Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 10 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i10.55797

PAPER

Real-Time Polyp Detection in Colonoscopy Using YOLOv8: A Fast and Accurate Deep Learning Approach

Kenza Redjimi, Mohammed Redjimi (✉), Asma Talaa

LICUS Laboratory, Université 20 Août 1955, Skikda, Algeria

m.redjimi@univ-skikda.dz

ABSTRACT

Polyps that may develop on the inner surfaces of the intestines or rectum are considered the primarily cause of colorectal cancer (CRC). To enhance survival rates, it is essential to focus on early detection, accurate prognosis, and timely treatment, typically involving surgical removal of polyps. The employment of advanced computer-aided diagnosis systems (CADx) that utilize appropriate machine learning techniques, particularly deep learning methods, aids physicians in achieving a highly relevant detection of abnormalities during internal examinations of the human body. In this context, this paper discusses a deep learning framework for automated polyp detection utilizing the you only look once (YOLO) model. This paper introduces a detection system based on the YOLOv8n model, designed for simplicity, effectiveness, cost-efficiency, and potential significant support for healthcare providers and patients in the realm of polyp detection. The results achieved are compared with those obtained using the YOLOv7 model and demonstrate enhanced performance.

KEYWORDS

you only look once (YOLO), computer-aided diagnosis systems (CADx), colonoscopy, colorectal cancer (CRC)

1 INTRODUCTION

Various illnesses, including esophagus, stomach, and colon cancers, can affect the human digestive system. According to estimates by the World Health Organization (WHO), colon cancer is the second leading cause of cancer-related deaths worldwide; these cancers result in over 1.9 million new diagnoses and 930,000 deaths in 2020 [1]. Endoscopic procedures, such as gastroscopy and colonoscopy, are considered the gold standard for examining the gastrointestinal (GI) tract. These methods are invasive and resource demanding, necessitating costly equipment and skilled professionals. In terms of preventing colorectal cancer (CRC), it is vital to identify

Redjimi, K., Redjimi, M., Talaa, A. (2025). Real-Time Polyp Detection in Colonoscopy Using YOLOv8: A Fast and Accurate Deep Learning Approach. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(10), pp. 77–93. https://doi.org/10.3991/ijoe.v21i10.55797

Article submitted 2025-03-29. Revision uploaded 2025-06-07. Final acceptance 2025-06-09.

and eliminate precancerous lesions through endoscopy, though detection rates can differ among physicians, influencing cancer risk. A precise and automated scoring system for endoscopic observations would contribute to reducing disparities, enhancing quality, and optimizing healthcare resources.

Artificial intelligence (AI) encompasses a wide area within computer science dedicated to developing intelligent machines that can carry out tasks normally requiring human intellect. AI technology has made considerable progress, particularly due to the creation of advanced techniques such as support vector machines and deep learning. By continually learning from data and gaining experience, machines improve their capabilities in task processing. AI has found increasing application in medical domains such as imaging, pathological diagnosis, disease management, and drug development, leading to advancements in genetics and molecular medicine. Numerous studies confirm this trend, displaying the potential of AI in identifying new treatment methods through algorithms for the advancement of molecular medicine. Additionally, AI systems that employ deep learning have shown exceptional effectiveness in diagnosing conditions such as esophageal cancer, glaucoma, and skin cancer via medical imaging. The last few years have also seen an increase in AI applications for the diagnosis and treatment of colorectal polyps and cancer, especially in gastroenterology, where AI assists in the detection of polyps during endoscopic procedures [2–5]. For example, ResYOLO [6], a rapid detection algorithm, accomplished remarkable accuracy in recognizing polyps from colonoscopy videos. Advancements in deep learning algorithms and computing capabilities are paving the way for real-time AI support during colonoscopy procedures. Many internal diseases, such as polyps—abnormal growths in the colon—can develop silently and lead to severe outcomes if not detected and treated promptly. Certain types of polyps have the potential to progress into CRC [7]. Detecting and removing these polyps at early stages can significantly reduce the risk of developing this serious illness through straightforward medical interventions. Unfortunately, in advanced stages, treatment options may no longer be effective, making early detection crucial for preventing CRC, which is responsible for numerous deaths worldwide [8]. Research in [9] highlights that adenoma size is a key factor; for instance, cancer develops in only 1% of adenomas less than 1 cm, but this increases to 10% for those between 1 cm and 2 cm and escalates to 50% for adenomas greater than 2 cm. One critical factor in the development of CRC is the size of colon polyps. Research shows that polyps larger than 10 mm have a significantly higher likelihood of turning cancerous [10]. There is a direct relationship between the stages of polyps and their sizes (see Figure 1). Essentially, the larger the polyp, the longer it has been present, making it easier to detect. However, accurately identifying smaller polyps can be challenging, even for experienced practitioners, as it requires meticulous examination of all colon tissue, a task that is often time-consuming. Unfortunately, not every examination allows for this level of attention.

The outlook for recovery from CRC diagnosed at advanced stages is quite grim, as malignant tumors can spread to various parts of the body (metastasis), often leading to fatal outcomes. Detecting polyps early can facilitate their removal and significantly improve patient outcomes. Regular screenings, especially for individuals over the age of 50, are vital, even if those younger may also benefit from preventive evaluations.

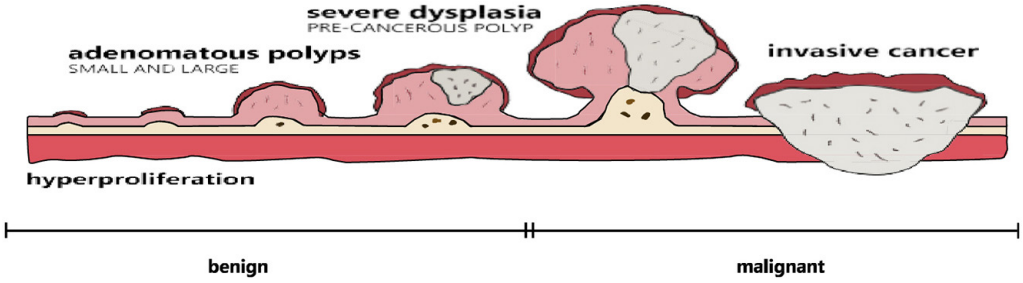

Fig. 1. Benign to malignant progression of colorectal polyps [11]

In this context, colonoscopy [12] remains the most widely employed method for investigating the colon and determining the presence of polyps. While it is generally an effective technique, it is considered highly invasive and may be performed with or without anesthesia, depending on the patient preference. Successful colonoscopy relies on adequate bowel preparation and the use of air insufflation to examine the bowel thoroughly [13]. This procedure is time-consuming and comes with significant financial costs. Moreover, its success highly depends on the gastroenterologist's expertise, and studies show that polyps, especially smaller ones, are often overlooked during examinations, which can lead to complications later on.

