


scientific data



OPEN

DATA DESCRIPTOR

Kvasir-Capsule, a video capsule endoscopy dataset

Pia H. Smedsrud ^{1,3,6,15}✉, Vajira Thambawita ^{1,2,15}, Steven A. Hicks^{1,2,15}, Henrik Gjestang ^{1,3}, Oda Olsen Nedrejord ^{1,3}, Espen Næss^{1,3}, Hanna Borgli ^{1,3}, Debesh Jha ^{1,7,15}, Tor Jan Derek Berstad⁶, Sigrun L. Eskeland⁴, Mathias Lux¹⁰, Håvard Espeland ⁶, Andreas Petlund ⁶, Duc Tien Dang Nguyen⁵, Enrique Garcia-Ceja ¹³, Dag Johansen⁷, Peter T. Schmidt^{8,9}, Ervin Toth¹⁴, Hugo L. Hammer^{1,2}, Thomas de Lange ^{4,6,11,12,15,16}, Michael A. Riegler ^{1,15,16} & Pål Halvorsen ^{1,2,15,16}

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. Existing work demonstrates the promising benefits of AI-based computer-assisted diagnosis systems for VCE. They also show great potential for improvements to achieve even better results. Also, medical data is often sparse and unavailable to the research community, and qualified medical personnel rarely have time for the tedious labelling work. We present *Kvasir-Capsule*, a large VCE dataset collected from examinations at a Norwegian Hospital. *Kvasir-Capsule* consists of 117 videos which can be used to extract a total of 4,741,504 image frames. We have labelled and medically verified 47,238 frames with a bounding box around findings from 14 different classes. In addition to these labelled images, there are 4,694,266 unlabelled frames included in the dataset. The *Kvasir-Capsule* dataset can play a valuable role in developing better algorithms in order to reach true potential of VCE technology.

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 30 m², including the villi's surface. As part of the digestive system, it plays a crucial role in absorbing nutrients¹. Therefore, disorders in the small bowel may cause severe growth retardation in children and nutrient deficiencies in children and adults¹. 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 the patients and the society, and a thorough examination of the lumen is frequently necessary to diagnose and treat them⁴. However, due to its anatomical location, the small bowel 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 bleeding⁴. A VCE consists of a small capsule containing a wide-angle camera, light sources, batteries, and other electronics. The patient swallows the capsule capturing a video as it moves passively throughout the GI tract. A recorder, carried by the patient or included in the capsule, stores the video before a medical expert examines it after the procedure.

VCE devices exist in various versions and brands such as Given Imaging (Medtronic), 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

¹SimulaMet, Oslo, Norway. ²Oslo Metropolitan University, Oslo, Norway. ³University of Oslo, Oslo, Norway.

⁴Department of Medical Research, Bærum Hospital, Gjetlum, Norway. ⁵University of Bergen, Bergen, Norway.

⁶Augere Medical AS, Oslo, Norway. ⁷UIT The Arctic University of Norway, Tromsø, Norway. ⁸Karolinska Institutet,

Department of Medicine, Solna, Sweden. ⁹Ersta Hospital, Department of Medicine, Stockholm, Sweden.

¹⁰Klagenfurt University, Wörthersee, Austria. ¹¹Medical Department, Sahlgrenska University Hospital-Mölnad

Hospital, Göteborg, Sweden. ¹²Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University

of Gothenburg, Göteborg, Sweden. ¹³SINTEF Digital, Oslo, Norway. ¹⁴Department of Gastroenterology, Skåne

University Hospital, Malmö Lund University, Malmö, Sweden. ¹⁵These authors contributed equally: Pia H. Smedsrud,

Vajira Thambawita, Steven A. Hicks, Debesh Jha, Thomas de Lange, Michael A. Riegler, Pål Halvorsen. ¹⁶These

authors jointly supervised: Thomas de Lange, Michael A. Riegler, Pål Halvorsen. ✉e-mail: pia@simula.no

Dataset	Findings	Size	Availability
KID ⁵⁴	Angiectasia, bleeding, inflammations, polyps	2,371 images + 47 videos	open academic*
GIANA 2017 ⁵⁵	Angiectasia†	600 images	by request
GIANA2018 ^{56,57}	Polyps and small bowel lesions†	8262 images + 38 videos	by request
CAD-CAP ^{58,59}	Normal frames, fresh blood, vascular lesion, ulcerative and inflammatory lesions	25,000 images	by request◇
Gastrolab ⁶⁰	Crohns diseases, small bowel (video)+ GI lesions	Few hundred images and videos	open academic*

Table 1. An overview of existing VCE datasets from the GI tract. †Including ground truth segmentation masks. *Not available anymore. ◇The Computer-Assisted Diagnosis for CAPSule endoscopy (CAD-CAP) Database - used for the angiectasia detection.

