir-Capsule* dataset can play a valuable role in developing better algorithms in order for VCE technology to reach its true potential.

Background & Summary

The small bowel constitutes the gastrointestinal (GI) tract's mid-part, situated between the stomach and the large bowel. It is three to four meters long and has a surface of about $30m^2$, including the surface of the villi, and plays a crucial role in absorbing nutrients[1]. Therefore, disorders in the small bowel may cause severe growth retardation in children and nutrient deficiencies in children and adults[1]. This organ may be affected by chronic diseases, like Crohn's disease, coeliac disease, and angiectasias, or malignant diseases like lymphoma and adenocarcinoma[2,3]. These diseases may represent a substantial health challenge for both patients and society, and a thorough examination of the lumen is frequently necessary to diagnose and treat them[4]. However, the small bowel, due to its anatomical location, is less accessible for inspection by flexible endoscopes commonly used for the upper GI tract and the large bowel. Since early 2000, video capsule endoscopy (VCE)[5] has been used, usually as a complementary test for patients with GI bleeding[4]. A VCE consists of a small capsule containing a wide-angle camera, light sources, batteries, and other electronics. The patient swallows the capsule, which then captures a video as it moves passively

Table 1. An overview of existing VCE datasets from the GI tract

Dataset	Findings	Size	Availability
KID[39]	Angiectasia, bleeding, inflammations, polyps	2,371 images + 47 videos	open academic•
GIANA 2017[40]	Angiectasia†	600 images	by request
GIANA2018[41,42]	Polyps and small bowel lesions†	8,262 images + 38 videos	by request
CAD-CAP[43,44]	Normal frames, fresh blood, vascular lesion, ulcerative and inflammatory lesions	25,000 images	by request◇

†Including ground truth segmentation masks •Not available anymore

◇The Computer-Assisted Diagnosis for CAPsule endoscopy (CAD-CAP) Database - used for the angiectasia detection

throughout the GI tract. A recorder, carried by the patient or included in the capsule, stores the video before a medical expert assesses it after the procedure.

VCE devices exist in multiple versions from manufacturers such as Given Imaging (Medtronics), Ankon Technologies, Chongqing Science, IntroMedic, CapsoVision, and Olympus. The frame rate typically varies between 1 and 30 frames per second, capturing in total between 50 and 100 thousand frames, with pixel-resolutions in the range of 256×256 to 512×512. Some of the vendors have software to remove duplicated frames due to slow movement. However, a large number of frames need to be analysed by a medical expert, resulting in a tedious and error-prone operation. In the related area of colonoscopy, operator variation and detection performance are reported problems[6–8] resulting in high miss rates[9]. In VCE analysis, essential findings are also missed due to lack of concentration, and insufficient experience and knowledge. Furthermore, physicians may have trouble reading the video and handling the associated technology, and infrequent VCE use leads to lack of confidence[10], all resulting in inter- and intra-observer variations in the assessments[11].

To improve the usefulness of VCE technology, both anomaly miss-rates and the amount of human labour must be reduced. The technical developments for automated image and video analysis have sky-rocketed, and multimedia solutions in medicine show great potential[12,13]. There is an increasing number of promising machine learning solutions being developed for automated diagnosis of colonoscopies[14–21] using open datasets[22–25]. Regarding automated analyses of VCE data, machine learning approaches also produce promising results regarding detection and classification rates[26–32]. Machine learning, or artificial intelligence (AI) in general, is therefore likely to have profound effects on future of VCE technology, not only for improving variation and detection rates, but also estimating the localisation of the capsule[10,33].

Regardless of promising initial results, there is room for improvements in detection rate, reduction of manual labour, and AI explainability. However, large amounts of data are needed, and access to these data are often scarce. As shown in Table 1, very few, small VCE datasets are published, and several have become unavailable. Datasets containing images from colonoscopies and esophagogastroscopies are not applicable because they do not depict the small bowel, characterised by the intestinal villi displaying a different surface than the rest of the bowel. In addition, the resolution of the images from VCE is much lower, and the movement of the capsule is uncontrolled in opposite to flexible endoscopes.

Therefore, we present a large VCE dataset, called *Kvasir-Capsule*, consisting of 118 videos with 4,820,739 frames. The data are collected from routine clinical VCE examinations performed at a Norwegian hospital. In total, the labelled images represent 13 classes of findings. In addition to the labelled images and their corresponding full videos, there are unlabelled videos included in the dataset. Recent work in the machine learning community has shown great improvements regarding unlabelled data value. Semi-supervised learning algorithms can now learn from sparsely labelled and unlabelled data, e.g., successfully applied in different medical image analyses[34,35] using self-learning[36,37] and neural graph learning[38]. Finally, we provide a baseline analysis and outline possible future research directions using *Kvasir-Capsule*.