Traditionally, the gastroenterologist uses a long, flexible tube inserted into the patient's rectum to visualize the colon and identify any abnormalities. Alternatively, there is a method involving a small wireless capsule endoscopy [14]. This technique utilizes a tiny camera housed within a capsule approximately twenty millimeters in size, resembling a standard medicine capsule. Integrated electronics allow the capsule to capture images of the internal tissues, while a battery provides power for the device. After the patient ingests the capsule, it travels through the digestive tract, transmitting thousands of images to external processing and storage devices. The main advantages of this method are its user-friendliness, less invasive nature compared to traditional endoscopy, and capability to film the internal lining of the small intestine effectively.

Virtual colonoscopy [15] is a cutting-edge, minimally invasive procedure that utilizes an X-ray technique to capture multiple images of the large intestine. These images are then processed by computer technology to create detailed two-dimensional or three-dimensional representations of the colon's interior [16].

In recent years, advancements in AI have led to various techniques and algorithms being implemented for the detection, monitoring, and diagnosis of several diseases. This study specifically explores the detection of polyps utilizing YOLOv8 [17, 18], alongside the Kvasir-SEG [19, 20] and BKAI-IGH NeoPolyp-Small datasets [21], and comparing the results to those from YOLOv7.

The remainder of this paper is organized as follows: Section 2 outlines the primary objectives of this study; Section 3 details our methodology, providing essential background information as well as the materials and methods used; Section 4 discusses the results obtained; and Section 5 wraps up the study with concluding remarks.

2 MAIN OBJECTIVES

The main objectives of this study are:

- To develop a fast and accurate polyp detection system using YOLOv8.
- To train and evaluate the model using Kvasir-SEG and BKAI-IGH NeoPolyp-Small datasets.

– To compare the performance of YOLOv8 with YOLOv7 in terms of detection accuracy and speed.
– To assess the impact of data augmentation on model robustness.

2.1 Backgrounds

Artificial intelligence is a rapidly expanding field with numerous applications emerging across various sectors. At its core, this technology involves equipping machines with specialized algorithms and hardware that allow them to replicate human behaviors [22]. One of the most sought-after capabilities is automatic vision, which entails setting up systems with cameras to capture images, along with algorithms designed to process these visual inputs. In the realm of image recognition, there are four key aspects to consider:

i. Object classification that predicts to what category an object belongs.
ii. Object localization, which identifies the specific location of one or more objects within an image.
iii. Object detection, focused on determining the presence of an object in an image.
iv. Object segmentation, which emphasizes the relevant information concerning the identified objects within the image.

Several studies have been conducted in this area, resulting in the development of various algorithms and specialized materials that have shown positive outcomes. However, these solutions often fall short in terms of precision and execution speed, particularly when it comes to real-time applications. The algorithms' complexity and long computation times pose significant challenges.

In recent years, advancements in hardware and software engineering have led to notable improvements. Specifically, new AI techniques, especially those based on deep learning, have increasingly found applications in machine vision. Within the object recognition domain, convolutional neural network (CNN)-based methods have led to significant implementations. There are several models for object detection, including region-based convolutional neural network (R-CNN) [23], fast R-CNN, and faster R-CNN [24–26].

In 2016, Joseph Redmon and his team introduced the innovative concept of 'you only look once' (YOLO) for object recognition [27]. YOLO operates differently from other deep convolutional networks such as the R-CNN models. While R-CNN models process an image in two stages: first identifying interesting regions and then detecting objects using a second neural network. This approach can be slow, especially with numerous regions. In contrast, YOLO is a deep neural network that analyzes the entire image in a single stage, combining both object localization and classification. This unique approach enables YOLO to achieve faster real-time detection in video contexts.

Several iterations of YOLO have been developed, each enhancing performance over the last. As of January 2023, Ultralytics released YOLOv8, which showcases advancements in accuracy, simplicity, and execution speed for real-time object detection, classification, and segmentation. Following that, YOLOv9 and YOLOv10 were introduced in 2024. YOLO v11 and YOLO v12 are available from 2025 [28].

2.2 Research methodology

Detecting polyps with the YOLO framework involves a few essential steps, which can be outlined as follows:

The data collection and annotation:

- Gather a set of medical images that feature polyps. Each image must be annotated with precise bounding boxes around every polyp.
- Verify that the annotations are accurate, preferably with the help of medical professionals, to ensure high-quality data for training.

The data preprocessing:

- Standardize image sizes to a consistent resolution suitable for input into the YOLO model.
- Normalize pixel values to a typical range (such as [0, 1] or [–1, 1]).
- If needed, enhance the dataset through augmentation methods to bolster model variability and robustness (e.g., rotating, flipping, or adding noise).

The model selection:

- Pick the right YOLO variant based on the computational capabilities and performance needs.
- Think about adjusting aspects of the YOLO architecture, such as anchor box sizes or layers, depending on how polyps appear in the medical images.

The model training:

- Start the YOLO model with weights that have been trained on a large dataset (like COCO) to help improve convergence speed and overall performance.
- Fine-tune the model using techniques of transfer learning on the polyp dataset.
- Monitor the model's performance during training with metrics such as mean average precision (mAP) to gauge its effectiveness.

The post-processing steps:

- Implement non-maximum suppression (NMS) to filter out redundant bounding box predictions and retain only the most confident detections.
- Adjust confidence thresholds to balance precision and recall based on specific application requirements.

Evaluation:

- Assess the trained model's performance on a separate validation set to measure its accuracy in polyp detection.
- Evaluate metrics such as precision, recall, and F1-score to quantify the model's effectiveness in detecting polyps.

The deployment and integration:

- Integrate the trained YOLO model into a deployment pipeline for real-time or batch processing of medical images.
- Ensure compliance with regulatory standards and ethical guidelines for medical AI applications.