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 rates⁹. In VCE analysis, essential findings are missed due to lack of concentration, insufficient experience and knowledge^{10–12}. Furthermore, physicians may have trouble handling the associated technology, and infrequent VCE use leads to lack of confidence¹³, resulting in inter-observer and intra-observer variations in the assessments¹².

The technical developments for automated image and video analysis have sky-rocketed, and multimedia solutions in medicine show tremendous potential^{14,15}. An increasing number of promising machine learning solutions are being developed for automated diagnosis of colonoscopies^{16–23} using open datasets^{24–27}. Regarding automated VCE data analyses, machine learning approaches also produce promising results regarding detection and classification rates^{28–35}. Machine learning, or artificial intelligence (AI) in general, is likely to have profound effects on the VCE technology's future, not only for improving variation and detection rates but also for estimating the capsule's localisation^{13,36}.

Regardless of promising initial results, there is room for improvements in detection rate, reduced manual labour, and AI explainability. Large amounts of data are needed^{37,38}, particularly annotated data³⁵, and access to these data are often scarce³⁹. As shown in Table 1, very few, small VCE datasets are made publicly available, and several have become unavailable. We have previously published the HyperKvasir dataset²⁷. Nevertheless, this and similar datasets containing images from *colonoscopies* and *esophagogastrosopies* 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. Also, the image resolution and the frame rate of VCEs are much lower. The small bowel is not air inflated during a VCE examination, as is the case with traditional colonoscopies. Different optics are also used, and the movement of the capsule is uncontrolled in contrast to flexible endoscopes used during manual examinations.

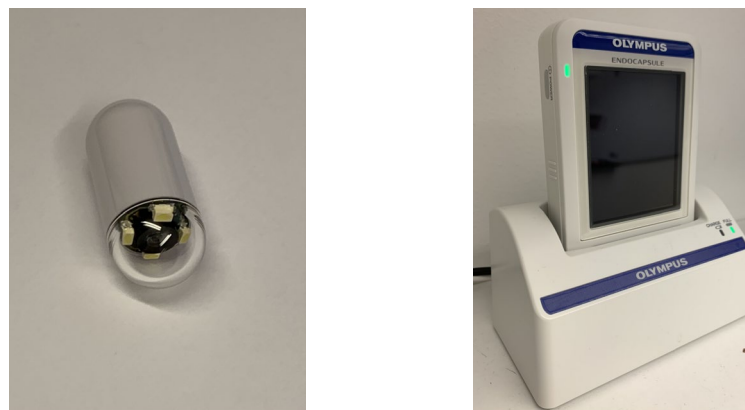
Therefore, we present a large VCE dataset, called *Kvasir-Capsule*, consisting of 117 videos with 4,741,504 frames and 14 classes of findings. The dataset contains labelled images and their corresponding full videos, and also unlabelled videos. Recent work in the machine learning community has shown significant improvements regarding sparsely labelled and unlabelled data value. Semi-supervised learning algorithms are successfully applied in different medical image analyses^{40,41} using self-learning^{42,43} and neural graph learning⁴⁴. Finally, we provide a baseline analysis and outline possible future research directions using *Kvasir-Capsule*.

Methods

The VCE videos were collected from consecutive clinical examinations performed at the Department of Medicine, Bærum Hospital, Vestre Viken Hospital Trust in Norway, which provides health care services to 490,000 people, of which about 200,000 are covered by Bærum Hospital. The examinations were conducted between February 2016 and January 2018 using the Olympus Endocapsule 10 System⁴⁵ including the Olympus EC-S10 endocapsule (Fig. 1a) and the Olympus RE-10 endocapsule recorder (Fig. 1b). Originally, the videos were captured at a rate of 2 frames per second, in a resolution of 336 × 336, and encoded using H.264 (MPEG-4 AVC, part 10). The videos were exported in AVI format using the Olympus system's export tool packaged and encapsulated in the same H.264 format, i.e., the frame formats are the same, but the frame rate specification is changed to 30 fps by the export tool.

Initially, a trained clinician analysed all videos using the Olympus software, selecting thumbnails from lesions and normal findings as part of their clinical work. In spring 2019, all the 117 anonymous videos and thumbnails were exported from a stand-alone workstation using the Olympus software. The Olympus video capsule system has user-friendly functionalities like Omni-selected Mode, skipping images that overlap with previous ones.

All metadata were removed and files renamed with randomly generated file names, before exporting the videos and thumbnails that were shared. Thus, data in the dataset are fully anonymized, as approved by Privacy Data Protection Authority and in accordance with relevant guidelines and regulations of the Regional Committee for Medical and Health Research Ethics - South East Norway. The data has not been pre-processed or augmented in any way apart from this. Subsequently, for clinical analyses of the videos, a central expert reader selected and categorized thumbnails with 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. Next, three master students labelled and marked the findings with bounding boxes for each frame. The bounding boxes were designed to include the entire lesion and as little as possible of the surrounding mucosa. If the students were unsure about the labelling, the expert reader verified the frames. All labels regarding anatomical structures and normal clean mucosa were then confirmed by one junior



(a) Olympus EC-S10 endocapsule **(b)** Olympus RE-10 endocapsule recorder

Fig. 1 VCE equipment used for data collection.