Methods

VCE videos were collected prospectively from consecutive clinical examinations performed at the Department of Medicine at Bærum Hospital, Vestre Viken Hospital Trust in Norway between February 2016 and January 2018 with an Olympus VCE system. Vestre Viken Hospital Trust provides health care services to 490,000 people, of which about 200,000 are covered by Bærum Hospital. Initially, a trained clinician analysed all videos, selecting thumbnails from lesions and normal findings as part of their clinical work. All the 118 capsule endoscopy videos were recorded with the Olympus Endocapsule 10 System* using the Olympus EC-S10 endocapsule (Figure 1a) and the Olympus RE-10 endocapsule recorder (Figure 1b). In spring 2019, anonymous videos and thumbnails were exported from the Olympus stand-alone workstation using the Olympus software. The

*https://www.olympus-europa.com/medical/en/Products-and-Solutions/Products/Product/ENDOCAPSULE-10-System.html

Olympus video capsule system has user-friendly functionalities like Omni-selected Mode, skipping images that overlap with previous ones.

(a) Olympus EC-S10 endocapsule

(b) Olympus RE-10 endocapsule recorder

Figure 1. VCE equipment used for data collection

All metadata were removed and files renamed with randomly generated file names, before exporting all videos and thumbnails. The data has not been pre-processed or augmented in any way apart from this. Subsequently, an expert endoscopist selected the thumbnails with the pathological findings. These thumbnails were traced to their corresponding video segments, and the videos were uploaded to a video annotation platform (provided by Augere Medical AS, Norway) for efficient viewing and labelling. Then, three master students labelled and marked the findings with bounding boxes for each frame. If the students were unsure about the labelling, the endoscopist verified the frames. Finally, the annotations were once more verified by two medical doctors, before the video frames were exported as images. A total of 44,228 frames are labelled.

The Privacy Data Protection Authority approved the export of anonymous images for the creation of the database. It was exempted from approval from the Regional Committee for Medical and Health Research Ethics - South East Norway. Since the data is anonymised and all metadata removed, the dataset is publicly shareable based on Norwegian and General Data Protection Regulation (GDPR) laws.

Data Records

The *Kvasir-Capsule* dataset is available from the Open Science Framework (OSF) accessible via the link `https://osf.io/dv2ag/`. Originally, the videos are captured using the Olympus EC-S10 endocapsule at a variable frame rate of 3-5 frames per second, in a resolution of 336x336, and encoded using H264 (MPEG-4 AVC, part 10). The videos are exported in AVI format using the Olympus system's export tool packaged and encapsulated in the same H264 format, i.e., the frame formats are the same, but the frame rate specification is changed to 30 fps by the export tool. Table 2 gives an overview of all data records in the dataset. In total, the dataset consists of 4,820,857 main data records, i.e., 44,228 images with labels and bounding box masks, the 44 corresponding labelled videos (the videos from which the images are extracted), and 74 unlabelled videos (from which labelled images have not been extracted). 4,776,479 unlabelled images can further be extracted from all the videos. All the various labelled classes are shown in Figure 2. The dataset has a total size of circa 94.7 GB. Note that the unlabelled images are not extracted and included in the uploaded data due to unnecessary duplication of data, but can easily be extracted from the videos.

Table 2. Overview of the data records in the *Kvasir-Capsule* dataset

Data Record	# Files
Labelled images	44,228
Labelled videos	44
Unlabelled images	4,776,479
Unlabelled videos	74

Labelled images

In total, the dataset contains 44,228 labelled images stored using the PNG format, where Figure 3 shows the 13 different classes representing the labelled images and the number of images in each class. The provided *metadata.csv* comma-separated value