Iterative improvements:

- Continuously refine the model based on feedback from medical experts and additional annotated data.
- Explore advanced techniques such as ensemble methods or domain-specific adjustments to enhance detection accuracy and robustness.

In the field of medicine, supervised machine learning techniques utilize established image and video datasets for training and testing specialized algorithms. Typically, in a dataset of a given size, one portion is dedicated to learning and another is reserved for testing. Several datasets are regularly employed in medical research. For instance, Table 1 lists some endoscopic image and video polyp datasets that are currently available. Based on our study, Kvasir stands out as one of the most frequently utilized datasets, featuring a substantial collection of annotated samples, which include both ground truth masks and bounding boxes essential for detection, localization, and segmentation tasks. Public datasets can be accessed directly; however, datasets requiring prior permission from their owners can lead to delays due to the need for agreement on their usage.

Table 1. Some polyp's datasets

Dataset	Size	Availability
Kvasir-SEG [20]	1000 images	Public dataset
CVC-ColonDB [29]	380 images	Public dataset
ETIS-Larib Polyp DB [30]	196 images	Public dataset
CVC-ClinicDB [31]	612 images	Public dataset
BKAI-IGH NeoPolyp-Small [21]	1200 images	Public dataset
BKAI-IGH NeoPolyp-Large [21]	6630 images	Public dataset
PolypGen [32, 33]	1537 images	Public dataset
CVC-VideoClinicDB [34]	11954 images	Need request
ASU-Mayo polyp database [35]	18781 images	Need request

2.3 The used datasets

In this study, we aimed to train and assess the YOLOv8 model for polyp detection using two datasets: KVASIR-SEG [18] and BKAI-IGH NeoPolyp-Small [20]. Merging these datasets not only increases the image variety but also enhances the reliability and general effectiveness of our findings.

The KVASIR-SEG dataset was employed for tasks involving detection, localization, and segmentation. As demonstrated in Figure 3, we present examples of images alongside their ground truth data and corresponding detections. This dataset was specifically designed to support transparent and reproducible research.

The Kvasir-SEG dataset comprises a well-annotated collection of images meticulously verified by seasoned medical professionals, particularly endoscopists. It includes various classes that display anatomical landmarks, pathological conditions, and endoscopic techniques within the GI tract. Each class contains hundreds of high-quality images, making the dataset suitable for image retrieval, machine learning, deep learning, and transfer learning applications. The images and their associated ground truth data serve well for segmentation tasks, while the bounding box information facilitates detection endeavors.

Anatomical landmarks featured in the dataset include the Z-line, pylorus, and cecum, as well as pathological conditions such as esophagitis, polyps, and ulcerative colitis. Also included are images connected to lesion removal procedures such as "dyed and lifted polyp" and "dyed resection margins."

Image resolutions vary from 720 × 576 to 1920 × 1072 pixels and are organized into distinct folders based on content. Certain images contain a green picture-in-picture overlay illustrating the endoscope's position and settings within the bowel, utilizing an electromagnetic imaging system (ScopeGuide, Olympus Europe) to enhance image interpretation (see Figure 2). While this supplementary information can support further investigations, care must be taken to ensure accurate detection of endoscopic findings.

This dataset is readily accessible for research purposes and can be downloaded via the following open-access link: https://datasets.simula.no/kvasir-seg/.

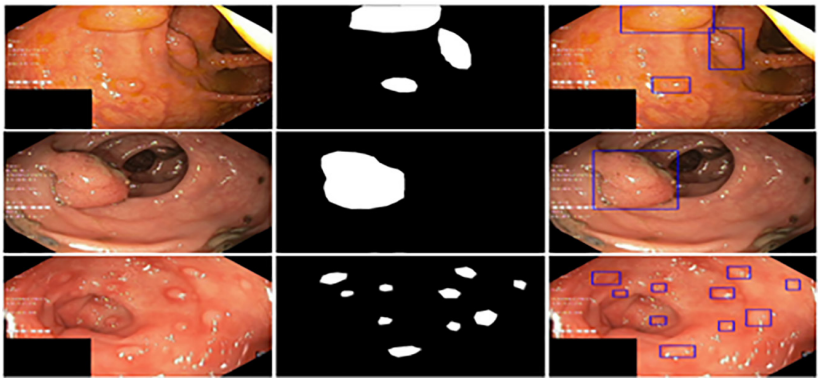

Fig. 2. Annotated masks and bounding boxes from the Kvasir-SEG dataset [19]

The BKAI-IGH NeoPolyp-Small dataset, which is publicly available (https://www.kaggle.com/c/bkai-igh-neopolyp/), consists of 1200 images. It is split into a training set with 1000 images and a test set containing 200 images. Polyps in this dataset are categorized as neoplastic (indicated by the color red) or non-neoplastic (indicated by the color green). This dataset is suitable for both polyp segmentation and detection tasks, and it also provides insights for identifying the neoplastic characteristics of polyps (see Figure 3).

The BKAI-IGH NeoPolyp-Small dataset was created through a partnership between BKAI, Hanoi University of Science and Technology, and the Institute of Gastroenterology and Hepatology (IGH) in Vietnam.

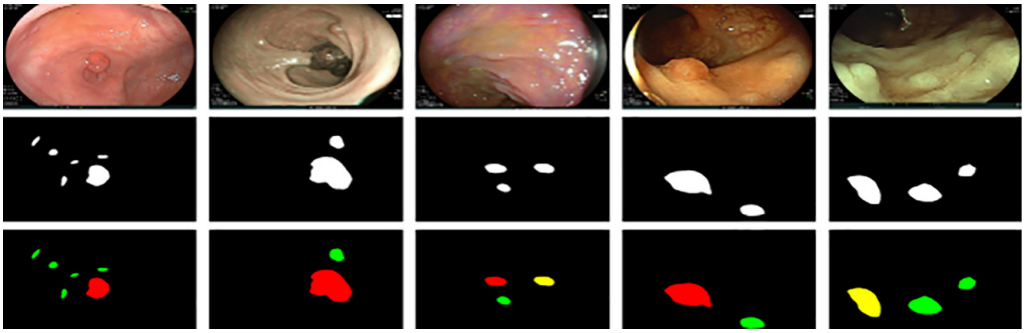

Fig. 3. Samples from the BKAI-IGH NeoPolyp-small dataset

Notes: Original images are shown in the first row. The second row shows the ground truths for polyp segmentation. The ground truths for neoplasm segmentation are shown in the third row. The undefined polyps are shown in yellow [21].

