Journal of Systems Architecture 177 (2026) 103851

Contents lists available at ScienceDirect

Journal of Systems Architecture

journal homepage: www.elsevier.com/locate/sysarc

WCE-SqueezeNet: Targeting on-capsule CNN inference for energy efficient microcontroller-based endoscopy☆

Paola Busia [a],*, Andrea Pinna [b], Paolo Meloni [a]

[a] DIEE, University of Cagliari, Cagliari, 09123, Italy
[b] Sorbonne University, CNRS, LIP6, Paris, 75005, France

ARTICLE INFO

Keywords:
Wireless Capsule Endoscopy
Near-sensor processing
Convolutional Neural Networks

ABSTRACT

Wireless Capsule Endoscopy (WCE) is emerging as an important alternative to traditional diagnostic approaches for gastrointestinal diseases. The advantages in terms of comfort and completeness of the inspection encourage the development of innovative and efficient solutions to address two main issues: the time-consuming analysis of the significant amount of data produced and the tight power budget connected to the battery life. In this work, we adopt the edge-processing paradigm to evaluate a suitable Convolutional Neural Network for the recognition of relevant diseases directly on the capsule, reducing the data stream to frames where signs of relevant diseases are detected, thus improving the efficiency of the capsule while assisting medical professionals in the analysis. The WCE-SqueezeNet model was evaluated on WCE frames from the Kvasir and ETIS-Larib, LD Polyp Video, CVC-ColonDB and CVC-CLinicDB datasets, reaching respectively 98.88%, 90.2%, and 80.4% accuracy. Direct measurements of a parallel implementation on the GAP9 multi-core platform demonstrate classification at a 16 fps rate, with 61 ms inference time and 30.6 mW average core power consumption.

1. Introduction

Endoscopy and colonoscopy represent the standard diagnostic instruments for the recognition of gastrointestinal (GI) diseases, which have been estimated to affect over 332 million people in Europe in the period 2000-2019 [1]. The inspection of the GI tract enables the early detection of these conditions, allowing the prevention of serious complications, such as the onset of cancerous tumors. Wireless Capsule Endoscopy (WCE) represents a non-invasive recent alternative to traditional techniques, allowing inspection of portions of the small bowel that are especially hard to reach with traditional colonoscopy [2] and offering to reduce the discomfort and pain for the patient undergoing the examination.

The capsule device is swallowed by the patient and transits along the GI tract, continuously capturing images through an integrated camera. This data is typically transferred to an external server to be examined by expert physicians, with the assistance of automated disease recognition tools supported by Artificial Intelligence (AI) [3–5], reducing the time needed to screen the amount of data resulting from the procedure. Despite the computational power provided by an external server, where image processing can be performed with complex and accurate models, medical applications are raising a growing interest in the AI-at-the-edge trend [6–8]. In the context of WCE, enabling intelligence-at-the-edge represents a key innovation to optimize the energy efficiency of the capsule, which is battery-powered and needs to ensure operation throughout the transit time, expected to be between 8 and 11 h, and varying from subject to subject. This paradigm shift represents the first fundamental step towards advanced WCE devices, aiming at integrating active robotic control and supporting targeted investigation and controlled drug delivery [9,10].

Fig. 1 depicts the envisioned WCE system considered in this research. We align with the edge computational paradigm, considering the capsule device to integrate an AI-capable processing module, along with the camera and the transmission module. The video recorded by the camera results in a continuous stream of frames that are individually processed for GI disease detection directly on the capsule, so that only suspicious frames are sent to an external device through the Bluetooth Low-Energy (BLE) protocol, to be examined by medical professionals. Due to the form factor and the power budget, a processing device suitable for integration in the capsule introduces strict constraints on the complexity of a candidate AI model, both in terms of memory requirements and the number of required operations to ensure real-time processing: recent AI-oriented platforms in the edge domain leverage up to a few MBs of available memory [11], although the working memory of most common microcontroller-based platforms is within 1 MB.

☆ This article is part of a Special issue entitled: 'DASIP 2024/25' published in Journal of Systems Architecture.

* Corresponding author.
E-mail address: paola.busia@unica.it (P. Busia).

https://doi.org/10.1016/j.sysarc.2026.103851
Received 31 July 2025; Received in revised form 14 April 2026; Accepted 19 May 2026
Available online 25 May 2026

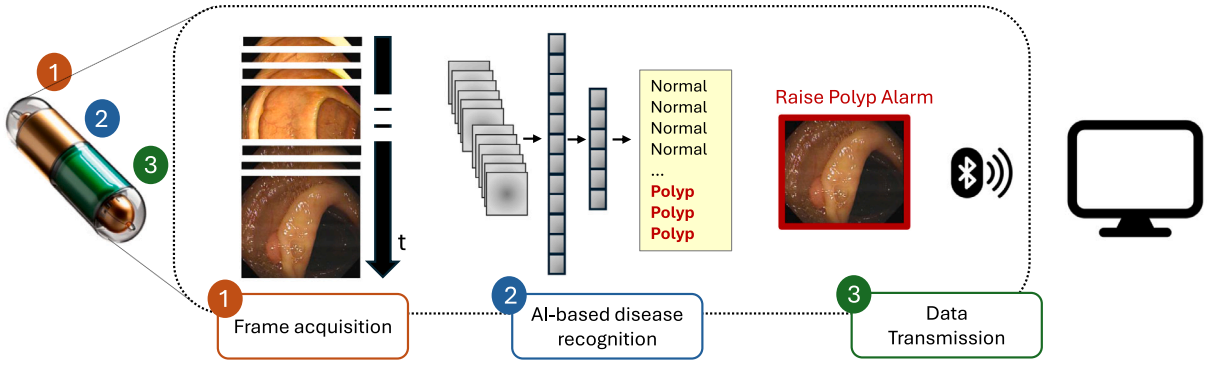

Fig. 1. Overview of the proposed WCE system: in-place AI processing enables smart frame transmission based on detected GI diseases requiring further inspection.

This work elaborates on the preliminary study published in [12], aiming at demonstrating the feasibility of enabling near-sensor image processing capabilities on the WCE device, with a lightweight Convolutional Neural Network (CNN) suitable for efficient inference execution on low-power microcontroller-based platforms, compatible with integration on the capsule.

In [12], we compared different network topologies and evaluated the most effective image resolution reduction representing the best trade-off between accuracy and computational complexity, resulting in the selection of the WCE-Squeezenet model. The classification performance was assessed on a small dataset for the recognition of four different GI conditions. The preliminary study also included the implementation of the model on the GAP9 platform [13], with direct measurements of power consumption, showing up to 16 fps processing throughput within a core power envelope of 30.6 mW, and demonstrating compatibility with the WCE task. This work further extends the study in [12], with the refinement of the model training on a significantly larger dataset and a thorough assessment of the classification performance on more diverse and unseen data, described in Sections 4.2 and 4.3. The new resources also include video acquisition, providing a benchmark closely resembling practical scenarios. Furthermore, we detail the analysis of the savings enabled by in-place processing, considering the time and energy required for data transmission with a suitable BLE module in different operating conditions. This analysis is summarized in Section 5.1.

The overall contributions of this research are listed in the following:

- the demonstration of the classification performance of the WCE-SqueezeNet model for real-time near-sensor WCE image classification, resulting from the assessment on three open-source datasets, reaching up to 98.88% accuracy in the recognition of three common GI conditions and 76% frame-level recall in polyp detection, with the detection of all examined polyp events;
- the estimation of up to 68% energy and bandwidth savings enabled by in-place processing, based on the online power profiler of the nRF52840 BLE module, selected as suitable for integration on the capsule device.