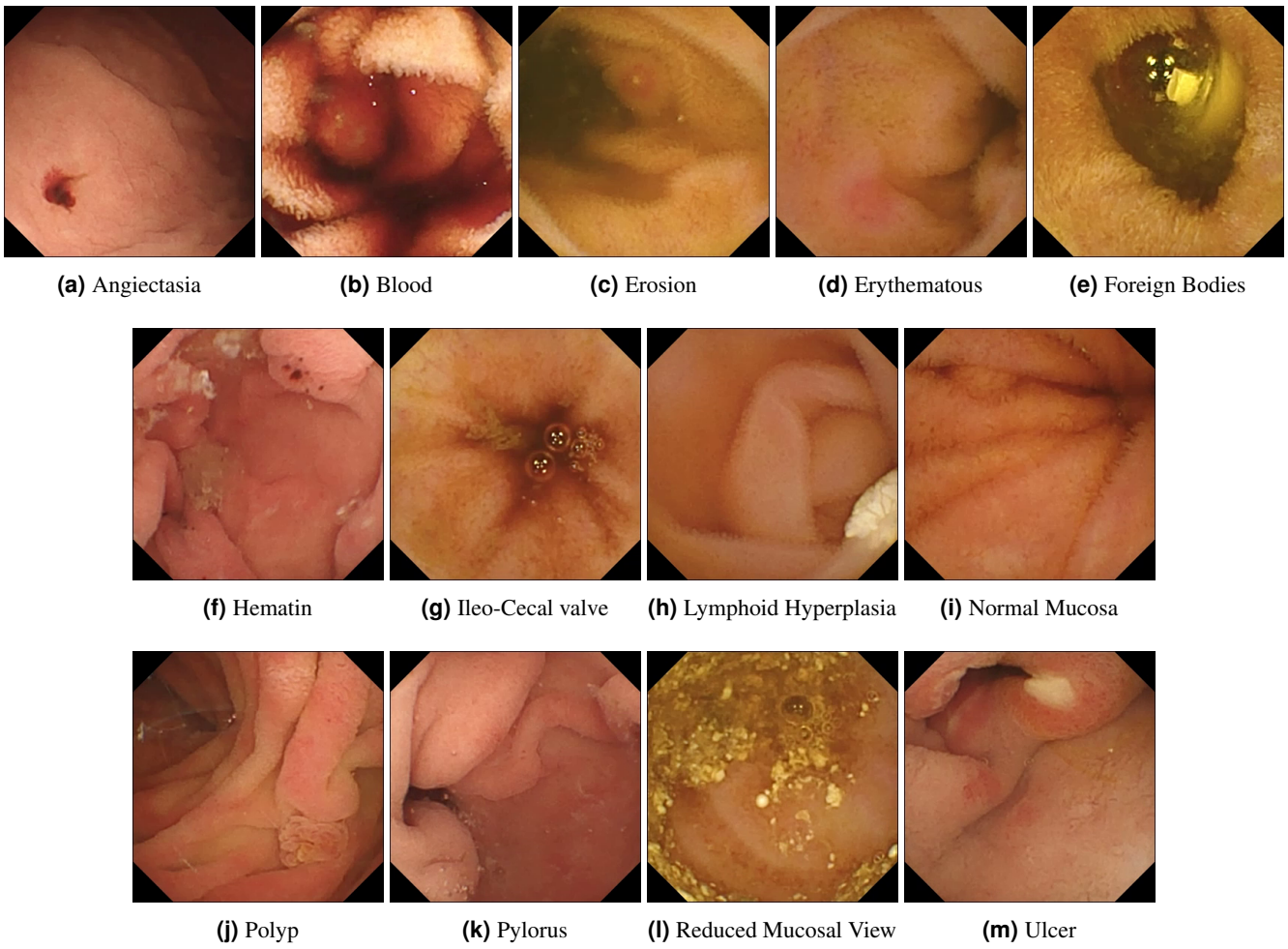

Figure 2. Image examples of the various labelled classes for images

(CSV) file gives the mapping between file name, the labelling for the image, the corresponding video, and the video frame number. Moreover, the CSV file gives information about the bounding box outlining the finding.

We defined three main categories of findings, namely anatomical landmarks, quality of mucosal view, and pathological findings. Each class and the images belonging to it is stored in the corresponding folder of the category it belongs to. As observed in Figure 3, the number of images per class is not balanced. This is a global challenge in the medical field because some findings occur more often than others, and adds a challenge for researchers since methods applied to the data should also be able to learn from a small amount of training data.

Anatomical landmarks

Anatomical landmarks are characteristics of the GI tract used for orientation during endoscopic procedures. Furthermore, they are used to confirm the complete extent of the examination. We have labelled two classes of *anatomical landmarks* which delineates the upper (proximal) and lower (distal) end of the small bowel. The **pylorus** is the anatomical junction between the stomach and small bowel and is a sphincter (circular muscle) regulating the emptying of the stomach into the duodenum. The **ileocecal valve** marks the transition from the small bowel to the large bowel and is a valve preventing reflux of colonic contents, stool, back into the small bowel.