2.4 Material and method

The YOLOv8 model brings several cutting-edge features that significantly boost its object detection performance. Notably, it introduces a revamped backbone network structure and anchor-free localization, which contribute to more efficient NMS. Additionally, the model includes a new loss function and replaces the previous C3 module with the C2f module. This enhancement allows the model to integrate outputs from two 3 × 3 convolutions using residual connections, rather than relying solely on the last bottleneck output. The design has also upgraded the initial 6 × 6 convolution in the backbone to a more streamlined 3 × 3 convolution block. YOLOv8 comes in five different sizes: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large). For this study, we chose the YOLOv8n due to its simplicity, effectiveness, and the available resources (see Figure 4).

2.5 The data preparation

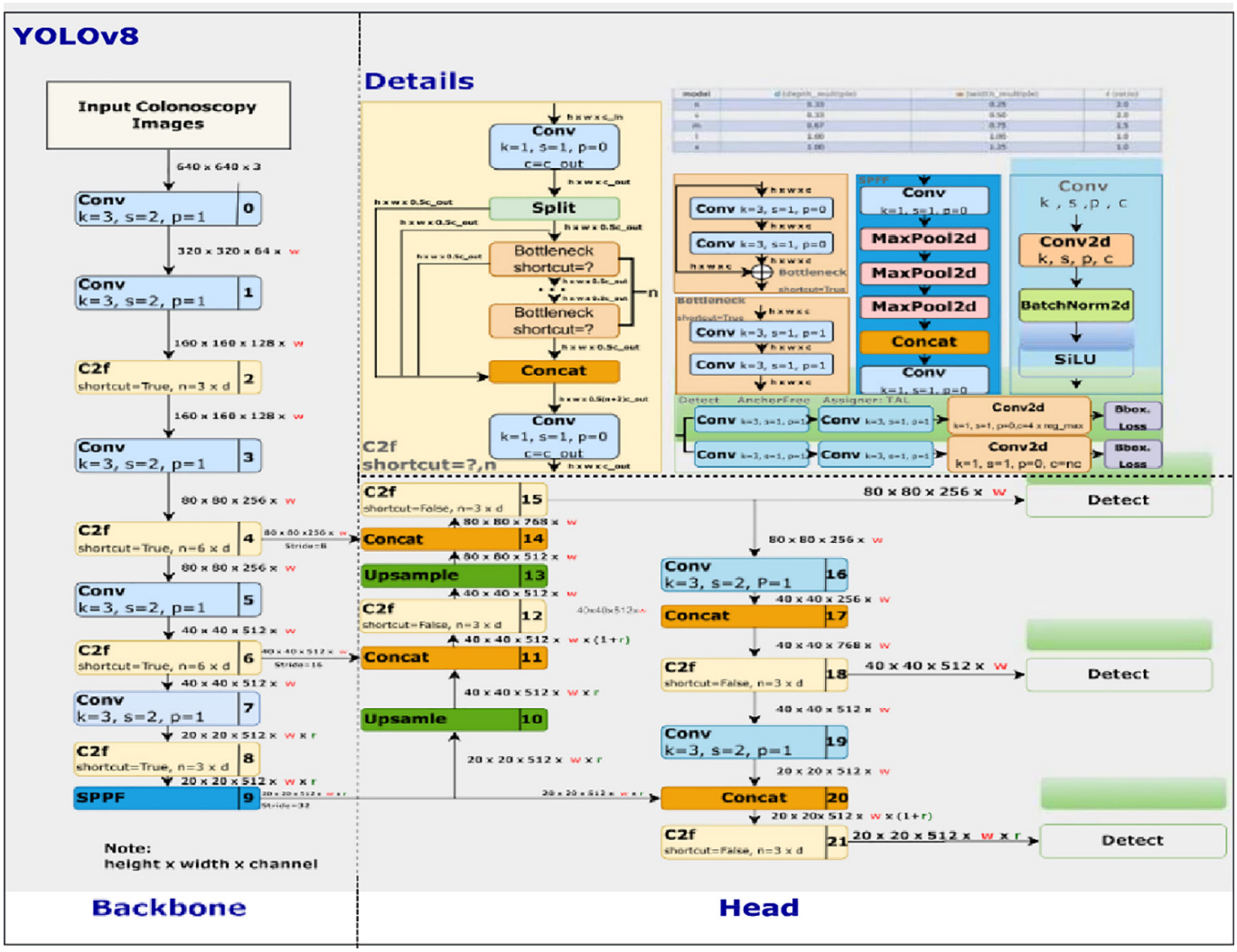

Fig. 4. YOLO V8: model architecture including backbone and head [36]

To set up the environment for our dataset in a format that works well with the YOLOv8 algorithm, we can repurpose some existing configuration files and tweak them as necessary. This method can simplify the process of preparing our dataset for YOLOv8.

The dataset configuration. The initial step is to create the dataset configuration file, named polyp.yaml. This file should outline the following parameters:

- Path train: Specify the training dataset path, which should point to a YAML file that includes the image paths and their respective labels.
- Path value: Indicate the validation dataset path, as well as a YAML file that contains the image paths and labels.
- Path test: Provide the testing dataset path, again a YAML file with the image paths and labels.
- Number of classes (NC): In our case, nc = 2, as the classes are defined as 'polyp' and 'not-polyp'.
- List names (ln): Include a list of the class names.

The data annotation. In the next phase, we focus on data annotations, which involve labeling a dataset to create a collection of data enriched with extra information, usually added by humans for added context and insight. These annotations can appear in various forms, including text tags, semantic labels, or visual cues, and they play a crucial role in helping machine learning algorithms comprehend the data better and enhance prediction accuracy. For our study, we utilized colonoscopy images that were labelled with bounding boxes around polyps, employing Roboflow's annotation tools [37]. Roboflow is an intuitive online platform designed for creating and managing annotation datasets specifically for computer vision tasks. It offers an easy-to-use interface that enables users to upload and label their images effortlessly. Once the annotations were finalized, we exported the data into a format compatible with the YOLOv8 machine-learning framework.