The paper is organized as follows: Section 2 summarizes the state of the art, Section 3 describes our proposed approach for the classification model training and evaluation, Section 4 reports the experimental results, Section 5 describes the on-hardware performance assessment and finally Section 6 summarizes the conclusions.

2. Related work

The reduced discomfort offered by WCE has encouraged the development of several commercial capsule devices approved in clinical practice [14–18]. Based on the analysis in [19], these devices typically acquire images of resolution varying from 256 × 256 to 512 × 512, at 2-6 fps, with a typical battery life ranging from 7 to 12 h.

A significant research effort has been made to address two of the most relevant issues of the approach: the need to reduce the examination time for the inspection of the massive number of images, up to 50 000, produced by each procedure, and to optimize power consumption to increase the battery life. Regarding the first problem, medical image processing benefits from the rapidly growing innovation offered by AI, as demonstrated by several works from the literature evaluating different solutions for GI condition recognition from endoscopy images. A common approach is to exploit state-of-the-art network models pre-trained on large datasets, e.g., ImageNet, and fine-tuned for WCE classification. The author of [3] presents a classification model, called MFuRE-CNN, for the recognition of 3 pathological conditions against healthy samples. The model results from the combination of truncated versions of the EfficientNetB0, MobileNetV2, and ResNet50V2 topologies, exploited as feature extractors prior to a Fusion Residual Block producing classification. The GastroNet model proposed in [4] leverages the EfficientNetV2B2 topology, adapted to the WCE task with the integration of a custom classification head combining Global Average Pooling and Dense layers. Similarly, the DCDS-Net model [5] exploited transfer-learning from the Xception model, integrating several blocks of separable convolutions prior to a classification block composed of three Dense layers. Additionally, state-of-the-art solutions for image segmentation have also been adapted to GI screening tasks, especially polyps detection and localization. Recent relevant examples are represented by the CE-YOLO model [7], inspired by YOLOv8, and the TPolyp model [20], based on the Swin transformer topology.

The systems considered in these works involve the continuous transmission of the acquired images to an external device having the necessary computational resources to run AI inference. This solution is already consolidated and applied in commercial devices, typically combined with a belt enabling storage for further processing. In this work, we consider a shift of paradigm, moving the processing to the capsule device. This approach represents an opportunity for improving the energy efficiency of the capsule, going from continuous transmission to the transmission of a subset of meaningful images. A preliminary pre-processing method based on RGB manipulation for the discard of non-significant images was evaluated in [21]. More recently, intelligent capsules integrating real-time image processing based on AI have been proposed [8], in combination with a CNN model for the recognition and detection of colorectal polyps designed considering the memory and efficiency constraints set by the capsule. Despite the reduced complexity, the assessment on the data acquired from 255 patients of Denmark's national screening program demonstrates the feasibility of running the intelligent task at the edge with high precision. An interesting effort towards AI integration in the camera pill is also presented in [22], describing the implementation of a lightweight CNN for polyp segmentation in Xilinx ZCU-102 FPGA, aiming to further miniaturization.

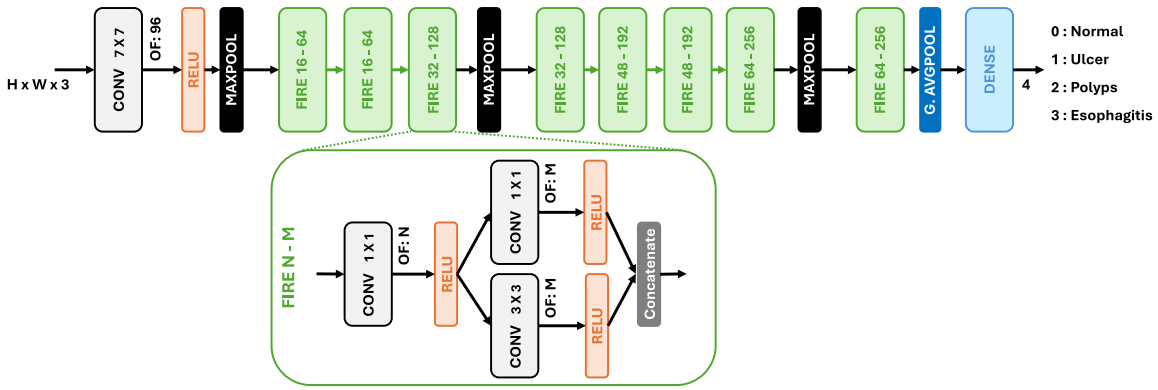

Fig. 2. WCE-SqueezeNet classification model.

Edge-processing thus represents a valid solution for extending the capsule battery-life, orthogonal to image compression approaches [19] and wireless power transfer [23]. Compared to the work of [8], we focus on the classification model assessment, further reducing the computational requirements with a model based on the SqueezeNet backbone. The classification performance of WCE-SqueezeNet is evaluated on diverse open-source datasets, and the suitability of real-time classification is demonstrated through direct measurements on a resource-constrained low-power hardware platform.

3. Method

This section describes the proposed AI-enhanced WCE system, presenting the details of the AI-capable hardware target, the candidate CNN classifier architecture, and the reference datasets considered for the assessment.

3.1. Hardware target

As the capsule needs to be swallowed, the size of the processing module must be compatible with the typical camera-pill dimensions of 11 mm × 26 mm. A suitable target for this constraint is represented by the GAP9 Parallel Ultra-Low-Power (PULP) platform [13], whose achievable energy efficiency on AI workloads, demonstrated as 0.033mW/GOP in the tiny-ML benchmarks [24], is especially promising to keep the power consumption within the budget established by the embedded battery. GAP9 is an advanced microcontroller-based platform, leveraging a cluster of nine parallel processors, that can be exploited for parallel processing and the acceleration of typical deep learning workloads, such as convolutional and fully connected layers. The memory hierarchy includes a shared 128 kB L1 memory and a 1.5 MB L2 memory, enough to effectively run lightweight edge-AI models. Based on these considerations, GAP9 was considered as the deployment target to validate the suitability of the proposed classification system for integration on a WCE capsule.

3.2. Network model

A relevant issue in AI development for medical applications is the size and the imbalance of typical medical datasets, making it difficult to meet the required accuracy standards by training a classifier from scratch. Due to these reasons, imaging tasks usually benefit from fine-tuning state-of-the-art models, having known feature extraction capabilities resulting from pre-training on large image datasets [25].

Based on this consideration, we searched for a suitable off-the-shelf model for the WCE image classification task, based on the computational and storage constraints resulting from the ultra-low-power tiny microcontrollers considered as hardware target. To meet the 1 MB memory footprint constraint, we selected as the backbone of our classification model the SqueezeNet topology [26], having less than 1 million parameters. The topology adapted for the WCE task is recalled in Fig. 2. Compared to the Vanilla topology, we replaced the classification head with a dense layer suitable for the new 4-class problem. In this configuration, the model accounts for 740k parameters; therefore, it is compatible with the constraint with 8-bit quantization.

For the fine-tuning of the model, we leveraged the Pytorch framework, exploiting CrossEntropy loss and SGD optimizer. The learning rate value, initially set to 0.001, was iteratively reduced after a 30-epoch patience since the last best validation loss. The early stop condition is met after 6 updates. The training was performed on Google Colaboratory, leveraging the T4 GPU.

3.3. Reference datasets

The results reported in this study were obtained considering multiple open-source WCE data resources. The topology exploration and optimization first leveraged the KVASIR [27] and the ETIS-Larib Polyp [28] databases. The data organization introduced in [3] includes 6000 images acquired through WCE, representing examples of three main pathological conditions of the gastrointestinal tract, such as ulcerative colitis, polyps, and esophagitis, as well as healthy/normal samples. Ref. [3] also introduced a standard training, validation, and test split, according to the scheme summarized in the first section of Table 1. As can be noticed, the size of the dataset and the balance in the representation of all four classes make it especially suitable for a topology exploration phase, where training efficiency is paramount.