Quality of mucosal view

A complete visualisation of the mucosa is crucial to ensure one discovers all pathological findings. For the *quality of mucosal view* assessment, we have labelled two classes from normal images without any other findings. **Normal mucosa** depicts relatively clean small bowel mucosa with healthy villi and no pathological findings. This class can also double as a "normal" class versus the pathological findings classes (see below). The class **reduced mucosal view** shows small bowel content reducing the view of the mucosa, like stool or bubbles, meaning the mucosa can not be adequately assessed.

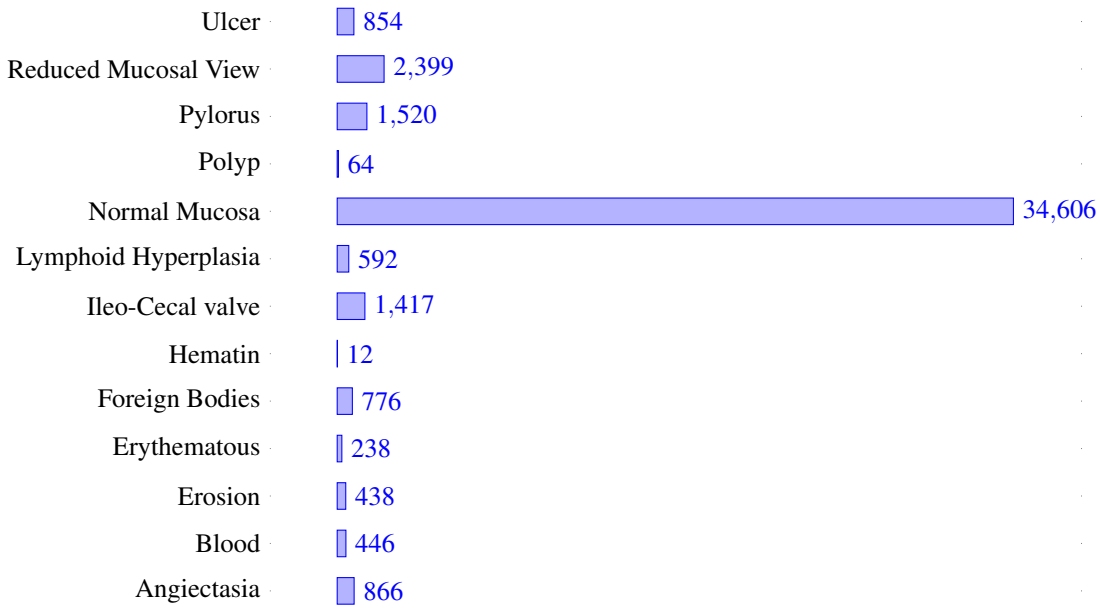

Figure 3. The number of images in the various Kvasir-Capsule labelled image classes

Pathological findings

All parts of the GI tract can be affected by abnormalities or findings due to disease, and the small bowel is no exception. Abnormalities, called *pathological findings*, in the small bowel can be seen both as irregular content in the mucosal lumen or as changes to the mucosal surface. These findings are classified according to the Minimal Standard Terminology, defined by the World Endoscopy Organization[45].

Normally, the small bowel contains only a certain amount of yellow or brown liquid. However, abnormalities in the upper GI tract or small bowel may bleed, causing the appearance of fresh **blood** which colour the liquid red. In cases with minimal bleeding, one may observe small black stripes called **hematin** on the mucosal surface. **Foreign bodies** like tablet residue or retained capsules can also be observed in the lumen. Typical mucosal changes sometimes cover larger segments, such as a reddish appearance called **erythematous mucosa**. The mucosal wall can also have different focal lesions, which can be flat, excavated or protruding compared to the surface of the normal mucosa. The flat lesions represented in the *Kvasir-Capsule* dataset are **angiectasias**; small superficial dilated vessels causing chronic bleeding and subsequently anaemia. It mostly occurs in people with chronic heart and lung diseases[46]. Excavated lesions erode to different extents the surface of the mucosa. Most common are **erosions**, covered by a tiny fibrin layer, while larger erosions are called **ulcers**. As an example, Crohn's disease is a chronic inflammation of the small bowel characterised by ulcers and erosions of the mucosa. It may cause strictures of the lumen, making the absorption and passage of nutrients difficult[47]. The classes of protruding lesions in this dataset are **polyps**, that may be precancerous lesions, and **lymphoid hyperplasia**, which represents normal lymphoid tissue in the mucosal wall.