Data Record	# Files
Labelled images	47,238
Labelled videos	43
Unlabelled images	4,694,266
Unlabelled videos	74

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

medical doctor and the expert reader. Finally, all the annotations were once more verified by the expert reader and subsequently validated by a second expert reader. If the second reviewer disagreed with the annotations, the first reviewer reassessed the images to see whether he then agreed with the second reviewer to get an agreement. After the validation process by the second reviewer there was a disagreement on twenty-six findings in seven examinations; nineteen concerning erroneous terminology of the class lymphoid hyperplasia which was changed to lymphangiectasia. The other seven were related to the interpretation of the finding. After reviewing these findings, the first reviewer agreed with the second one to finally reach a perfect agreement. After this procedure, the video frames were exported as images. Hence, a total of four medical persons have selected, analysed and verified the data, and a total of 47,238 frames are labelled.

The Norwegian Privacy Data Protection Authority approved the export of anonymous images for the creation of the database, without consent from participants. 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)⁴⁶. Table 2 gives an overview of all data records in the dataset. In total, the dataset consists of 4,741,621 main data records, i.e., 47,238 images with labels and bounding box masks, the 43 corresponding labelled videos (the videos from which the images are extracted), and 74 unlabelled videos (from which labelled images have not been extracted). 4,694,266 unlabelled images can further be extracted from all the videos combined. All the various labelled classes are shown in Fig. 2. The dataset has a total size of circa 89 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.

The dataset is stored according to the data records listed above, and described in more detail below. We have a “labelled images” catalogue which contains archive files of each labelled class of images. We have a “labelled videos” catalogue which contains all the videos where we have annotated findings from, and an “unlabelled videos” catalogue containing the videos that are not annotated.

Labelled images. In total, the dataset contains 47,238 labelled images stored using the PNG format, where Fig. 3 shows the 14 different classes representing the labelled images and the number of images in each class. The provided *metadata.csv* comma-separated value (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. Some samples are given in Fig. 4 where the first line gives the type of each element in the lines below. This means that the file *filename* of the labelled image which is the frame *frame_number* extracted from the *video_id* video. Moreover, the finding is from the category *finding_category* and class *finding_class*. Finally, the four x_i, y_i pairs are the four pixel coordinates for the bounding box, e.g., in the first three lines they are empty, meaning that there is no finding with a bounding box in this labelled image. There is one line in the file per each labelled image.