To further assess the classification performance of the model, particularly on the crucial polyp recognition problem, we additionally considered the LDPolyp Video BenchMark [29], the CVC-ColonDB [30], and the CVC-ClinicDB [31]. The LDPolyp Video BenchMark dataset collects 33 834 annotated polyp images, extracted from 160 colonoscopy videos showing 200 different polyp examples. The frames are arranged into a train/validation and a test set, and they are accompanied by the information about the position of the polyps in the images. Alongside these selected frames, the dataset also provides 61 videos showing tissue without polyps, and 42 non-annotated videos with polyps. For the training refinement, we considered a subset of the annotated training frames, including normal and polyp examples, combined with the normal frames extracted from 3 videos without polyps, at a 5 fps rate. The test folder and the remaining normal videos were considered for the performance assessment, addressing polyp recognition against the other conditions learned during training.

Similarly, the CVC-ClinicDB and CVC-ColonDB collect, respectively, 612 and 380 frames showing polyp occurrences from 44 different videos. These frames were considered as an additional test set for the generalization assessment. The detail of the considered datasets composition is summarized in Table 1.

Table 1
Composition and organization of the considered WCE datasets.

Dataset ID	Dataset Name	Image Resolution	Class	Train	Valid	Test
1	Kvasir, ETIS-Larib	720 × 576	Normal - N	800	500	200
			Ulcer - U	800	500	200
			Polyps - P	800	500	200
			Esophagitis - E	800	500	200
			Tot	**3200**	**2000**	**800**
2	LDPolyp Video	560 × 480	Normal - N	3035	758	38 586
			Polyps - P	749	187	12 933
			Tot	**3784**	**945**	**51 519**
3	CVC-ClincDB, CVC-ColonDB	384 × 288, 574 × 500	Normal - N	–	–	0
			Polyps - P	–	–	992
			Tot	-	-	**992**

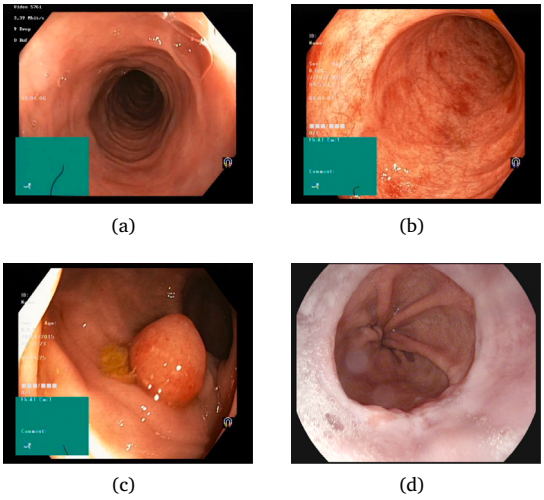

Fig. 3. Sample images from the reference Dataset 1, including (a) normal sample, (b) ulcerative colitis, (c) polyp, (d) esophagitis.

An example of images resulting from WCE acquisition is reported in Fig. 3, showing samples from each of the targeted classes in datasets [27,28]. All the images in the dataset were pre-processed, in order to standardize their size to a common resolution, and normalized to align with ImageNet pre-training. The selection of the input resolution was then the subject of a dedicated exploration, which is described in detail in Section 4.

4. Experimental results

This section includes the description of the preliminary assessment of the proposed classification model, demonstrating its suitability for in-place WCE image classification. Additionally, we explore in more depth the achievable performance on larger and unseen datasets by leveraging data augmentation and considering more diverse training and test examples. The trained models and the classification performance evaluation code are available open source.[1]

4.1. Validation of the proposed classification systems

The WCE-Squeezenet model described in Section 3 was preliminarily evaluated on Dataset 1. The classification performance obtained on the test set is reported in the first line of Table 2, which refers to the ImageNet pretrained model, finetuned for WCE classification, applied to images of 224 × 224 resolution. As can be observed, the different GI conditions are recognized with precision and recall over 95%, demonstrating a good discrimination capability. This solution was compared to the same topology trained from scratch, and to a hybrid model obtained from the combination of the pre-trained SqueezeNet feature extractor and a Support Vector Machine (SVM) trained on the resulting features. The comparison with the previous solution shows a drop in the achievable classification accuracy when ignoring the general knowledge acquired on ImageNet and relying only on the WCE dataset for the learning of the new task. Additionally, CNN standalone classification demonstrates higher performance than hybrid alternatives combining the fixed feature extractor with a trainable classifier.

To provide a comparison with a slightly more complex alternative, we also considered the MobileNetV2 [32] model as the feature-extracting backbone. This solution results in a network model with little more than 2M parameters, thus exceeding, not by far, the target memory constraint, yet providing a reliable upper bound for the expected classification performance. We performed a symmetrical evaluation, considering the pretrained CNN feature extractor, finetuned and combined with a dense classification head, or fixed and combined with a trained SVM classifier. The result of the comparison is consistent with the findings of the previous exploration, confirming the advantages of the standalone CNN solution. As can be noticed, the classification performance is only slightly improved compared to the more lightweight WCE-SqueezeNet model.

For further optimization, after a suitable topology definition, we evaluated the impact of different image resolutions on the classification outcome. We considered a sweep of the image size from the original 224 × 224 resolution, exploited in [3], to progressively compressed representations, such as 128 × 128, and 64 × 64. Fig. 4 summarizes the outcome of the exploration. Each candidate design point was trained and tested on images having one of the target resolutions, and it is placed in the plot based on the classification accuracy and the complexity in terms of the required number of operations (GOPs) per inference. As can be observed, a compression of the resolution to a 128 × 128 introduces only a negligible drop in the accuracy, while resulting in a significant reduction, by a factor of 3×, of the computational workload. On the contrary, the performance degrades significantly when the resolution is reduced further. Based on this first assessment, and considering the low-power hardware targeted for the deployment, we selected the model trained to perform classification on the 128 × 128 images.

4.2. Classification performance assessment

The topology in Fig. 2 reached up to 99% accuracy on the test set selected from Dataset 1, with full-precision 32-bit floating-point inference. This result was further improved thanks to training data augmentation, exploiting random rotations, width and height shifts, zoom, and horizontal flips. The augmented dataset included up to 4

[1] https://github.com/paobusia/WCE-SqueezeNet

Table 2
Comparison of the achievable classification performance Assessment on Dataset 1. PT column indicates whether the training started from ImageNet-trained weights.

Model	PT	Accuracy	Ulcer Precision	Ulcer Recall	Polyps Precision	Polyps Recall	Esophagitis Precision	Esophagitis Recall
WCE-SqueezeNet	✓	**98.13**	98.45	95	95.12	97.5	100	100
WCE-SqueezeNet	✗	84.63	64.84	88.5	81.97	50	98	100
SqueezeNet + SVM	✓	96.5	95	95.5	95.29	91	100	100
WCE-MobileNetV2	✓	**98.5**	97	97	97	97	100	100
MobileNetV2 + SVM	✓	96.87	96.39	93.5	94	94	100	100

Table 3
Test classification performance of WCE-Squeezenet model on Dataset 1.