The data augmentation. Data augmentation is a widely used technique in machine learning that enhances the training dataset's size by applying various transformations to the existing data. This approach can significantly boost the performance of a model, particularly when the available training data is scarce. Moreover, data augmentation plays a vital role in mitigating overfitting, a scenario where the model becomes overly attuned to the training data and struggles with new, unseen information. By introducing variations—such as rotating or flipping images—data augmentation compels the model to generalize and recognize relevant patterns for the task, rather than merely memorizing the specific examples encountered during training. An illustrative example of data augmentation can be seen in Figure 5.

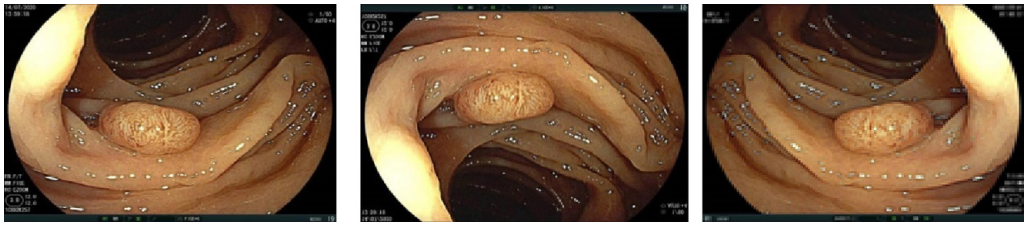

Fig. 5. Examples of the data augmentation visualization using horizontal and vertical flips

Training the model. During the model-training phase, we employed the task-aligned assigner from task-aligned one-stage object detection (TOOD) [38] to

distribute positive and negative samples. This method identifies positive samples by assessing a combination of weighted scores based on both classification and regression metrics, as outlined in equation (1).

$$t = s^{\alpha} \cdot u^{\beta} \tag{1}$$

To implement object detection, we adopted a stratified approach to distribute the dataset effectively. In this context 's' represents the predicted score for each detected class, while 'u' indicates the intersection over union (IoU) between the predicted and ground truth bounding boxes. The model features distinct branches dedicated to classification and regression tasks, with α and β controlling their influence on the anchor alignment metric.

For our dataset, we divided it into training, validation, and test sets to maintain a consistent class distribution. Specifically, 70% of the total dataset was designated for training, with the remaining 20% for validation and 10% for testing. We utilized the Google Colaboratory cloud platform [39] for model training, which provides robust GPU capabilities and requires no extensive setup. The dataset was conveniently downloaded onto Google Colaboratory via a Roboflow-generated zip folder.

The YOLOv8 PyTorch implementation employed specific parameters during training, with the model trained over 35 epochs at an image resolution of 640 × 640. Throughout the training process, we assessed the model's performance on the validation dataset using mAP as our primary metric. Additionally, visual inspections of the model's predictions were conducted to enhance our understanding of its performance. Upon completion of the training, the model was evaluated on a separate test set, using the same evaluation metrics as the validation phase. This thorough evaluation process ensured that the model was not overfitting the training data and maintained its effectiveness on unseen data.

3 RESULTS AND ANALYSIS

3.1 Used metrics

In this study, we utilized four key metrics to assess the performance of the polyp detection models: precision, recall, accuracy, and F1 score. These metrics rely on the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

The TP metric indicates the number of polyps accurately detected in the images. In contrast, the FP metric counts the non-polyp items mistakenly labelled as polyps. The TN metric reflects the correct identification of non-polyp objects, while the FN metric accounts for polyps that went undetected by the model.

Precision is defined as the ratio of true positive predictions to the total number of positive predictions, whereas recall measures the proportion of TF that the model correctly identified from the overall actual positive samples. Below are the formulas for calculating precision and recall, given in (2) and (3), respectively:

$$precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$recall = \frac{TP}{(TP + FN)} \tag{3}$$

Accuracy serves as a vital performance metric that calculates the ratio of correct predictions made by a model in relation to the total predictions made (4):

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{4}$$

The F1 score represents a balance between precision and recall by averaging them through the harmonic mean, considering both FP and FN in the process. The formula for calculating the F1 score is illustrated below (5):

$$F1_{score} = 2\frac{precision * recall}{precision + recall} \tag{5}$$

The mAP evaluates the effectiveness of a system in retrieving relevant items based on a user's query. This metric is determined by averaging the precision scores at different recall levels (6):

$$mAP = \sum_{k=1}^{k=n} AP_k \tag{6}$$

3.2 Results and discussion

The research focused on creating a polyp detection model utilizing a specially designed dataset. This model demonstrated outstanding capabilities, reaching a mAP of 96.6%. This impressive score suggests it effectively identifies polyps within the dataset. In addition, Figure 6 illustrates various components of the training evaluation metrics.