Labelled videos

Labelled videos are the full 44 videos from which we extracted the above mentioned labelled image classes. In total, these videos correspond to approximately 19 hours of video and 2,034,910 video frames. As previously mentioned, one can find the frame number and video of origin of each extracted image in the CSV-file. Even though we already have extracted the most interesting frames (images) found by the clinicians from these videos, they do contain a large number of non-labelled frames that could be interesting in future research, or if one wants to extract the video sequences around the various findings.

Unlabelled videos

We also provide 74 videos, which contains approximately 25 hours of video and 2,785,829 video frames, without any labels. They are stored in the same format as the ones used for labelling. As previously mentioned, unlabelled data can still have great value. As mentioned in the background, sparsely labelled data can be important for recently emerging semi-supervised learning algorithms. These videos are of the same type and quality as the labelled videos, except we do not provide any annotations.This

Table 3. Results for all classification experiments. Experiments were done with and without weighted cross-entropy loss (CEL) and using a weighted sampling technique. Bold numbers represent the best average value of that column

		Macro average			Micro average			
	Method	Precision	Recall	F1-score	Precision	Recall	F1-score	MCC
Normal CEL	DensNet-161 (fold 0)	0.2266	0.2228	0.1848	0.7444	0.7444	0.7444	0.2756
	DensNet-161 (fold 1)	0.3029	0.2646	0.2367	0.7370	0.7370	0.7370	0.3695
	Average	0.2647	0.2437	0.2108	0.7407	0.7407	0.7407	0.3226
	ResNet-152 (fold 0)	0.2975	0.2375	0.1885	0.7406	0.7406	0.7406	0.2504
	ResNet-152 (fold 1)	0.2620	0.2353	0.2245	0.7537	0.7537	0.7537	0.3663
	Average	0.2797	0.2364	0.2065	**0.7472**	**0.7472**	**0.7472**	0.3083
Weighted CEL	DensNet-161 (fold 0)	0.3004	0.2386	0.1978	0.7237	0.7237	0.7237	0.2691
	DensNet-161 (fold 1)	0.3199	0.2695	0.2439	0.7306	0.7306	0.7306	0.3550
	Average	**0.3101**	0.2540	**0.2208**	0.7271	0.7271	0.7271	0.3121
	ResNet-152 (fold 0)	0.2656	0.2945	0.2126	0.6905	0.6905	0.6905	0.2822
	ResNet-152 (fold 1)	0.2591	0.2383	0.2143	0.7078	0.7078	0.7078	0.3222
	Average	0.2623	**0.2664**	0.2134	0.6992	0.6992	0.6992	0.3022
Weighted sampling	DensNet-161 (fold 0)	0.2608	0.2328	0.1786	0.7294	0.7294	0.7294	0.2598
	DensNet-161 (fold 1)	0.3212	0.2753	0.2474	0.7444	0.7444	0.7444	0.3793
	Average	0.2910	0.2540	0.2130	0.7369	0.7369	0.7369	0.3195
	ResNet-152 (fold 0)	0.2206	0.2372	0.1878	0.7317	0.7317	0.7317	0.2836
	ResNet-152 (fold 1)	0.2925	0.2776	0.2390	0.7566	0.7566	0.7566	0.3973
	Average	0.2565	0.2574	0.2134	0.7442	0.7442	0.7442	**0.3405**

means that users of the dataset can either use local medical experts to provide further labels, or use the data in unsupervised or semi-supervised learning approaches.

Technical Validation

To evaluate the technical quality of *Kvasir-Capsule*, we performed a series of classification experiments. We trained two CNN-based classifiers to classify the labelled data. Both architectures have previously shown excellent performance on classifying GI-related imagery from traditional colonoscopies[48,49], and should be a good benchmark for VCE-related data. The two algorithms are based on standard CNN architectures, namely DenseNet-161[50] and ResNet-152[51]. All experiments were performed over two-fold cross-validation using categorical cross-entropy loss with and without class weighting. We also tried using weighted sampling, which balances the dataset by removing and adding images for each class based on a given set of weights. To ensure a fair and robust evaluation, no video is shared between splits. The purpose of these experiments is two-fold. First, we create a baseline for future researchers using the *Kvasir-Capsule* dataset. Second, by using an algorithm that has previously shown good results on classifying GI images, we evaluate how challenging the task of categorizing VCE-related data is. Note that for the classification experiments, we removed the hematin and polyp classes due to the small number of findings. The results for the two classification algorithms are shown in Table 3 and confusion matrices for the best average MCC value in Figure 4. Considering the results, we see that classifying VCE data is quite a challenging task. For example, several of the classes are erroneously predicted as **Normal mucosa**. On the other hand, the class with the most accurate predictions is also **Normal mucosa**, reaching 75% in fold one and 81% in fold two. This is expected as the class makes up approximately 78% of the labelled images. This points out the challenges of making reliable systems as there are multiple aspects to consider, e.g., the resolution of VCE frames are lower compared to gastro- or colonoscopies, and many of the findings are subtle where even clinicians have difficulties differentiating between the classes. As seen when comparing the images in Figure 2, several findings are hard to see and easily mixed. For example, erosions can often be mistaken as small residues, and it can be difficult to differentiate normal mucosa from slight erythema. Thus, these results show the potential of AI-based analysis, but also further motivates the need to publish this dataset for more investigations and research into better specific algorithms for VCE data. The code used to conduct all experiments, produce all plots, and the images contained in each split is available on GitHub (`https://github.com/simula/kvasir-capsule`).