(a) With data augmentation

True Labels \ Predicted Labels	N	U	P	E
Normal	**200**	0	0	0
Ulcer	0	**200**	2	0
Polyps	1	2	**197**	0
Esophagitis	0	0	0	**200**

(b) After refinement on Dataset 2

True Labels \ Predicted Labels	N	U	P	E
Normal	**200**	0	0	0
Ulcer	0	**199**	1	0
Polpys	3	4	**193**	0
Esophagitis	0	0	1	**199**

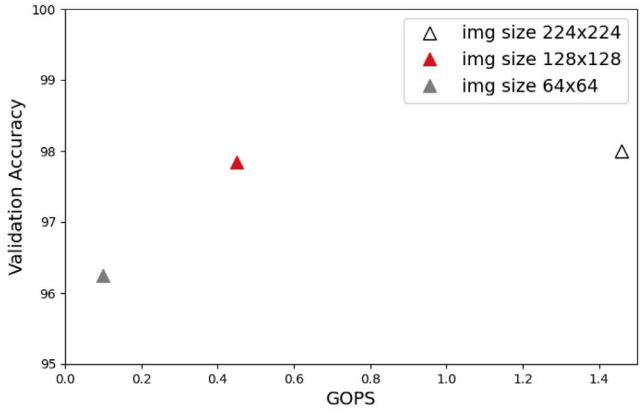

Fig. 4. Exploration of the input image resolution considering the validation set.

copies of each training instance, resulting in over 3200 examples per class. The network trained with augmentation reached 99.6% accuracy, with perfect recall on normal, ulcer, and esophagitis recognition, and only 3 errors in polyps classification. The detailed confusion matrix on the test set of Dataset 1 is reported in Table 3a.

Despite these encouraging results, demonstrating the possibility of exploiting the proposed classification system for GI disease recognition during WCE procedures, the size of the dataset considered for the performance assessment is far from adequately representing the complexity and variability of the data resulting from real-time acquisition in practical scenarios. To provide a more thorough evaluation of the expected classification performance, we focused on the crucial task of polyp recognition and considered Dataset 2 as an additional resource. A first run of the refined model after data augmentation on six normal videos, with frames extracted at a 5 fps rate, showed an accuracy drop to 38%, with over 60% false alarms, suggesting the need for accessing a larger and more diverse learning database for the training of the model.

We thus repeated the training process, including the training set from Dataset 2 and the augmented training set from Dataset 1. The resulting model demonstrated a slight drop in Dataset 1 accuracy, to 98.88%, but with a significant improvement to 99% accuracy on the previously considered six normal videos. The comprehensive summary of the classification performance on Dataset 1 is reported in the confusion matrix in Table 3b, whereas the results on the Dataset 2 test set are reported in Table 4a. These results demonstrate an accuracy of 89.3% in the recognition of polyps against normal tissue, with 84.9% precision and 76.2% recall. The specificity, or recall of the normal class, is 93.7%, showing an adequate discrimination capability, especially relevant as no frame selection was performed based on the acquisition from the video: the test set included several extremely noisy images, such as the ones reported in Fig. 5. For reproducibility purposes, the IDs of the videos from which the images were acquired are reported in the caption. The composition of the considered Dataset should provide a realistic scenario of the quality of the images that would be acquired and processed in a practical scenario, affected by different sources of noise, such as the movement of the camera. The sampling frequency selected for image capture emulates a 5 fps acquisition, which is competitive with the 1 fps and 2 fps acquisition frequencies envisioned in [8].

The video and sequential data in Dataset 2 provide another interesting opportunity with a considerable relevance for practical scenarios, i.e., exploiting the temporal information to refine the classification outcome. We applied a post-processing based on the majority voting of 5 consecutive frames, thus considering the 4 previous classifications to confirm the current one. Starting from the raw results in Table 4, the majority voting enables the correction of 353 errors in the classification of normal frames, and 78 in the recognition of polyps. With this refinement, the classification performance improves to 90.2% accuracy, with 86.99% precision and 76.8% recall in polyp recognition, and 94.7% specificity. Finally, the event-level analysis of the polyp detection results shows that all of the 60 polyp sequences in the test dataset are successfully detected to some extent, despite single polyp frames being occasionally misclassified. The detail of the detection performance is reported in Fig. 6, showing how most of the test polyps are detected with confidence, with some exceptions represented especially by test sequences 109, 148, and 150. A very interesting direction for further analysis would be represented by consultation with expert physicians on the specific features of the polyps shown in the sequences that are not classified correctly, that could provide relevant insights to improve the explainability of the system. The analysis of the classification output shows the ulcer alarm is raised for most of the frames included in sequences 109 and 148, whereas most of the images in sequence 149 were classified as normal by the model.

4.3. Generalization assessment

To assess the generalization capability of the proposed classification system, we extended the evaluation to an additional dataset, completely

Table 4
Test classification performance of WCE-Squeezenet model on Datasets 2 and 3.

(a) Test on Dataset 2

True Labels \ Predicted Labels	N	U	P	E
Normal	**36169**	523	1759	135
Ulcer	0	**0**	0	0
Polyps	1508	1521	**9859**	45
Esophagitis	0	0	0	**0**

(b) Test on Dataset 3.

True Labels \ Predicted Labels	N	U	P	E
Normal	**0**	0	0	0
Ulcer	0	**0**	0	0
Polyps	153	16	**798**	25
Esophagitis	0	0	0	**0**

Table 5
Comparison with the state of the art of WCE image classification models.

Model	Dataset	Accuracy	Precision	Recall	Edge Target	Parameters Memory	GOPS[a]
MFuRE-CNN [3]	1	97.75%	97.75%	97.75%	✗	19.2 MB	7.8
ResNet50 [4]	1	98%	98.1%	98%	✗	89 MB[b]	7.6[b]
EfficientNetV2B1 [4]	1	98.5%	98.5%	98.5%	✗	25.9 MB[b]	2[b]
GastroNetV1 [4]	1	99.2%	99.3%	99.3%	✗	32.8 MB[b]	2.3[b]
DCDS-Net [5]	1	99.33%	99.37%	99.32%	✗	83.8 MB	9[b]
this work 32-bit	**1**	**98.88%**	**98.88%**	**98.88%**	✓	**750 kB**	**0.45**
this work 8-bit		**98.76%**	**98.76%**	**98.75%**			
YOLO-based [8]	custom	99.5 mAP	N.R.	N.R.	✓	3.2 MB	0.9[b]
CNN [22]	custom	80 mAP	N.R.	N.R.	✓	157 kB[b]	0.8[b]
TPolyp [20]	2	99.5 mAP	99%	98.9%	✗	138 MB[b]	62[b]
CE-YOLO-Tiny [7]	2	N.R.	72.99%	48.07%	✓	3.41 MB	2.1
this work 32-bit	**2**	**90.2%**	**86.99%**	**76.8%**	✓	**750 kB**	**0.45**
this work 8-bit		**88.67%**	**86.58%**	**73.4%**			
CE-YOLO-Tiny [7]	3[c]	N.R.	88.97%	83.89%	✓	3.41 MB	2.1
this work 32-bit	**3**[c]	**87.4%**	N.A.	**87.4%**	✓	**750 kB**	**0.45**
this work 8-bit		**85.62%**		**85.62%**			

[a] The number of operations refers separately to multiplications and additions: 1 MAC = 2 OPS.
[b] Estimated from paper.
[c] Limited to the CVC-ClinicDB.