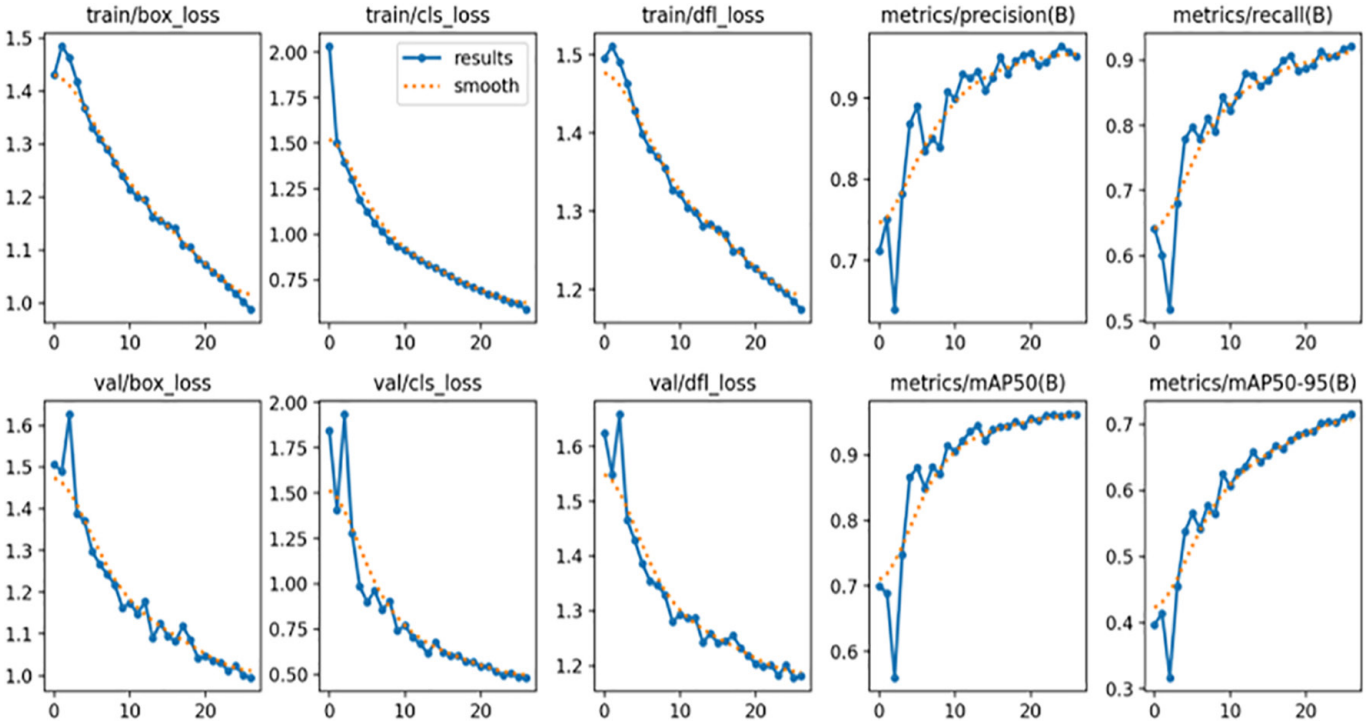

Fig. 6. The training results (x-axis: epochs)

Figure 6 presents the graphs illustrating the bounding box loss (box_loss), classification loss (cls_loss), and deformable convolution layer loss (dfl_loss) for both training and validation datasets. The bounding box loss during prediction quantifies the difference between the predicted bounding box and the actual one. Essentially, a lower box loss indicates that the algorithm is more proficient at predicting a bounding box that closely aligns with the actual one. The deformable convolution layer loss represents a recent enhancement in the YOLO architecture, specifically in YOLOv8. This loss quantifies errors within the deformable convolution layers, which are tailored to boost the model's capacity to detect objects across a variety of scales and aspect ratios. A reduced "dfl_loss" signifies that the model excels at managing variations in object shapes and appearances. Classification loss assesses the accuracy of the predicted class p probabilities for each object in the image against the ground truth. A lower "cls_loss" indicates heightened accuracy in class prediction. Notably, the model exhibited marked improvements in precision, recall, and mAP after about 20 epochs, with rapid declines in bounding box loss and overall object loss observed around the 10-epoch mark. Furthermore, the model achieved impressive average precision and recall rates of 96% and 92.9%, respectively, highlighting its robustness in identifying objects even in challenging conditions. The F1 score, recorded at 94.4%, further underscores the model's strong generalization capabilities, suggesting its proficiency in accurately detecting objects in new, unseen images with high precision and recall.

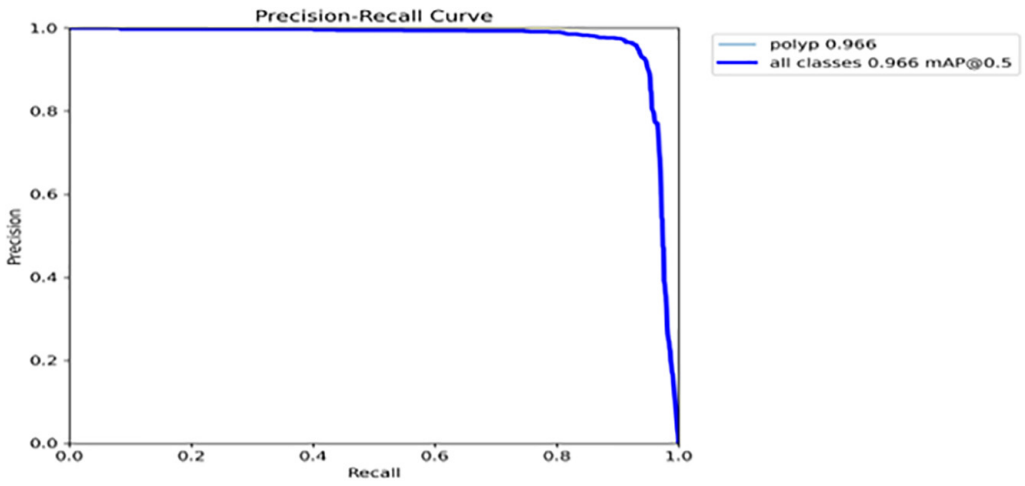

Fig. 7. The precision-recall curve

Figure 7 presents the precision recall (PR) curve for the YOLOv8 model trained on our dataset. The graph reveals an impressive balance between precision and recall, achieving an impressive mAP@0.5 value of 0.966, which highlights a favorable area under the PR curve. These results demonstrate the effectiveness of our approach in training an object detection model tailored for a custom dataset. Additionally, the findings indicate that the model could be applied in various practical scenarios within the targeted domain. The outcomes of this study hold significant value for researchers and practitioners who are keen on developing object detection models for unique datasets where pre-trained models might not suffice. Table 2 outlines the results of our model with and without data augmentation, while Table 3 compares the performance of YOLOv7 and YOLOv8.

Table 2. The performances of yolov8 trained on both Kvasir-seg and Bkai-IGH NeoPolyp-Small datasets

Dataset	mAP_50	Precision	Recall	F1_score
Kvasir-seg without augmentation	89.26%	87.86%	81.21%	84.40%
BKAI-IGH NeoPolyp-Small without augmentation	84.89%	94.93%	78.09%	85.69%
Kvasir-seg with augmentation	90.31%	93.39%	76.27%	84.76%
BKAI-IGH NeoPolyp-Small with augmentation	88.32%	92.65%	82.79%	87.44%
Kvasir-seg & BKAI-IGH NeoPolyp-Small without augmentation	92.03%	91.52%	86.10%	88.72%
Kvasir-seg & BKAI-IGH NeoPolyp-Small with augmentation	**96.63%**	**96.00%**	**92.92%**	**94.43%**

Table 3. Results obtained from validation of YOLOv8 and YOLOv7 models trained on both Kvasir-seg and Bkai-IGH NeoPolyp-Small datasets.

