www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

Mapping the colon through the colonoscope's coordinates – The Copenhagen Colonoscopy Coordinate Database

Kristoffer Mazanti Cold[1,2] ✉, Anishan Vamadevan[1], Amihai Heen[1], Andreas Slot Vilmann[1,3], Mustafa Bulut[1], Bojan Kovacevic[1], Morten Rasmussen[4,5], Lars Konge[1,2] & Morten Bo Søndergaard Svendsen[1,6]

Colonoscopy is the leading endoscopic technique when it comes to implementing artificial intelligence-based tools to optimize the procedure. However, no database consisting of the colonoscope's coordinates exists, allowing for a mapping with timestamps of the colonoscope path through the colon. The colonoscope contains coils that, through electromagnetic radiance, are translated into magnetic endoscopic imaging of the position while inside the patient, so the entire length of the colonoscope's position of the colonoscopy can be mapped. Such data have already been used to develop the colonoscopy retraction score, which correlates with the adenoma detection rate and the colonoscopy progression score, which correlates with pain experienced pain. Therefore, we provide a database consisting of 1400 clinical colonoscopies and 100 colonoscopies from a simulated setting. These data are freely available and could be used to map the mucosal inspection of the colon, generate heatmaps to ensure an equally distributed inspection, etc.

Background & Summary

In 2018, colorectal cancer (CRC) was diagnosed in approximately 1.3 million individuals worldwide and ranking as the third leading cause of cancer-related deaths[1]. Colonoscopy remains the premier method for identifying and treating precancerous conditions, significantly reducing mortality rates through effective screening initiatives[2]. The adenoma detection rate (ADR) serves as a critical measure of endoscopist competence, showing an inverse relationship with the incidence of post colonoscopy colorectal cancer (PCCRC)[3]. In particular, ADR exhibits considerable variability among endoscopists, ranging from 8.2% to 68.1% in the control group of randomised controlled trials (RCTs)[4]. These variabilities in endoscopist performance is concerning, with estimated adenoma miss rates of 26%[5]. To address these disparities, numerous initiatives have been launched to improve endoscopist's ADR[6], with advances in AI that have already been tested. Computer Aided Detection (CADe) that automatically detects polyps increases the ADR[7]. Computer Aided Quality assurance (CAQ), like a withdrawal speedometer that can alert the endoscopist when withdrawing too quickly, can also increase the ADR[8]. All these AIs try to complement the endoscopists with fully inspecting the mucosal wall to miss any adenomas. In colonoscopy, databases focus on image recognition[9,10], and the development of future AI-based systems rely on the availability, complexity and format of the data to serve as valuable training and validation resources[11]. The colonoscope contains magnetic coils that allow full traction of the scope within the patient's colon doing the procedure. Based on these coordinates, the CAQ systems colonoscopy retraction score (CoRS), has been developed and correlates to ADR[12] and the colonoscopy progression score (CoPS) correlates to patient experienced pain[13]. Using the colonoscope's coordinates to assess endoscopist's competence have therefore already been applied, but they also provide

[1]Copenhagen Academy for Medical Education and Simulation (CAMES) Rigshospitalet, 2100, Copenhagen, Denmark. [2]University of Copenhagen, Faculty of Health Sciences, Copenhagen, Denmark. [3]Department of Gastrointestinal and Hepatic Diseases, Copenhagen University Hospital - Herlev and Gentofte, Herlev, Denmark. [4]Danish Colorectal Cancer Screening Database (DCCSD) Steering Committee, Aarhus, Denmark. [5]Bispebjerg University Hospital, Copenhagen, Denmark. [6]University of Copenhagen, Faculty of Science, Department of Computer Science, Copenhagen, Denmark. ✉e-mail: kristoffer.mazanti.cold.01@regionh.dk

Databases	Colonoscopies	Type of Data
Simulation dataset (SDS)	100	Logfile of events.
Clinical dataset 1 (CDS1)	400	Progression and withdrawal time, ADR.
Clinical dataset 2 (CDS2)	1000	Logfile of events.

Table 1. Datasets. All colonoscopies contain coordinate datasets throughout the procedure. The data collection and patient demographics are described for CDS1 and CDS2 in the following studies:[12,17].

a path through the entire colon with the displacement of the individual coils helping in mapping the entire colon and understanding its flexible nature[14]. This study introduces the Copenhagen Colonoscopy Coordinate Database, a unique dataset designed to track and map the position of colonoscopes during both simulated and clinical procedures. The database comprises 100 simulated colonoscopies and 1400 clinical procedures, recorded using an electromagnetic colonoscope with positional tracking. The data collected includes timestamps and the 3D coordinates of the colonoscope, enabling precise mapping of its movement through the colon[15].

Technical validation confirmed the high fidelity of the coordinate data, which was processed to reduce noise and ensure smooth visualisation. Smoothing techniques and cubic spline interpolation were applied to refine the data, providing an accurate and continuous representation of the colonoscope's trajectory. The processed data maintain a high degree of spatial resolution, closely matching what the endoscopist in real time.

The database offers a variety of potential applications, including the development of quality assurance tools such as the Colonoscopy Retraction Score (CoRS) and Colonoscopy Progression Score (CoPS), which correlate with key clinical outcomes such as the adenoma detection rate and patient discomfort. Furthermore, the dataset provides a foundation for future research on colonoscopy performance, training, and the development of AI-based systems that could further optimise the effectiveness and safety of colonoscopy procedures. Open access to this dataset promotes further innovation in endoscopic technology and technique.