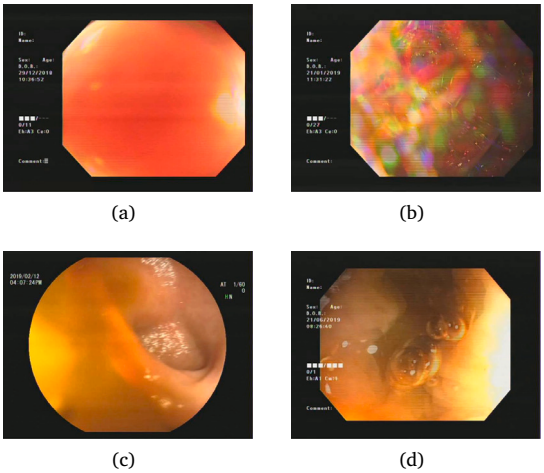

Fig. 5. Example of noisy images resulting from frame acquisition from normal videos in Dataset 2. The reported images were extracted from the videos identified with the IDs (a) 0_4, (b) 0_18, (c) 0_33, (d) 0_61.

unseen during the training of the model, Dataset 3. All the images in the dataset include one or more polyps, which the model should detect without the need for any additional training. The inference confusion matrix is reported in Table 4b, showing an accuracy of 80.4%, corresponding to the recall of polyps recognition. The proposed model is thus adequate to recognize polyp occurrence in various conditions, by processing frames resulting from different acquisition systems.

4.4. Discussion

Table 5 summarizes recent works from the state of the art, addressing WCE-image classification. The table indicates in the *Edge Target* column whether edge deployment was considered as target in the definition of the model, thus introducing complexity and footprint constraints impacting on the achievable classification and detection performance.

The first section of the Table refers to CNN-based classification models, evaluated on Dataset 1 for 4-class classification. Despite reaching accuracy values as high as 99.3%, the models in [3–5] require at least 20 MB storage resources, and over 2 GOPS per inference run. Due to this reason, they do not represent suitable candidates for integration in an intelligent WCE device. Nonetheless, WCE-SqueezeNet reaches the third-best accuracy, less than 1% lower than the absolute best, while significantly reducing the storage and processing requirements.

As a reference for analyzing the performance obtained on Datasets 2 and 3, we can consider some recent works from the literature where polyps recognition is addressed as a segmentation or detection task [7, 20]. This problem is slightly more complex than classification, requiring the model to correctly locate the polyp occurrence within the frame, whereas classification only evaluates whether a polyp appears in any position in the frame. However, precision and recall are considered to be the most relevant metrics for the evaluation of polyps detection [7], even more than the mean average precision (mAP). We thus focus on these metrics for the comparison, with two main differences compared to the segmentation works we reference: (1) we consider a single bounding box as big as the frame, and (2) we do not distinguish any two polyps present in the same frame.

The transformer model presented in [20] reaches a remarkable 99% precision and recall, ensuring extremely precise detection with

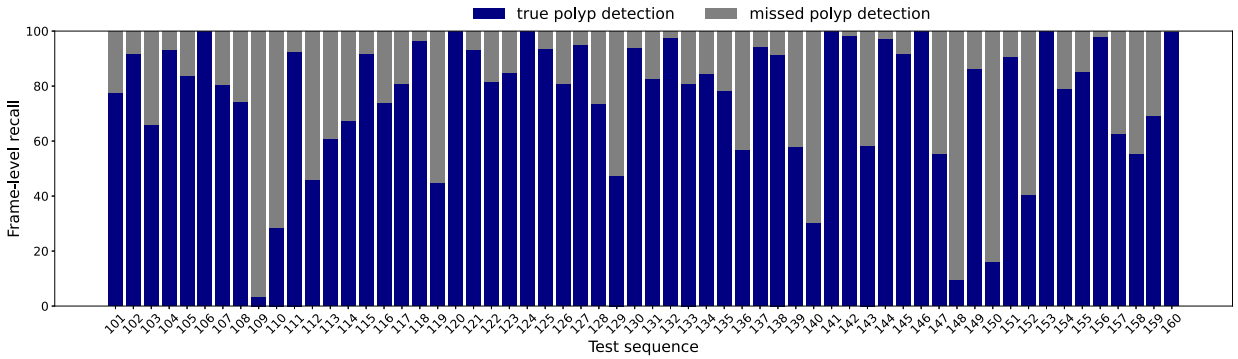

Fig. 6. Detail of the polyp detection recall on the test sequences from Dataset 2.

a complex architecture inspired by the Swin model, which is incompatible with real-time frame processing on edge devices. It can thus represent a valuable solution only when continuous frame transmission to the server is enabled, but it cannot be run on the WCE capsule. However, despite a reduced precision, our solution enables the event-level detection of all the polyp occurrences in Dataset 2, with a 5.3% false alarm rate resulting in additional transmissions, yet enabling 68% transmission energy savings, as detailed in Section 5.1.

Most of the works discussed so far focus on classification performance assessment, without discussing the required inference time and any implications on the constraints of the WCE capsule. The rest of the Section discusses works targeting edge deployment. The assessment on Dataset 2 of the YOLO-based solution proposed in [7] is limited to only 600 polyp frames selected to ensure a diverse representation of the different occurrences in the dataset, whereas the evaluation of WCE-SqueezeNet considered over 50 000 different frames. Despite solving a slightly less complex problem, our model provides a competitive solution, with significantly higher precision and recall levels. Similarly, the comparison on Dataset 3 demonstrates an adequate recognition accuracy. Furthermore, despite considering edge inference as a target and exploiting mixed-precision quantization to reduce the complexity of the model, CE-YOLO-Tiny requires over 4× the memory resources and number of operations. These metrics are provided by the authors as general hardware-independent complexity figures, whereas power measurements are not reported.

Finally, the YOLO-based solution in [8] represents a deployable alternative, already implemented on the custom camera-pill device proposed by the authors. The proposed device integrates a K210 Kendryte chip, specialized for AI workload execution and integrating 8 MB of RAM, a dual-core 64-bit CPU and a dedicated neural network processor. The proposed system processes images of size 240 × 240. Quantization to 8-bit precision was considered to fit the target device. The power consumption of the AI chip was measured at 40 MHz clock frequency, reporting an average power consumption of 50 mW. The required inference time under the selected frequency scaling is not reported. Nonetheless, the number of parameters is still significantly high, over 3 million, compared to WCE-SqueezeNet. On the contrary, the model proposed in [22] exploits a limited number of parameters, resulting in memory requirements well within 160 kB. The assessment of polyp detection was completed on the data from the Oden University Hospital, thus, the reported precision cannot be compared directly with the results of the evaluation of our WCE-SqueezeNet. The authors target edge deployment on the Xilinx ZCU-102 FPGA, but the peak power consumption estimated for the system is around 0.7 W, introducing significant requirements on the battery choice. The size and complexity of the model are compatible with deployment on different targets, such as GAP9, although further optimization on the image size, and consequently the network topology, would be needed to match the required execution time of WCE-SqueezeNet.

5. Deployment

To enable efficient deployment on the target platform, the WCE-SqueezeNet model was quantized to 8-bit precision. The results of the classification performance assessment on the integer version of the model are reported in Table 6. As can be observed, the classification performance is only marginally impacted, with a classification accuracy of 98.75% on Dataset 1, 87.6% on Dataset 2, and 79.7% on Dataset 3. The most significant drop, of approximately 2%, is observed on Dataset 2. These results are included in Table 5, considering the benefits of majority-based post-processing also on the quantized model. The quantization was performed with the TensorflowLite utilities, allowing to meet the 1 MB constraint set for the memory requirements. We exploited the GAP9 SDK proprietary code generation tool to obtain an efficient implementation of WCE-SqueezeNet on GAP9, mapping the computational load to the eight cores of the computing clusters. We considered 224 × 224 and 128 × 128 input resolution and evaluated the required inference time in the two scenarios. At 370 MHz working frequency, inference execution requires 200 ms in the first case, providing an expected throughput of 5 fps with an average computational efficiency of 9 OPS/cycle. In the second case, inference time is reduced to only 61 ms, reaching an expected throughput of 16 fps thanks to an average computational efficiency of 10 OPS/cycle.