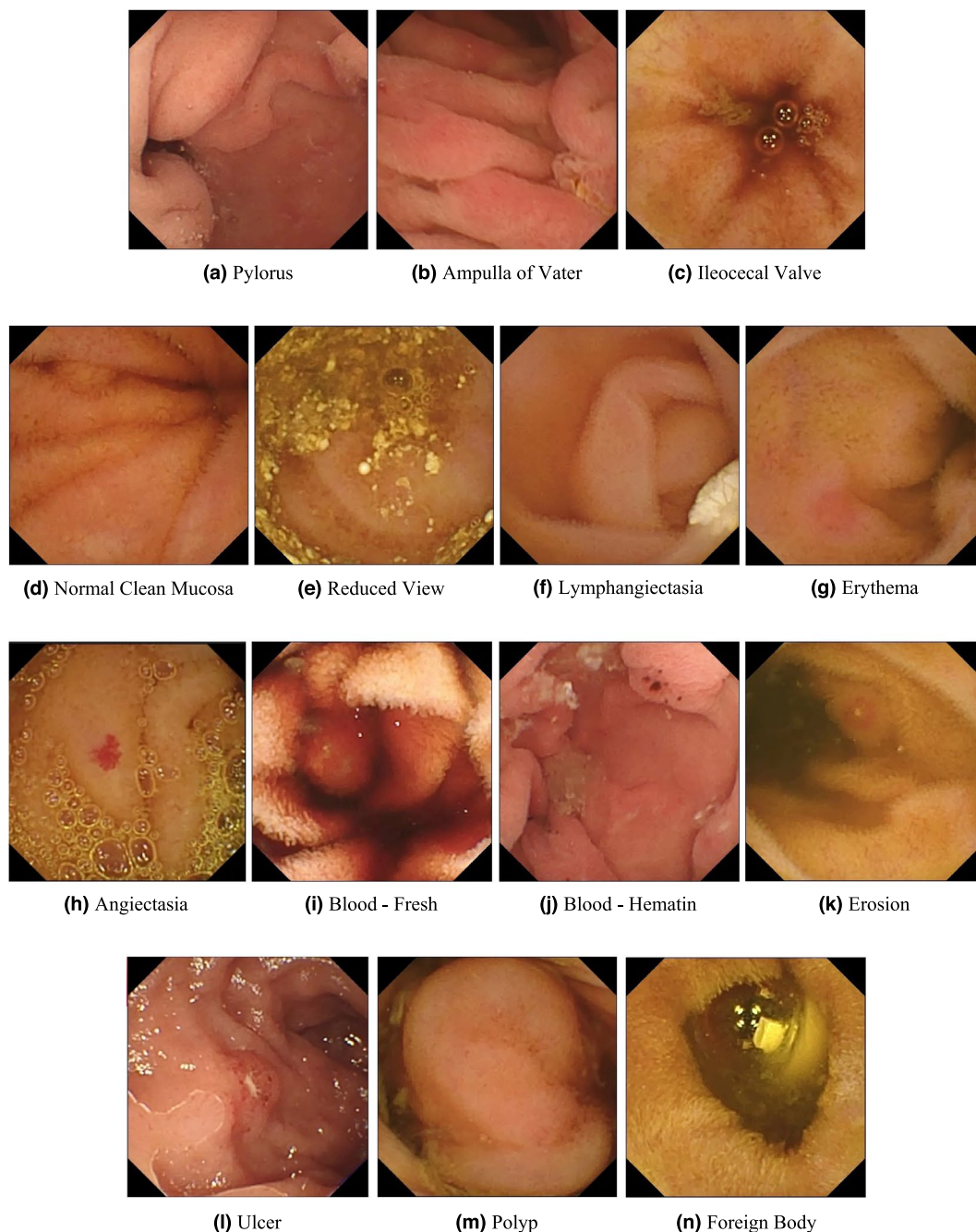


Fig. 2 Image examples of the various labelled classes for images. Images (a) to (c) are from the *Anatomy* category, and images (d) to (n) are from the *Luminal findings* category.

We defined two main categories of findings, namely anatomy and luminal findings. Each category, their classes and belonging images are stored in their corresponding folder. As observed in Fig. 3, the number of images per class is not balanced. This is a global challenge in the medical field because some findings are more common than others, which adds a challenge for researchers since methods applied to the data should also be able to learn from a small amount of training data.

Categories of findings. We have organised the dataset in two main categories with their corresponding classes according to the World Endoscopy Association Minimal Standard Terminology version 3.0 (MST 3.0), though we have not included the subcategories or intermediate level to simplify the dataset⁴⁷.

Anatomy. The category of *Anatomy* contains anatomical landmarks characterising the GI tract. These landmarks may be used for orientation during endoscopic procedures. However, for small bowel VCE their role is to verify the passage of the capsule through the entire small bowel to confirm a complete examination. We have labelled three anatomical landmarks, the first two delineate the upper (proximal) and lower (distal) end of the

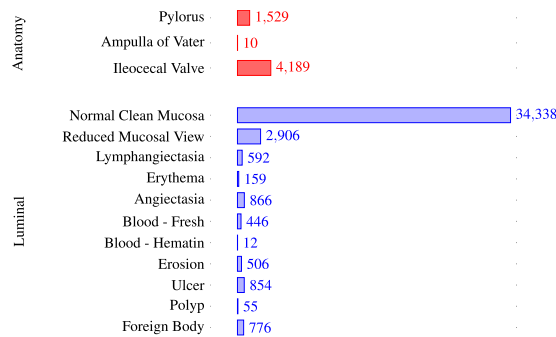


Fig. 3 The number of images in the various Kvasir-Capsule labelled image classes.

```
filename;video_id;frame_number;finding_category;finding_class;x1;y1;x2;y2;x3;y3;x4;y4
...
0728084c8da942d9_22805.jpg;0728084c8da942d9;22805;Luminal;Normal clean mucosa;;;;;;;;;
0728084c8da942d9_22806.jpg;0728084c8da942d9;22806;Luminal;Normal clean mucosa;;;;;;;;;
0728084c8da942d9_22807.jpg;0728084c8da942d9;22807;Luminal;Normal clean mucosa;;;;;;;;;
...
0728084c8da942d9_28789.jpg;0728084c8da942d9;28789;Luminal;Erosion;195;226;244;226;244;265;195;265
0728084c8da942d9_28798.jpg;0728084c8da942d9;28798;Luminal;Erosion;183;212;213;212;213;265;183;265
0728084c8da942d9_28799.jpg;0728084c8da942d9;28799;Luminal;Erosion;197;213;229;213;229;267;197;267
...
```

Fig. 4 Samples from the *metadata.csv* CSV file.

small bowel, respectively. 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. The third one, the **ampulla of Vater**, is the junction between the duodenum and the gall duct.