(a) Confusion matrix for model evaluated on split 0

(b) Confusion matrix for model evaluated on split 1

Figure 4. Confusion matrices for the best average MCC value which is from the weighted sampling technique. The labeling of the classes is as follows: (A) Angiectasia; (B) Blood; (C) Erosion; (D) Erythematous; (E) Foreign Bodies; (F) Ileo-cecal valve; (G) Lymphoid Hyperplasia; (H) Normal mucosa; (I) Pylorus; (J) Reduced Mucosal View; (K) Ulcer

Usage Notes

To the best of our knowledge, we have collected the largest and most diverse public available VCE dataset. *Kvasir-Capsule* is made available to enable researchers to develop detection or classification methods of various GI findings using for example computer vision and machine learning approaches. As the labelled findings also include bounding boxes, areas of potential use are analysis, classification, segmentation, and retrieval of images and videos of particular findings or properties. Moreover, the ground truths of various findings by the expert gastroenterologists provide a unique and diverse learning set for future clinicians, i.e., the labelled data can be used for teaching and training in medical education.

The unlabelled data is well suited for semi-supervised and unsupervised methods, and, if even more ground truth data is needed, the users of the data can use their own local medical experts to provide the needed labels. In this respect, recent work has shown remarkable improvements in the area of semi-supervised learning, also successfully applied in medical image analyses[34]. Instead of learning from a large set of annotated data, algorithms learn from sparsely labelled and unlabelled data. Self-learning[36,37] and neural graph learning[38] are both examples using unlabelled data in addition to a small amount of labelled data to extract additional information[35–37]. In an area with scarce data, these new algorithms might be the technology needed to make AI truly useful for medical applications.

There is currently a lot of research being performed in the field of GI image and video analysis, and we welcome and encourage future contributions in this area. This is not limited to using the dataset for comparisons and reproducibility of experiments, but also publishing and sharing new data in the future.

Code Availability

In addition to releasing the data, we also publish code used for the baseline experiments. All code and additional data required for the experiments are available on GitHub via `http://www.github.com/simula/kvasir-capsule`.

Acknowledgements

We would like to acknowledge various people at Bærum Hospital for making the data available. Moreover, the work is partially funded by the Research Council of Norway, project number 282315 (AutoCap).

Author Contributions Statement

S.A.H., V.T., P.H., M.A.R., P.H.S., and T.d.L. conceived the experiment(s), S.A.H. and V.T. conducted the experiment(s), P.H.S., H.G., O.O.N., E.N., V.T., S.A.H., M.A.R., P.H. and T.d.L. prepared and cleaned the data for publication, and all authors analysed the results and reviewed the manuscript.

Competing interests

Authors P.H.S., T.J.D.B., H.E., A.P., T.d.L., M.A.R., and P.H. all own shares in the Augere Medical AS company developing AI solutions for colonoscopies. The Augere video annotation system was used to label the data. There is no commercial interest from Augere regarding this publication and dataset. Otherwise, the authors declare no competing interests.

Ethical Approval

In this study, we used fully anonymized data approved by Privacy Data Protection Authority. It was exempted from approval from the Regional Committee for Medical and Health Research Ethics - South East Norway. Furthermore, we confirm that all experiments were performed in accordance with the relevant guidelines and regulations of the Regional Committee for Medical and Health Research Ethics - South East Norway, and the GDPR.