Fig. 7 and 8 show in the top bar the distribution of the computational workload across the network layers, estimated in terms of the number of required operations. As can be observed, convolutional layers represent the most relevant load, as they are the building blocks of the typical Fire modules. The image resolution choice, although impacting the overall execution time, does not influence the workload composition. A symmetric analysis about the composition of the inference time is reported in the bottom bar. The graphs closely resemble the expected distribution based on OPS, demonstrating there is no significant inefficiency in the implementation of the most relevant operands.

To assess the energy efficiency, we measured the average core power consumption during inference execution, exploiting the Nordic Power Profiler II. The direct measurement showed a 30.6 mW average core power, resulting in a 1.9 mJ inference energy contribution of the model working on 128 × 128 frames. This result demonstrates that real-time GI disease recognition on the WCE device is feasible with WCE-SqueezeNet, within a limited power budget compatible with battery-powered solutions.

5.1. Bandwidth and energy savings

In the following, we assess the bandwidth and energy savings enabled by in-place AI-based processing on the capsule. We consider three different practical scenarios:

- **case 1 - continuous transmission**: each newly acquired frame is continuously streamed to an external device for offline processing on a dedicated server;

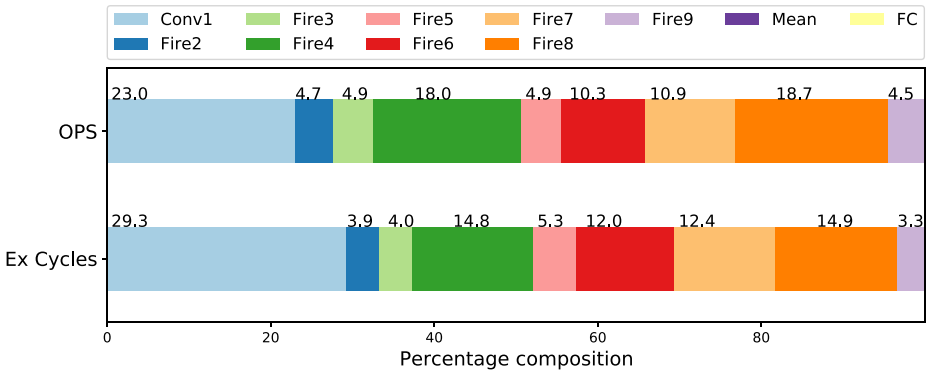

Fig. 7. Computational workload of the different layers in WCE-SqueezeNet model evaluated in terms of number of required operations and required execution cycles on the GAP9 platform, for 224 × 224 input resolution.

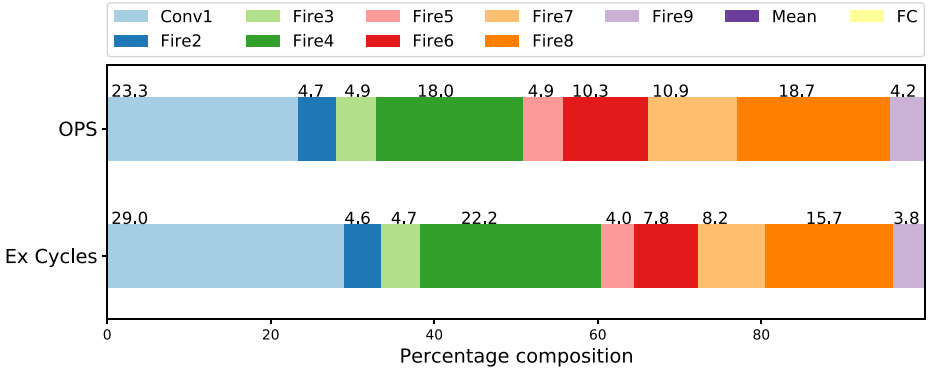

Fig. 8. Computational workload of the different layers in WCE-SqueezeNet model evaluated in terms of number of required operations and required execution cycles on the GAP9 platform, for 128 × 128 input resolution.

Table 6
Test classification performance of 8-bit WCE-Squeezenet model on Datasets 1, 2, and 3.

(a) Test on Dataset 1.

True Labels \ Predicted Labels	N	U	P	E
Normal	**200**	0	0	0
Ulcer	0	**199**	1	0
Polyps	3	5	**192**	0
Esophagitis	0	0	1	**199**

(b) Test on Dataset 2.

True Labels \ Predicted Labels	N	U	P	E
Normal	**35707**	932	1759	188
Ulcer	0	**0**	0	0
Polyps	1329	2113	**9413**	78
Esophagitis	0	0	0	**0**

(c) Test on Dataset 3.

True Labels \ Predicted Labels	N	U	P	E
Normal	**0**	0	0	0
Ulcer	0	**0**	0	0
Polyps	132	40	**791**	29
Esophagitis	0	0	0	**0**

- **case 2 - anomaly transmission**: the frames are processed in real-time on the capsule device and are only transmitted when a pathological condition is recognized;
- **case 3 - anomaly history transmission**: the frames are processed in real-time on the capsule device, and when a pathological condition is recognized, the image is transmitted to an external device along with the history of n previous acquisitions.

To estimate the energy consumption associated with data transmission, we considered the BLE protocol, particularly the Nordic Semiconductor nRF52840 module [33] integrated in the camera-pill described in [8]. The definition of these practical scenarios accounts for the operating modes considered in [8]: the authors propose a low-rate continuous transmission (case 1 with 0.2 fps) as long as no anomalies are detected, whereas, as soon as a pathological condition alarm is raised, acquisition frequency is increased to 2 fps, with continuous transmission towards the external server. To determine the required transmission time and energy in the three scenarios, we exploited the online power profiler [34], which reports a typical voltage supply of 3 V and current draw of 6.7 mA during 2Mbps data transmission in peripheral connection mode. Therefore, the typical transmission power P_{TX} is of 20.1 mW.

Table 7
Estimation of required transmission time and energy with the nRF52840 module for the three practical scenarios considered at different acquisition rates.

	1 fps			2 fps			5 fps		
	case 1	case 2	case 3	case 1	case 2	case 3	case 1	case 2	case 3
Byte TX	1.3 GB	128 MB	420 MB	2.6 GB	257 MB	838 MB	6.6 GB	643 MB	2.04 GB
Time TX	1 h 38 min	9 min 20 s	30 min 32 s	3 h 16 min	18 min 41 s	1 h 54 s	8 h 10 min	46 min 43 s	2 h 31 min
Energy TX	118.5 J	11.3 J	36.8 J	236 J	22.5 J	73.5 J	592.3 J	56.4 J	183.3 J
Savings TX	–	90%	68%	–	90%	68%	–	90%	68%
Savings tot	–	44%	22%	–	44%	22%	–	44%	22%

To estimate the number of images transmitted in each scenario we considered Eq. (1), which is based on a few assumptions, listed in the following: (1) we consider a typical WCE transit time of 8 hours [35]; (2) we focus on polyp detection and consider a number N_{polyp} of 10 occurrences per examined patient, which is regarded as a threshold to identify patients with numerous polyps [36]; (3) we estimate the capsule to transit for approximately 2 min, T_{polyp}, around each polyp. Based on these assumptions, we are able to estimate the total number of images to be transmitted in case 1 starting from the transit time in seconds, T_{wce}, and the selected acquisition rate acq_{rate}. For cases 2 and 3, we consider a first contribution accounting for the frames including true polyp occurrences, and one accounting for false alarm detections. This last contribution considers 5.3% of the acquired images as false alarms (FP_{rate}), based on the specificity demonstrated by the WCE-SqueezeNet model. In case 3, we also need to account for the overhead of log transmission, which is negligible for true detections, while impacting significantly on the contribution of false positives: we did not make any assumption on the order of occurrence of false alarms; therefore, we considered the transmission of n frames for every detection, with n selected equal to 5. This condition represents a worst-case scenario, as isolated false alarms are removed by the majority post-processing, thus it is unlikely that each false positive requires a 5-frame overload.