Luminal findings. Endoscopic examinations may detect various *luminal findings*, this include the subcategories content of the bowel lumen, the aspect of the mucosa and mucosal lesions (pathological findings) that could be either flat, elevated or excavated. These subcategories are not shown in the dataset. Normally, the small bowel contains only a certain amount of yellow or brown liquid considered as clean mucosa. However, larger amounts of content may preclude a complete visualisation of the mucosa crucial to verify normal mucosa and detection of all pathological (abnormal) findings. For the lumen content assessment, we have labelled five classes. **Normal clean mucosa** depicts clean small bowel with no or small amount of fluid and mucosa with healthy villi and no pathological findings. This class can also double as a “normal” class versus the pathological luminal finding class (see below). The class **reduced mucosal view** shows small bowel content reducing the view of the mucosa, like stool or bubbles. However, lesions in the upper GI tract or small bowel may bleed, causing the appearance of **blood - fresh** colouring the liquid red. In cases with minimal bleeding, one may observe small black stripes called **blood - hematin** on the mucosal surface. The **foreign body** class include tablet residue or retained capsules which can also be observed in the lumen.

Abnormalities, called lesions or pathological findings, in the small bowel can be seen as changes to the mucosal surface. Typical mucosal changes sometimes cover larger segments, such as a reddish appearance called erythematous mucosa, is labelled as **erythema**. The mucosal wall can also have different focal lesions. The classes of 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 diseases⁴⁸. 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 difficult⁴⁹. **Lymphangiectasia**, which represents dilated lymphoid vessels in the mucosal wall, and **polyps**, which may be precancerous lesions, are visible as protruding from the mucosal wall.

Labelled videos. Labelled videos are the full 43 videos from which we extracted the above mentioned labelled image classes. In total, these videos correspond to approximately 19 hours of video and 47,238 labelled video frames. Several segments of each video was labelled, and these segments are what was exported as the labelled images. 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 1,932,047 non-labelled frames that could be interesting in future research. One could also extract the video sequences around the various findings.

Method		Macro average			Micro average			
		Precision	Recall	F1-score	Precision	Recall	F1-score	MCC
Normal CEL	DensNet-161 (fold 0)	0.2165	0.2341	0.1923	0.7375	0.7375	0.7375	0.3707
	DensNet-161 (fold 1)	0.3493	0.3158	0.2996	0.7327	0.7327	0.7327	0.4604
	Average	0.2829	0.2749	0.2459	0.7351	0.7351	0.7351	0.4156
	ResNet-152 (fold 0)	0.3302	0.2401	0.1970	0.7203	0.7203	0.7203	0.3520
	ResNet-152 (fold 1)	0.3431	0.2805	0.2789	0.7481	0.7481	0.7481	0.4718
	Average	0.3367	0.2603	0.2379	0.7342	0.7342	0.7342	0.4119
Weighted CEL	DensNet-161 (fold 0)	0.2933	0.2939	0.2523	0.7195	0.7195	0.7195	0.3998
	DensNet-161 (fold 1)	0.3163	0.2914	0.2581	0.6991	0.6991	0.6991	0.4054
	Average	0.3048	0.2927	0.2552	0.7093	0.7093	0.7093	0.4026
	ResNet-152 (fold 0)	0.2136	0.2872	0.2186	0.6568	0.6568	0.6568	0.3588
	ResNet-152 (fold 1)	0.3033	0.2799	0.2478	0.6890	0.6890	0.6890	0.3966
	Average	0.2585	0.2836	0.2332	0.6729	0.6729	0.6729	0.3777
Weighted sampling	DensNet-161 (fold 0)	0.2525	0.2794	0.2315	0.7332	0.7332	0.7332	0.4111
	DensNet-161 (fold 1)	0.3463	0.2830	0.2806	0.7400	0.7400	0.7400	0.4547
	Average	0.2994	0.2812	0.2560	0.7366	0.7366	0.7366	0.4329
	ResNet-152 (fold 0)	0.2637	0.2930	0.2334	0.7324	0.7324	0.7324	0.4088
	ResNet-152 (fold 1)	0.3088	0.2619	0.2417	0.7316	0.7316	0.7316	0.4520
	Average	0.2862	0.2774	0.2375	0.7320	0.7320	0.7320	0.4304

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.

Unlabelled videos. We also provide 74 videos, which contain approximately 25 hours of video and 2,762,219 video frames, without any labels. As previously mentioned, unlabelled data can still have great value. Sparsely labelled or unlabelled data can be important for recently emerging semi-supervised learning algorithms. These videos are of the same format and quality as the labelled videos, except we do not provide any annotations. This means that users of the dataset can either use 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^{50,51}, and should be a good benchmark for VCE-related data. The two algorithms are based on standard CNN architectures, namely DenseNet-161⁵² and ResNet-152⁵³. All experiments were performed over two-fold cross-validation using categorical cross-entropy loss with and without class weighting. We also used 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. Thus, the frames used for training were independent from the frames used for validation. This also means that there are frames depicting the same finding in each split which then are related to each other, but no related frames distributed across the splits. The effect should therefore be similar to traditional data augmentation techniques used by many researchers today such as multiple rotations, angles and crops.