Rights and permissions

References

1. Greenwood-Van Meerveld, B., Johnson, A. C. & Grundy, D. Gastrointestinal physiology and function. In *Gastrointestinal Pharmacology*, 1–16 (Springer, 2017).
2. McLaughlin, P. D. & Maher, M. M. Primary malignant diseases of the small intestine. *Am. J. Roentgenol.* **201**, W9–W14 (2013).
3. Thomson, A. *et al.* Small bowel review: diseases of the small intestine. *Dig. diseases sciences* **46**, 2555–2566 (2001).
4. Enns, R. A. *et al.* Clinical practice guidelines for the use of video capsule endoscopy. *Gastroenterology* **152**, 497–514 (2017).
5. Costamagna, G. *et al.* A prospective trial comparing small bowel radiographs and video capsule endoscopy for suspected small bowel disease. *Gastroenterology* **123**, 999–1005 (2002).
6. Hewett, D. G., Kahi, C. J. & Rex, D. K. Efficacy and effectiveness of colonoscopy: how do we bridge the gap? *Gastrointest. Endosc. Clin.* **20**, 673–684, DOI: 10.1016/j.giec.2010.07.011 (2010).
7. Lee, S. H. *et al.* Endoscopic experience improves interobserver agreement in the grading of esophagitis by los angeles classification: conventional endoscopy and optimal band image system. *Gut liver* **8**, 154 (2014).
8. Van Doorn, S. C. *et al.* Polyp morphology: an interobserver evaluation for the paris classification among international experts. *The Am. J. Gastroenterol.* **110**, 180–187, DOI: 10.1038/ajg.2014.326 (2015).
9. Kaminski, M. F. *et al.* Quality indicators for colonoscopy and the risk of interval cancer. *New Engl. J. Medicine* **362**, 1795–1803, DOI: 10.1056/NEJMoa0907667 (2010).
10. Cave, D. R., Hakimian, S. & Patel, K. Current controversies concerning capsule endoscopy. *Dig. Dis. Sci.* **64**, 3040–3047, DOI: 10.1007/s10620-019-05791-4 (2019).
11. Rondonotti, E. *et al.* Can we improve the detection rate and interobserver agreement in capsule endoscopy? *Dig. Liver Dis.* **44**, 1006–1011 (2012).
12. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. medicine* **25**, 44 (2019).
13. Riegler, M. *et al.* Multimedia and medicine: Teammates for better disease detection and survival. In *Proceedings of the 24th ACM International Conference on Multimedia*, MM '16, 968–977, DOI: 10.1145/2964284.2976760 (ACM, New York, NY, USA, 2016).

14. Riegler, M. *et al.* EIR - efficient computer aided diagnosis framework for gastrointestinal endoscopies. In *Proceedings of the IEEE International Workshop on Content-Based Multimedia Indexing (CBMI)*, 1–6, DOI: 10.1109/CBMI.2016.7500257 (2016).

15. Alammari, A. *et al.* Classification of ulcerative colitis severity in colonoscopy videos using cnn. In *Proceedings of the ACM International Conference on Information Management and Engineering (ICIME)*, 139–144, DOI: 10.1145/3149572.3149613 (2017).

16. Wang, Y., Tavanapong, W., Wong, J., Oh, J. H. & De Groen, P. C. Polyp-alert: Near real-time feedback during colonoscopy. *Comput. Methods Programs Biomed.* **120**, 164–179, DOI: https://doi.org/10.1016/j.cmpb.2015.04.002 (2015).

17. Hirasawa, T., Aoyama, K., Fujisaki, J. & Tada, T. 113 application of artificial intelligence using convolutional neural network for detecting gastric cancer in endoscopic images. *Gastrointest. Endosc.* **87**, AB51 (2018).

18. Wang, L., Xie, C. & Hu, Y. Iddf2018-abs-0260 deep learning for polyp segmentation (2018).

19. Mori, Y. *et al.* Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Annals internal medicine* (2018).

20. Bychkov, D. *et al.* Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Reports* **8**, 1–11, DOI: 10.1038/s41598-018-21758-3 (2018).

21. Min, M. *et al.* Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. *Sci. reports* **9**, 2881 (2019).

22. Bernal, J. & Aymeric, H. Miccai endoscopic vision challenge polyp detection and segmentation. https://endovissub2017-giana.grand-challenge.org/home/ (2017). Accessed: 2017-12-11.

23. Tajbakhsh, N., Gurudu, S. R. & Liang, J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Med. Imaging* **35**, 630–644, DOI: 10.1109/TMI.2015.2487997 (2016).