To estimate the overall data transfer, we exploited Eq. (2), considering an image size S_{img} of 128 × 128 × 3 B. In order to optimize the efficiency of the transmission, we consider the maximum payload size $S_{payload}$ of 251 B, resulting in a payload transmission time $T_{payload}$ of 1044 µs [34]. We are thus able to estimate the number of packages to be transmitted, based on the number of images N_{img} and the image size. Each package introduces an energy contribution $E_{payload}$ of 21 μJ, thus the overall energy consumption connected to transmission can be derived with Eq. (3).

$$N_{img} = \begin{cases} T_{wce} \times \mathrm{acq}_{rate}, & \text{in case 1} \\ N_{polyp} \times T_{polyp} \times \mathrm{acq}_{rate} + \mathrm{FP}_{rate} \times T_{wce} \times \mathrm{acq}_{rate}, & \text{in case 2} \\ N_{polyp} \times T_{polyp} \times \mathrm{acq}_{rate} + \mathrm{FP}_{rate} \times T_{wce} \times \mathrm{acq}_{rate} \times n, & \text{in case 3} \end{cases} \tag{1}$$

$$B_{TX} = N_{img} \times S_{img} \tag{2}$$

$$E_{TX} = N_{img} \times (S_{img}/S_{payload}) \times E_{payload} \tag{3}$$

The results of this analysis are reported in Table 7, detailing the number of bytes transmitted, as well as the required transmission time, and transmission energy for the complete transit time in the three scenarios and considering different acquisition rates. As can be observed, in-place processing enables over 68% bandwidth and transmission energy savings in case 3, when a set of images around the detection is transmitted for further analysis, and up to over 90% when only frames associated with positive detections are streamed. Considering an additional contribution of 1.9 mJ for every processed image, we can also estimate the energy consumption of AI inference, which accounts for 54 J, 109 J, or 274 J, based on the acquisition rate. The last row of the Table reports the overall savings when the AI contribution is included in cases 2 and 3, up to respectively 44% and 22%, neglecting any processing contribution that might be needed to assist continuous transmission in case 1. This analysis shows that energy savings are possible up to 60 true polyp occurrences per patient. Compared to the dual-mode operation described in [8], limiting transmission to detected anomalies improves the energy efficiency of the capsule significantly. Nonetheless, a possible advantage of the low-rate continuous transmission envisioned in [8] lies in the increased trustworthiness of the system, which can represent a relevant concern until the AI-enhanced processing can provide sufficient reliability and explainability.

6. Conclusions

In this work, we evaluated the feasibility of real-time GI disease recognition on WCE capsules with WCE-SqueezeNet. The classification performance was evaluated on three open-source datasets, demonstrating up to 98.88% accuracy in the recognition of normal tissue against ulcerative colitis, polyps, and esophagitis, and up to 86% precision and 76% recall in polyp detection. The inference time and average power measurements showed the suitability of real-time recognition on low-power AI-processing platforms such as GAP9, within a core power envelope compatible with possible integration in a WCE device. Additionally, the analysis of the achievable savings in bandwidth and energy consumption shows the advantages of integrating intelligence on board. As a future development, we consider that further optimization of the classifier model will enable even greater savings and reduce the computational requirements, thereby increasing the accessibility of a similar device.

CRediT authorship contribution statement

Paola Busia: Writing – original draft, Validation, Software, Methodology, Investigation. **Andrea Pinna:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Paolo Meloni:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - Call for tender No. 3277 published on December 30, 2021 by the Italian Ministry of University and Research (MUR) funded by the European Union – NextGenerationEU. Project Code ECS0000038 – Project Title eINS Ecosystem of Innovation for Next Generation Sardinia – CUP F53C22000430001- Grant Assignment Decree No. 1056 adopted on June 23, 2022 by the Italian Ministry of University and Research (MUR). This research was also supported by NextGenerationEU Mission 4, Component 2, Investment 1.5, CUP B83C22002820006—Project

METBIOTEL—Innovation Ecosystem ECS 0000024 ROME TECHNO POLE SPOKE 1, and SPOKE 6, by the French National Research Agency (ANR) under the LabCom program 2021 - V2 (ICI-Lab), and by the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement GA 101140052 (H2TRAIN).

References

[1] Tackling the burden of digestive disorders in Europe, Lancet Gastroenterol. Hepatol. 8 (2023) 95, http://dx.doi.org/10.1016/S2468-1253(22)00431-9.

[2] M. Bordbar, M.S. Helfroush, H. Danyali, F. Ejtehadi, Wireless capsule endoscopy multiclass classification using three-dimensional deep convolutional neural network model, BioMedical Eng. OnLine 22 (2023) 124, http://dx.doi.org/10.1186/s12938-023-01186-9.

[3] F.J.P. Montalbo, Diagnosing gastrointestinal diseases from endoscopy images through a multi-fused CNN with auxiliary layers, alpha dropouts, and a fusion residual block, Biomed. Signal Process. Control. 76 (2022) 103683, http://dx.doi.org/10.1016/j.bspc.2022.103683, URL https://www.sciencedirect.com/science/article/pii/S1746809422002051.

[4] S. Rajkumar, C.S. Harini, J. Giri, V.A. Sairam, N. Ahmad, A.S. Badawy, G.K. Krithika, P. Dhanusha, G.E. Chandrasekar, V. Sapthagirivasan, GastroNet: A CNN based system for detection of abnormalities in gastrointestinal tract from wireless capsule endoscopy images, AIP Adv. 14 (8) (2024) 085223, http://dx.doi.org/10.1063/5.0208691.

[5] S. Asif, M. Zhao, F. Tang, Y. Zhu, DCDS-net: Deep transfer network based on depth-wise separable convolution with residual connection for diagnosing gastrointestinal diseases, Biomed. Signal Process. Control. 90 (2024) 105866, http://dx.doi.org/10.1016/j.bspc.2023.105866, URL https://www.sciencedirect.com/science/article/pii/S1746809423012995.

[6] P. Busia, M.A. Scrugli, V.J.-B. Jung, L. Benini, P. Meloni, A tiny transformer for low-power arrhythmia classification on microcontrollers, IEEE Trans. Biomed. Circuits Syst. (2024) 1–11, http://dx.doi.org/10.1109/TBCAS.2024.3401858.

[7] C. Xu, Z. Fan, Y. Ma, Y. Huang, J. Wang, Y. Xu, D. Wu, CE-YOLO: A channel-efficient YOLO algorithm for colorectal polyp detection, Biomed. Signal Process. Control. 104 (2025) 107529, http://dx.doi.org/10.1016/j.bspc.2025.107529, URL https://www.sciencedirect.com/science/article/pii/S1746809425000400.

[8] A. Sahafi, Y. Wang, C.L.M. Rasmussen, P. Bollen, G. Baatrup, V. Blanes-Vidal, J. Herp, E.S. Nadimi, Edge artificial intelligence wireless video capsule endoscopy, Sci. Rep. 12 (2022) 13723, http://dx.doi.org/10.1038/s41598-022-17502-7.

[9] Q. Cao, R. Deng, Y. Pan, R. Liu, Y. Chen, G. Gong, J. Zou, H. Yang, D. Han, Robotic wireless capsule endoscopy: recent advances and upcoming technologies, Nat. Commun. 15 (2024) 4597, http://dx.doi.org/10.1038/s41467-024-49019-0.