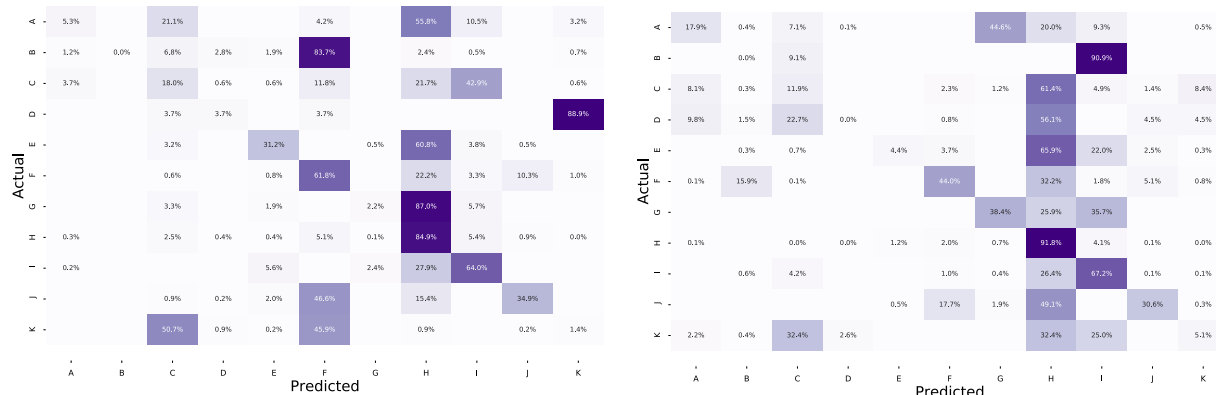
The purpose of these experiments is two-fold. First, we create a baseline for future researchers using the *Kvasir-Capsule* dataset. Second, by using an algorithm that has previously shown good results on classifying GI images, we evaluate how challenging the task of categorizing VCE-related data is. Note that for the classification experiments, we removed the blood - hematin, ampulla of Vater, 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 Fig. 5. We estimated micro-averaged and macro-averaged values for precision, recall and F1-score for each method. The Matthews correlation coefficient (MCC) was calculated using the multi-class generalization, also called the R_K . In short, if TP, TN, FP, and FN are the true positives, true negatives, false positives, and false negatives, respectively, these metrics are defined as follows²⁶:

Precision. This metric is also frequently called the *positive predictive value*, and shows the ratio of samples that are correctly identified as positive among the returned samples (the fraction of retrieved samples that are relevant):

$$precision = \frac{TP}{\# \text{ of all returned samples}} = \frac{TP}{TP + FP}$$

Recall. This metric is also frequently called *sensitivity*, *probability of detection* and *true positive rate*, and it is the ratio of samples that are correctly identified as positive among all existing positive samples:

$$recall = \frac{TP}{\# \text{ of all positives}} = \frac{TP}{TP + FN}$$



(a) Confusion matrix for model evaluated on split 0.

(b) Confusion matrix for model evaluated on split 0.

Fig. 5 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 - fresh; (C) Erosion; (D) Erythema; (E) Foreign Body; (F) Ileocecal valve; (G) Lymphangiectasia; (H) Normal clean mucosa; (I) Pylorus; (J) Reduced Mucosal View; (K) Ulcer.

F1 score (F1). A measure of a test's accuracy by calculating the harmonic mean of the precision and recall:

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall} = \frac{2TP}{2TP + FP + FN}$$

Matthews correlation coefficient (MCC). MCC takes into account true and false positives and negatives, and is a balanced measure even if the classes are of very different sizes. For the multiclass classification generalization, it is often called the R_k statistic. In following equation, t_k is the number of times class k actually occurred, p_k is the number of times class k was predicted, c is the total number of samples correctly predicted, and s is the total number of samples:

$$MCC = \frac{c \times s - \sum_k^K p_k \times t_k}{\sqrt{(s^2 - \sum_k^K p_k^2) \times (s^2 - \sum_k^K t_k^2)}}$$

The micro and macro averages are different ways to average metrics calculated over multiple classes. The macro average is the arithmetic mean of all the scores of different classes, i.e., calculates the metric per class and then calculates the average of these over the number of classes. For example, it is defined for precision as the sum of precision scores for all classes ($precision_1 + \dots + precision_n$) divided by the number of classes (n). The micro average is not counting class wise first, but looking at the total number of true and false findings. For example, for precision, it is defined as sum of true positives ($TP_1 + \dots + TP_n$) for all the n classes divided by the all returned positive predictions ($TP_1 + FP_1 + \dots + TP_n + FP_n$).