Models	mAP_50	Precision %	Recall %	F1_score
YOLOv7	83.72%	86.54%	79.98%	83.13%
YOLOv8	96.63%	96%	92.92%	94.43%

We carried out a comparative analysis of YOLOv8 alongside another well-known object detection model, YOLOv7, utilizing a dataset we developed. The findings, summarized in Table 3, indicate that YOLOv8 outshines YOLOv7 in both object detection and recognition, exhibiting a reduced number of errors. Like its counterparts, YOLOv8 was trained using a loss function, but it distinguishes itself as the fastest general-purpose object detection model currently available, solidifying its status as the leading technology in this domain.

In this study, we trained a YOLOv8 nano model (YOLOv8n) from scratch. Inference was performed at a speed of 0.99 milliseconds per image on an NVIDIA A100 GPU using TensorRT-optimized weights. This variant was selected due to its fast prediction capabilities, making it well suited for real-time polyp detection and localization tasks.

3.3 Comparison

The current state-of-the-art in CRC detection and prediction is typically a combination of deep learning models tailored for medical imaging, namely segmentation-based CNNs (e.g., UNet, UNet++, ResUNet, or TransUNet) and classification models (e.g., ResNet, EfficientNet, or more recently Vision Transformers) [40–42]. The models tend to be trained on endoscopic or histopathology images and fine-tuned for high sensitivity and specificity in the detection of polyps or tumors.

YOLOv8, while being a general object detection model, has great advantages for CRC detection in real-time colonoscopy videos or high-resolution images due to its speed (real-time inference, low latency), compact architecture (which eases the deployment on low-resource devices), and improved accuracy over previous YOLO models (due to anchor-free detection and decoupled head).

While older models such as UNet-based architectures may offer slightly improved pixel-level segmentation performance, YOLOv8 is significantly better at fast and accurate localization of CRC-related features, making it highly suitable for real-time clinical support during colonoscopy procedures.

4 CONCLUSION

The use of AI in developing computer-aided diagnosis (CADx) systems, particularly through advanced machine learning and deep learning methods, is proving to be invaluable for healthcare professionals aiming to detect medical abnormalities. In this study, we applied the YOLOv8 model, which yielded impressive results in detecting polyps. To enhance our dataset, we employed two techniques: merging the Kvasir_SEG and BKAI-IGH NeoPolyp-Small datasets, alongside data augmentation strategies. Our evaluation indicates that the YOLOv8 model represents a notable advancement over earlier YOLO versions, demonstrating competitive accuracy and speed when placed alongside other leading object detection models. However, it is important to note that implementing a YOLO-based deep learning system poses challenges, including significant storage requirements, computing power, and time commitments for deployment. Moving forward, our focus will be on improving both the model's speed and accuracy and developing practical applications in the medical field.

In future work, we aim to extend this system for real-time video-based colonoscopy analysis to assist doctors during live procedures.

5 REFERENCES

[1] World Health Organization (WHO), "Colorectal cancer," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer [Accessed: Mar. 25, 2025].

[2] M. Misawa *et al.*, "Artificial intelligence-assisted polyp detection for colonoscopy: Initial experience," *Gastroenterology,* vol. 154. no. 8, pp. 2027–2029, 2018. https://doi.org/10.1053/j.gastro.2018.04.003

[3] I. Barua *et al.*, "Artificial intelligence for polyp detection during colonoscopy: A systematic review and meta-analysis," *Endoscopy,* vol. 53, no. 3, pp. 277–284, 2021. https://doi.org/10.1055/a-1201-7165

[4] F. Oliveira *et al.*, "Automatic detection of polyps using deep learning," in *International Conference on Wireless Mobile Communication and Healthcare*, A. Cunha, A. Paiva, and S. Pereira, Eds., Springer, Charm, 2023, pp. 254–263. https://doi.org/10.1007/978-3-031-60665-6_19

[5] F. Chadebecq *et al.*, "Artificial intelligence and automation in endoscopy and surgery," *Nature Reviews Gastroenterology & Hepatology*, vol. 20, pp. 171–182, 2023. https://doi.org/10.1038/s41575-022-00701-y

[6] F. He *et al.*, "Colonoscopic image synthesis for polyp detector enhancement via gan and adversarial training," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI),* 2021, pp. 1887–1891. https://doi.org/10.1109/ISBI48211.2021.9434050

[7] O. L. Pop *et al.*, "An overview of gut microbiota and colon diseases with a focus on adenomatous colon polyps," *Int. J. Mol. Sci.*, vol. 2020, no. 19, p. 7359, 2020. https://doi.org/10.3390/ijms21197359

[8] M. S. Hossain *et al.*, "Colorectal cancer: A review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies," *Cancers*, vol. 14, no. 7, p. 1732, 2022. https://doi.org/10.3390/cancers14071732

[9] V. Conteduca *et al.*, "Precancerous colorectal lesions," *International Journal of Oncology*, vol. 43, no. 4, pp. 973–984, 2013. https://doi.org/10.3892/ijo.2013.2041

[10] Daniel Yetman, "Colon polyp sizes and types," healthline, 2023. [Online]. Available: https://www.healthline.com/health/colorectal-cancer/colon-polyp-size-chart#treatment [Accessed: Mar. 26, 2025].

[11] Harvard Health Publishing, "They found colon polyps: Now what?" 2023. [Online]. Available: https://www.health.harvard.edu/diseases-and-conditions/they-found-colon-polyps-now-what [Accessed: Mar. 26, 2025].

[12] Y. Saito *et al.*, "Colonoscopy screening and surveillance guidelines," *Digestive Endoscopy,* vol. 33, no. 4, pp. 486–519, 2021. https://doi.org/10.1111/den.13972

[13] V. O. Millien and N. M. Mansour, "Bowel preparation for colonoscopy in 2020: A look at the past, present, and future," *Current Gastroenterology Reports,* vol. 22, pp. 1–9, 2020. https://doi.org/10.1007/s11894-020-00764-4

[14] V. Blanes-Vidal, G. Baatrup, and E. S. Nadimi, "Addressing priority challenges in the detection and assessment of colorectal polyps from capsule endoscopy and colonoscopy in colorectal cancer screening using machine learning," *Acta Oncologica*, vol. 58, pp. 529–536, 2019. https://doi.org/10.1080/0284186X.2019.1584404

[15] N. Gluck, B. Shpak, R. Brun, T. Rösch, N. Arber, and M. Moskowitz, "A novel prepress X-ray imaging capsule for colon cancer screening," *Gut*, vol. 65, pp. 371–373, 2016. https://doi.org/10.1136/gutjnl-2015-310893

[16] M. G. Demirci and Y. M. Kesgin, "Virtual colonoscopy: Retrospective comparison of the findings in supine and prone positions," *Surgical Innovation,* vol. 32, no. 3, pp. 242–248, 2025. https://doi.org/10.1177/15533506251325349

[17] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond," *arXiv preprint arXiv:2304.00501,* 2023. http://dx.doi.org/10.48550/arXiv.2304.00501

[18] Ultralytics, "Explorer Ultralytics YOLOv8," 2025. https://docs.ultralytics.com/fr/models/yolov8/ [Accessed: Mar. 26, 2025].