[10] C.-C. Su, C.-K. Chou, A. Mukundan, R. Karmakar, B.F. Sanbatcha, C.-W. Huang, W.-C. Weng, H.-C. Wang, Capsule endoscopy: Current trends, technological advancements, and future perspectives in gastrointestinal diagnostics, Bioengineering 12 (6) (2025) http://dx.doi.org/10.3390/bioengineering12060613, URL https://www.mdpi.com/2306-5354/12/6/613.

[11] M. Scherer, L. Macan, V.J.B. Jung, P. Wiese, L. Bompani, A. Burrello, F. Conti, L. Benini, Deeploy: Enabling energy-efficient deployment of small language models on heterogeneous microcontrollers, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 43 (11) (2024) 4009–4020, http://dx.doi.org/10.1109/TCAD.2024.3443718.

[12] P. Busia, A. Pinna, P. Meloni, Endoscopy image classification for wireless capsules with CNNs on microcontroller-based platforms, in: J. Lorandel, A. Kamaleldin (Eds.), Design and Architecture for Signal and Image Processing, Springer Nature Switzerland, Cham, 2025, pp. 57–68.

[13] Greenwaves, Ultra low power GAP processors, 2024, URL https://greenwaves-technologies.com/low-power-processor/.

[14] Medtronic, PillCam™ SB 3 capsule endoscopy system, 2025, URL https://www.medtronic.com/en-us/healthcare-professionals/products/digestive-gastrointestinal/capsule-endoscopy/endoscopy-systems/pillcam-sb-3-capsule-endoscopy-system.html.

[15] Olympus, 65 years of GI endoscopy innovation in one powerful little capsule, 2025, URL https://medical.olympusamerica.com/products/endocapsule.

[16] IntroMedic, Digestive endoscopy capsule MiroCam, 2025, URL https://www.medicalexpo.it/prod/intromedic/product-77770-469488.html.

[17] Jinshan, OMOM HD capsule endoscopy platform, 2025, URL https://www.jinshangroup.com/product/omom-hd-capsule-endoscopy-camera/.

[18] Anx Robotics, Introducing NaviCam XS – Smaller capsule, bigger impact, 2025, URL https://www.anxrobotics.com/products/navicam-sb-capsule-system/.

[19] B. Sushma, P. Aparna, Recent developments in wireless capsule endoscopy imaging: Compression and summarization techniques, Comput. Biol. Med. 149 (2022) 106087, http://dx.doi.org/10.1016/j.compbiomed.2022.106087, URL https://www.sciencedirect.com/science/article/pii/S0010482522007958.

[20] E. Liu, B. He, D. Zhu, Y. Chen, Z. Xu, Tiny polyp detection from endoscopic video frames using vision transformers, Pattern Anal. Appl. 27 (2024) 38, http://dx.doi.org/10.1007/s10044-024-01254-3.

[21] M.A. Khorsandi, N. Karimi, S. Samavi, M. Hajabdollahi, S.M. Soroushmehr, K. Ward, K. Najarían, Hardware image assessment for wireless endoscopy capsules, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2016, pp. 2050–2053, http://dx.doi.org/10.1109/EMBC.2016.7591130.

[22] S. Yoo, Y. Wang, J.-M. Braun, E.S. Nadimi, Accelerating FPGA-implementations for mobile medical devices with high-level AI libraries: an object detection model for colorectal polyp images, in: 2021 IEEE International Conference on Imaging Systems and Techniques, IST, 2021, pp. 1–6, http://dx.doi.org/10.1109/IST50367.2021.9651361.

[23] H. Zhang, Z. Li, C.-K. Lee, Transmitter adaptation and wireless power control for capsule endoscopy, IEEE Trans. Power Electron. 39 (4) (2024) 4884–4894, http://dx.doi.org/10.1109/TPEL.2023.3345277.

[24] M. Commons, Inference: tiny. v1.0 Results, 2024, URL https://mlcommons.org/en/inference-tiny-10/. (Accessed: 30 October 2024).

[25] X. Li, M. Cen, J. Xu, H. Zhang, X.S. Xu, Improving feature extraction from histopathological images through a fine-tuning ImageNet model, J. Pathol. Informatics 13 (2022) 100115, http://dx.doi.org/10.1016/j.jpi.2022.100115, URL https://www.sciencedirect.com/science/article/pii/S221353392200709X.

[26] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size, 2016, arXiv:1602.07360.

[27] K. Pogorelov, K.R. Randel, C. Griwodz, S.L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P.T. Schmidt, M. Riegler, P. Halvorsen, KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection, in: Proceedings of the 8th ACM on Multimedia Systems Conference, MMSys '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 164–169, http://dx.doi.org/10.1145/3083187.3083212.

[28] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer, Int. J. Comput. Assist. Radiol. Surg. 9 (2014) 283–293, http://dx.doi.org/10.1007/s11548-013-0926-3.

[29] Y. Ma, X. Chen, K. Cheng, Y. Li, B. Sun, LDPolypVideo benchmark: A large-scale colonoscopy video dataset of diverse polyps, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International Conference, Strasbourg, France, September 27 – October 1, 2021, Proceedings, Part V, Springer-Verlag, Berlin, Heidelberg, 2021, pp. 387–396, http://dx.doi.org/10.1007/978-3-030-87240-3_37.

[30] J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, Comput. Med. Imaging Graph. : Off. J. Comput. Med. Imaging Soc. 43 (2015) 99–111, http://dx.doi.org/10.1016/j.compmedimag.2015.02.007.

[31] J. Bernal, J. Sánchez, F. Vilariño, Towards automatic polyp detection with a polyp appearance model, Pattern Recognit. 45 (9) (2012) 3166–3182, http://dx.doi.org/10.1016/j.patcog.2012.03.002, Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011). URL https://www.sciencedirect.com/science/article/pii/S0031320312001185.

[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[33] Nordic Semiconductor, nRF52840, 2025, URL https://www.nordicsemi.com/Products/nRF52840/Modules.

[34] Nordic Semiconductor, Online power profiler for bluetooth LE, 2025, URL https://devzone.nordicsemi.com/power/w/opp/2/online-power-profiler-for-bluetooth-le.

[35] H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa, Y. Yagi, Detection of contractions in adaptive transit time of the small bowel from wireless capsule endoscopy videos, Comput. Biol. Med. 39 (1) (2009) 16–26, http://dx.doi.org/10.1016/j.compbiomed.2008.10.005, URL https://www.sciencedirect.com/science/article/pii/S0010482508001522.

[36] J.H. Park, S.W. Hong, S.W. Hwang, S.H. Park, D.-H. Yang, B.D. Ye, S.-J. Myung, S.-K. Yang, J.-S. Byeon, Clinical outcomes of colonoscopic polypectomy with strategic surveillance colonoscopies in patients with 10 or more polyps, Sci. Rep. 13 (2023) 2604, http://dx.doi.org/10.1038/s41598-023-29604-x.

Paola Busia received the Ph.D. degree in Electronics and Computer Engineering from Università degli studi di Cagliari, Italy, in 2023. She is currently assistant professor at Università degli studi di Cagliari. Her main research interests involve power efficient processing architectures for embedded applications and for at-the-edge AI.

Andrea Pinna is Full Professor at Sorbonne Université (since 2023) and a member of the LIP6 laboratory (Since 2011). His research focuses on edge computing and hardware-accelerated artificial intelligence (AI), with a strong emphasis on optimizing AI models for deployment on embedded devices. He explores innovative approaches to balance performance, energy efficiency, and latency in distributed intelligent systems. His work aims to make AI more accessible and sustainable, particularly in resource-constrained environments.

Paolo Meloni is an associate professor at the University of Cagliari. His research activity is on the development of advanced digital systems, on the application-driven design and programming of multi-core on-chip architectures and FPGAs. He is the author of a significant track of international research papers. He teaches Microcontroller-based systems and Advanced embedded systems at Università degli studi di Cagliari.