Considering the results, we experience that classifying VCE data is quite a challenging task. For example, several of the classes are erroneously predicted as **Normal clean mucosa**. On the other hand, the class with the most accurate predictions is also **Normal clean mucosa**, reaching 85% in fold one and 91% in fold two. This is expected as the class comprise approximately 73% 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 noticed when comparing the images in Fig. 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 are available on GitHub (<https://github.com/simula/kvasir-capsule>), i.e., to increase reproducibility and facilitate researches to perform comparable experiments on the *Kvasir-Capsule* dataset.

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 machine learning methods, and, if even more ground truth data is needed, the users of the data can have medical experts 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 analyses⁴⁰. Instead of learning from a large set of annotated data, algorithms learn from sparsely labelled and unlabelled data. Self-learning^{42,43} and neural graph learning⁴⁴ are both examples using unlabelled data in addition to a small amount of labelled data to extract additional information^{41–43}. In an area with scarce data, these new algorithms might be the technology needed to make AI truly useful for medical applications.

An important note in general for this type of AI-based detection systems is that one should be careful about how the dataset is split into for example training and test sets in order to avoid having related frames in several of the sets. This will give an unfair effect on the results. Thus, the splits should be completely different, probably even at the level of patients. As described below, an example of such a split is found in our GitHub repository (see below in the Code Availability section).

Currently, there is substantial research in GI image and video analysis. We welcome future contributions such as using the dataset for comparisons and reproducibility of experiments and further encourage publishing and sharing of new data. *Kvasir-Capsule* is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original authors and the source.

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, including our splits into training and test datasets, are available on GitHub via <http://www.github.com/simula/kvasir-capsule>.

Received: 13 August 2020; Accepted: 15 April 2021;

Published online: 27 May 2021

References

- Greenwood-Van Meerveld, B., Johnson, A. C. & Grundy, D. Gastrointestinal physiology and function. In *Gastrointestinal Pharmacology*, 1–16 (Springer, 2017).
- McLaughlin, P. D. & Maher, M. M. Primary malignant diseases of the small intestine. *American Journal of Roentgenology* **201**, W9–W14 (2013).
- Thomson, A. *et al.* Small bowel review: diseases of the small intestine. *Digestive diseases and sciences* **46**, 2555–2566 (2001).
- Enns, R. A. *et al.* Clinical practice guidelines for the use of video capsule endoscopy. *Gastroenterology* **152**, 497–514 (2017).
- 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).
- Hewett, D. G., Kahi, C. J. & Rex, D. K. Efficacy and effectiveness of colonoscopy: how do we bridge the gap? *Gastrointestinal Endoscopy Clinics* **20**, 673–684 (2010).
- 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 and liver* **8**, 154 (2014).
- Van Doorn, S. C. *et al.* Polyp morphology: an interobserver evaluation for the paris classification among international experts. *The American Journal of Gastroenterology* **110**, 180–187 (2015).
- Kaminski, M. F. *et al.* Quality indicators for colonoscopy and the risk of interval cancer. *New England Journal of Medicine* **362**, 1795–1803 (2010).
- Zheng, Y., Hawkins, L., Wolff, J., Goloubeva, O. & Goldberg, E. Detection of lesions during capsule endoscopy: physician performance is disappointing. *American Journal of Gastroenterology* **107**, 554–560 (2012).
- Chetcuti, S. Z. & Sidhu, R. Capsule endoscopy—recent developments and future directions. *Expert review of gastroenterology & hepatology* **15**, 127–137 (2021).
- Rondonotti, E. *et al.* Can we improve the detection rate and interobserver agreement in capsule endoscopy? *Digestive and Liver Disease* **44**, 1006–1011 (2012).
- Cave, D. R., Hakimian, S. & Patel, K. Current controversies concerning capsule endoscopy. *Digestive Diseases and Sciences* **64**, 3040–3047 (2019).
- Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine* **25**, 44 (2019).
- Riegler, M. *et al.* Multimedia and medicine: Teammates for better disease detection and survival. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, 968–977 (2016).
- 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 (2016).
- 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 (2017).
- Wang, Y., Tavanapong, W., Wong, J., Oh, J. H. & De Groen, P. C. Polyp-alert: Near real-time feedback during colonoscopy. *Computer Methods and Programs in Biomedicine* **120**, 164–179 (2015).
- Hirasawa, T., Aoyama, K., Fujisaki, J. & Tada, T. 113 application of artificial intelligence using convolutional neural network for detecting gastric cancer in endoscopic images. *Gastrointestinal Endoscopy* **87**, AB51 (2018).
- Wang, L., Xie, C. & Hu, Y. Iddf2018-abs-0260 deep learning for polyp segmentation. *BMJ Publishing Group* (2018).
- Mori, Y. *et al.* Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Annals of internal medicine* **169**, 357–366 (2018).
- Bychkov, D. *et al.* Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific Reports* **8**, 1–11 (2018).
- Min, M. *et al.* Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. *Scientific reports* **9**, 2881 (2019).
- Bernal, J. & Aymeric, H. Miccai endoscopic vision challenge polyp detection and segmentation. *Web-page of the 2017 Endoscopic Vision Challenge*, <https://endovissub2017-giana.grand-challenge.org/home/> (2017).
- Tajbakhsh, N., Gurudu, S. R. & Liang, J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medical Imaging* **35**, 630–644 (2016).
- Pogorelov, K. *et al.* Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the ACM on Multimedia Systems Conference (MMSYS)*, 164–169 (2017).
- Borgli, H. *et al.* Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data* **7**, 1–14 (2020).