24. Pogorelov, K. *et al.* Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, 164–169 (ACM, 2017).

25. Borgli, H. *et al.* HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Springer Nat. Sci. Data* (2020).

26. Yuan, Y. & Meng, M. Q.-H. A novel feature for polyp detection in wireless capsule endoscopy images. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5010–5015 (2014).

27. Yuan, Y. & Meng, M. Q.-H. Deep learning for polyp recognition in wireless capsule endoscopy images. *Med. Phys.* **44**, 1379–1389 (2017).

28. Karargyris, A. & Bourbakis, N. G. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. *IEEE Transactions on Biomed. Eng.* **58**, 2777–2786 (2011).

29. Leenhardt, R. *et al.* A neural network algorithm for detection of gi angiectasia during small-bowel capsule endoscopy. *Gastrointest. endoscopy* **89 1**, 189–194 (2019).

30. Pogorelov, K. *et al.* Deep learning and handcrafted feature based approaches for automatic detection of angiectasia. In *Proceedings of IEEE Conference on Biomedical and Health Informatics (BHI)*, 365–368, DOI: 10.1109/BHI.2018.8333444 (2018).

31. Pogorelov, K. *et al.* Bleeding detection in wireless capsule endoscopy videos—color versus texture features. *J. applied clinical medical physics* **20**, DOI: https://doi.org/10.1002/acm2.12662 (2019).

32. Rahim, T., Usman, M. A. & Shin, S. Y. A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging (2019). 1910.00265.

33. Yang, Y. J. The future of capsule endoscopy: The role of artificial intelligence and other technical advancements. *Clin. Endosc.* (2020).

34. Cheplygina, V., de Bruijne, M. & Pluim, J. P. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Analysis* **54**, 280–296, DOI: 10.1016/j.media.2019.03.009 (2019).

35. He, K., Fan, H., Wu, Y., Xie, S. & Girshick, R. Momentum contrast for unsupervised visual representation learning. *arXiv preprint arXiv:1911.05722* (2019).

36. Hénaff, O. J., Razavi, A., Doersch, C., Eslami, S. & Oord, A. v. d. Data-efficient image recognition with contrastive predictive coding. *arXiv preprint arXiv:1905.09272* (2019).

37. Misra, I. & van der Maaten, L. Self-supervised learning of pretext-invariant representations. *arXiv preprint arXiv:1912.01991* (2019).

38. Bui, T. D., Ravi, S. & Ramavajjala, V. Neural graph learning: Training neural networks using graphs. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*, 64–71, DOI: 10.1145/3159652.3159731 (2018).

39. Koulaouzidis, A. *et al.* Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes. *Endosc. international open* **5**, E477–E483, DOI: 10.1055/s-0043-105488 (2017).

40. Bernal, J. & Aymeric, H. Gastrointestinal Image ANAlysis (GIANA) Angiodysplasia D&L challenge. https://endovissub2017-giana.grand-challenge.org/home/ (2017). Accessed: 2017-11-20.

41. Angermann, Q. *et al.* Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis. In *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures*, 29–41 (Springer, 2017).

42. Bernal, J. *et al.* Polyp detection benchmark in colonoscopy videos using gtcreator: A novel fully configurable tool for easy and fast annotation of image databases (2018).

43. Computer-assisted diagnosis for capsule endoscopy (cad-cap) database. https://giana.grand-challenge.org/WCE/ (2019).

44. Leenhardt, R. *et al.* Cad-cap: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy. *Endosc. international open* **8**, E415 (2020).

45. Aabakken, L. *et al.* Standardized endoscopic reporting. *J. Gastroenterol. Hepatol.* **29**, 234–240, DOI: 10.1111/jgh.12489 (2014).

46. Chetcuti Zammit, S. *et al.* Overview of small bowel angioectasias: clinical presentation and treatment options. *Expert. review gastroenterology & hepatology* **12**, 125–139 (2018).

47. Gomollón, F. *et al.* 3rd european evidence-based consensus on the diagnosis and management of crohn's disease 2016: part 1: diagnosis and medical management. *J. Crohn's Colitis* **11**, 3–25 (2017).

48. Thambawita, V. *et al.* An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification. *ACM Trans. Comput. Healthc.* **1**, DOI: 10.1145/3386295 (2020).

49. Thambawita, V. *et al.* The medico-task 2018: Disease detection in the gastrointestinal tract using global features and deep learning. In *Proceedings of the MediaEval 2018 Workshop* (2018).

50. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proc. of IEEE CVPR*, 2261–2269 (2017).

51. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. of IEEE CVPR*, 770–778 (2016).