[19] D. Jha *et al.*, "Kvasir-seg: A segmented polyp dataset," in *MultiMedia Modeling: 26th International Conference, MMM 2020*, Daejeon, South Korea, January 5–8, 2020, Springer International Publishing, 2019, pp. 451–462. https://doi.org/10.1007/978-3-030-37734-2_37

[20] Meet Nagadia, "Kvasir Dataset," *kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/meetnagadia/kvasir-dataset [Accessed: Mar. 26, 2025].

[21] Dinh Sang, "BKAI-IGH NeoPolyp," *kaggle*, 2021. [Online]. Available: https://www.kaggle.com/c/bkai-igh-neopolyp/ [Accessed: Mar. 26, 2025].

[22] P. Klare *et al.*, "Automated polyp detection in the colorectum: A prospective study (with videos)," *Gastrointestinal Endoscopy,* vol. 89, no. 3, pp. 576–582, 2018. https://doi.org/10.1016/j.gie.2018.09.042

[23] C. Chen, M. Y. Liu, O. Tuzel, and J. Xiao, "R-CNN for small object detection," in *Lecture Notes in Computer Science*, vol. 10115, S. H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds., Springer, Cham, 2017. https://doi.org/10.1007/978-3-319-54193-8_14

[24] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169

[25] A. B. Amjoud and M. Amrouch, "Object detection using deep learning, CNNs and vision transformers: A review," *IEEE Access*, vol. 11, pp. 35479–35516, 2023. https://doi.org/10.1109/ACCESS.2023.3266093

[26] P. Bharati and A. Pramanik, "Deep learning techniques—R-CNN to mask R-CNN: A survey," in *Computational Intelligence in Pattern Recognition: Proceedings of CIPR*, vol. 999, A. Das, J. Nayak, B. Naik, S. Pati, and D. Pelusi, Eds., Springer, Cham, 2019, pp. 657–668. https://doi.org/10.1007/978-981-13-9042-5_56

[27] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

[28] Ultralytics, "Models Supported by Ultralytics," 2024. [Online]. Available: https://docs.ultralytics.com/models/ [Accessed: Mar. 26, 2025].

[29] Long Việt, "CVC-ColonDB," *kaggle*, 2024. [Online]. Available: https://www.kaggle.com/datasets/longvil/cvc-colondb [Accessed: Mar. 26, 2025].

[30] Francis Jesmar Montalbo, "WCE curated colon disease dataset deep learning," *kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning [Accessed: Mar. 26, 2025].

[31] Balraj Ashwath, "CVC-ClinicDB," *kaggle*, 2020. [Online]. Available: https://www.kaggle.com/datasets/balraj98/cvcclinicdb [Accessed: Mar. 26, 2025].

[32] paperswithcopde, "PolypGen," 2025. https://paperswithcode.com/dataset/polypgen [Accessed: Mar. 26, 2025].

[33] Debesh Jha, "PolypGen Video Sequence," *kaggle*, 2023. [Online]. Available: https://www.kaggle.com/datasets/debeshjha1/polypgen-video-sequence [Accessed: Mar. 26, 2025].

[34] S. Ali *et al.*, "Assessing generalizability of deep learning-based polyp detection and segmentation methods through a computer vision challenge," *Scientific Reports*, vol. 14, 2024. https://doi.org/10.1038/s41598-024-52063-x

[35] Truong Dang Manh, "Where can I download the ASU-MAYO database for polyp segmentation?" *kaggle*, 2021. [Online]. Available: https://www.kaggle.com/discussions/questions-and-answers/239094 [Accessed: Mar. 26, 2025].

[36] M. Lalinia and A. Sahafi, "Colorectal polyp detection in colonoscopy images using YOLO-V8 network," *Signal, Image and Video Processing*, vol. 18, pp. 2047–2058, 2024. https://doi.org/10.1007/s11760-023-02835-1

[37] roboflow, "Everything you need to build and deploy computer vision applications," 2025. https://roboflow.com/ [Accessed: Mar. 26, 2025].

[38] C. Feng *et al.*, "Tood: Task-aligned one-stage object detection," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 3490–3499. https://doi.org/10.1109/ICCV48922.2021.00349

[39] https://colab.research.google.com/ [Accessed: Mar. 26, 2025].

[40] P. Rajesh *et al.*, "Lung cancer diagnosis and treatment using AI and mobile applications," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 14, no. 17, pp. 189–203, 2020. https://doi.org/10.3991/ijim.v14i17.16607

[41] A. Saeed *et al.*, "Improving the imbalanced data accuracy using CNN and ReLU," *IETI Transactions on Data Analysis and Forecasting (iTDAF)*, vol. 2, no. 3, pp. 50–58, 2024. https://doi.org/10.3991/itdaf.v2i3.51013

[42] N. I. Nife and M. Chtourou, "Improved detection and tracking of objects based on a modified deep learning model (YOLOv5)," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 21, pp. 145–160, 2023. https://doi.org/10.3991/ijim.v17i21.45201

6 AUTHORS

Kenza Redjimi obtained a PhD. in computer science from Université 20 Août 1955, Skikda, (2022), Skikda, Algeria. She is an associate professor and member

of the LICUS (Laboratoire d'Informatique et de Communication de l'Université de Skikda) at this university. (E-mail: k.redjimi@univ-skikda.dz).

Mohammed Redjimi is a full professor at Université 20 Août 1955, Skikda, Algeria. He received a doctorate in computer science from University of Lille1, France (E-mail: m.redjimi@univ-skikda.dz).

Asma Talaa holds a Master degree in computer science from Université 20 Août 1955, Skikda, Algeria (E-mail: asmatalaa4@gmail.com).