28. Yuan, Y. & Meng, M. Q.-H. A novel feature for polyp detection in wireless capsule endoscopy images. In *Proceedings of the IEEE/RISJ International Conference on Intelligent Robots and Systems*, 5010–5015 (2014).
29. Yuan, Y. & Meng, M. Q.-H. Deep learning for polyp recognition in wireless capsule endoscopy images. *Medical Physics* **44**, 1379–1389 (2017).
30. Karargyris, A. & Bourbakis, N. G. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. *IEEE Transactions on Biomedical Engineering* **58**, 2777–2786 (2011).
31. Leenhardt, R. *et al.* A neural network algorithm for detection of gi angiectasia during small-bowel capsule endoscopy. *Gastrointestinal endoscopy* **89** **1**, 189–194 (2019).
32. 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 (2018).
33. Pogorelov, K. *et al.* Bleeding detection in wireless capsule endoscopy videos—color versus texture features. *Journal of applied clinical medical physics* **20** (2019).
34. 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. *Computerized Medical Imaging and Graphics* **85**, 101767 (2020).
35. Soffer, S. *et al.* Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. *Gastrointestinal Endoscopy* (2020).
36. Yang, Y. J. The future of capsule endoscopy: The role of artificial intelligence and other technical advancements. *Clinical Endoscopy* **53**, 387 (2020).
37. Park, J. *et al.* Recent development of computer vision technology to improve capsule endoscopy. *Clinical endoscopy* **52**, 328 (2019).
38. Iakovidis, D. K. & Koulaouzidis, A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nature Reviews Gastroenterology & Hepatology* **12**, 172–186 (2015).
39. Jani, K. K. & Srivastava, R. A survey on medical image analysis in capsule endoscopy. *Current Medical Imaging* **15**, 622–636 (2019).
40. 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. *Medical Image Analysis* **54**, 280–296 (2019).
41. He, K., Fan, H., Wu, Y., Xie, S. & Girshick, R. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738 (2020).
42. Henaff, O. Data-efficient image recognition with contrastive predictive coding. In *International Conference on Machine Learning*, 4182–4192 (PMLR, 2020).
43. Misra, I. & Maaten, L. V. D. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6707–6717 (2020).
44. 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 (2018).
45. Olympus. The endocapsule 10 system. *Olympus homepage*, <https://www.olympus-europa.com/medical/en/Products-and-Solutions/Products/Product/ENDOCAPSULE-10-System.html> (2013).
46. Thambawita, V. *et al.* The kvasir-capsule dataset. *Open Science Framework* <https://doi.org/10.17605/OSF.IO/DV2AG> (2020).
47. Aabakken, L. *et al.* Standardized endoscopic reporting. *Journal of Gastroenterology and Hepatology* **29**, 234–240 (2014).
48. Chetcuti Zammit, S. *et al.* Overview of small bowel angioectasias: clinical presentation and treatment options. *Expert review of gastroenterology & hepatology* **12**, 125–139 (2018).
49. 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. *Journal of Crohn's and Colitis* **11**, 3–25 (2017).
50. 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 Transactions on Computing for Healthcare* **1**, 1–29 (2020).
51. 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).
52. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269 (2017).
53. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778 (2016).
54. Koulaouzidis, A. *et al.* Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes. *Endoscopy international open* **5**, E477–E483 (2017).
55. Bernal, J. & Aymeric, H. Gastrointestinal Image ANALysis (GIANA) Angiodysplasia D&L challenge. *Web-page of the 2017 Endoscopic Vision Challenge*, <https://endovissub2017-giana.grand-challenge.org/home/> (2017).
56. 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).
57. 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. In *Proceedings of 32nd CARS conference* (2018).
58. Computer-assisted diagnosis for capsule endoscopy (cad-cap) database. *The 2019 GIANA Grand Challenge web-page*, <https://giana.grand-challenge.org/WCE/> (2019).
59. Leenhardt, R. *et al.* Cad-cap: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy. *Endoscopy international open* **8**, E415 (2020).
60. Gastrolab. *The Gastrointestinal Site*, <http://www.gastrolab.net/index.htm> (1996).

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 (RCN), project number 282315 (AutoCap), and our experiments have been performed on the Experimental Infrastructure for Exploration of Exascale Computing (eX3) also supported by RCN, contract 270053.

Author contributions

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., D.J., 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.

Additional information

Correspondence and requests for materials should be addressed to P.H.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021