This study will establish a comprehensive database consisting of scope coordinates and timestamps through 100 simulated and 1400 clinical colonoscopy procedures.

Methods

To obtain the scope coordinates, a standardised setup was used for all data collection; a colonoscope (CFH180DL, CF-EX1500DL, Evis Exera II video centre CV-180 or Evis X1, manufacturer Olympus Medical System Ltd, Tokyo, Japan), the Magnetic Endoscopy Imaging (MEI) system called ScopeGuide (UPD-3, Olympus Medical System Ltd, Tokyo, Japan) and a computer interface to record colonoscope's coordinates, without a commercial name, we designate it the Olympus receiver box (UCES 3, Olympus Medical System Ltd, Tokyo, Japan). The receiver box was connected to a computer via USB cable, the coordinates we recorded using Olympus debugging software, wrapped in custom-made automation scripts using Python (Python Software Foundation. Python Language Reference, version 3 Available at http://www.python.org) and AutoIt (AutoIt, AutoIt Consulting Ltd, available at www.autoitscript.com/).

The database consists of three datasets: Table 1. Simulation dataset (SDS), 100 simulated colonoscopies[16]: The colonoscope coordinates were collected from 11 April to 19th of April 2024 in a simulated setting at Copenhagen Academy for Medical Education and Simulation (CAMES), Rigshospitalet Denmark. Four experienced colonoscopists (>500 colonoscopies) were included, each performed five colonoscopies in each of the phantom's first five cases (Kyoto Colonoscopy Training Model, Kyoto, Japan). The first author (KC) made a timestamp when the procedure started, at each flexure, when the cecum was reached, and finally when the colonoscopy ended (Fig. 1).

Clinical data set 1 (CDS1), 400 patient colonoscopies[16]: The colonoscope coordinates were collected through a prospective study[12] in 2021 with colonoscopies for colorectal cancer screening at three different sites in the capital region of Denmark. Only colonoscopies that reached the cecum were included due to availability. A blinded researcher calculated the progression and withdrawal time for each colonoscopy[12]. Additionally, the historical ADR of each operator was saved for each colonoscopy.

Clinical data set 2 (CDS2), 1000 clinical colonoscopies[16]: The colonoscope's coordinates were collected through screening colonoscopies for colorectal cancer through a cluster RCT (clinicaltrials.gov: NCT04862793) from 2022 to 2024 at three different sites in the Capital Region of Denmark. An on-site designated medical student registered the following timestamps: Beginning of procedure, left flexur, right flexur, cecum, ileum, retraction, polypectomies, patient rotation, and flushing (Fig. 1).

For both CDS1 and CDS2, the participation of patients and endoscopists was voluntary[16]. Informed consent was collected for participation and sharing of data. The CDS1 and CDS2 data was collected through the following studies[12,17], which both received regional ethical approval through the Capital Region of Denmark: journal no. H-17032652 and H-21019106, respectively[16].

Data Records

The dataset is available at Harvard Dataverse, available at https://doi.org/10.7910/DVN/F9B4SG[16].

As described in the methods section, the data repository consists of 100 procedures with colonoscope coordinates from a simulation setting (SDS)[16], 400 procedures with colonoscope coordinates from a clinical setting obtained through an observational trial (CDS1)[12] and 1000 procedures with colonoscope coordinates from a clinical setting obtained through a cluster RCT (CDS2)[16,17] - see methods for more details. Both clinical trials received local ethics approval. The repository does not contain patient-sensitive details. In addition to the

```
Synchronized relative time;event;
-25;Logging started;
-23;Start position is on the left;
29;Endoscopy started;
348;Fleksur L.;
408;Rotate to back;
475;Transverse;
876;Fleksur D.;
1041;Polyp;
1082;Rotate to left;
1565;Rotate to back;
1632;Cecum;
1746;Retraction;
1749;Cecum;
1758;Biopsy;
1812;Biopsy;
1918;Flush;
2040;Fleksur D.;
2077;Flush;
2097;Rotate to left;
2245;Polypectomi;
2329;Flush;
2465;Transverse;
2483;Flush;
2536;Fleksur L.;
2685;Polypectomi;
2755;Flush;
2790;Flush;
2910;Flush;
2911;Retroflex;
2981;Recording ended;
2994;End reason - Completed;
```

Fig. 1 Logfile with timestamps in seconds of events doing the colonoscopy.

endoscope coordinates, each of the datasets contains additional information explained in the Methods section. Examples of the data and levels of details in space and time can be seen in Figs. 2–4.

The coordinate data is stored in CSV files, and procedure meta-data, stored in timestamped text-files. Coordinate data consists of 4 columns, timestamp (T), X,Y, and Z coordinates, where XYZ are repeated along the interpolated points of the endoscope. The metadata file, consists of a timestamp column and event column, where the event represent the live annotated completed during the procedure. Files are archived in a split zipped directory (*.7z), so all files are required to be downloaded in order to be able to unzip.

Each CDS2 procedure includes a logfile documenting key procedural timestamps, including anatomical landmarks and procedure end reasons (e.g., completed satisfactorily, incomplete due to poor bowel preparation or patient discomfort)[16]. These help contextualize anatomical variation and procedural outcomes, offering indirect insights into factors such as colonic redundancy or surgical alterations. Total registration of events is available in Table 2.

While patient demographics (e.g., age, gender, bowel preparation quality) are not included in the shared metadata, each procedure for CDS2 includes a logfile with detailed timestamps of anatomical landmarks and procedural events, which can help infer lesion location and procedural context. Full study details with patient demographics are available in the original publications[12,16,17].

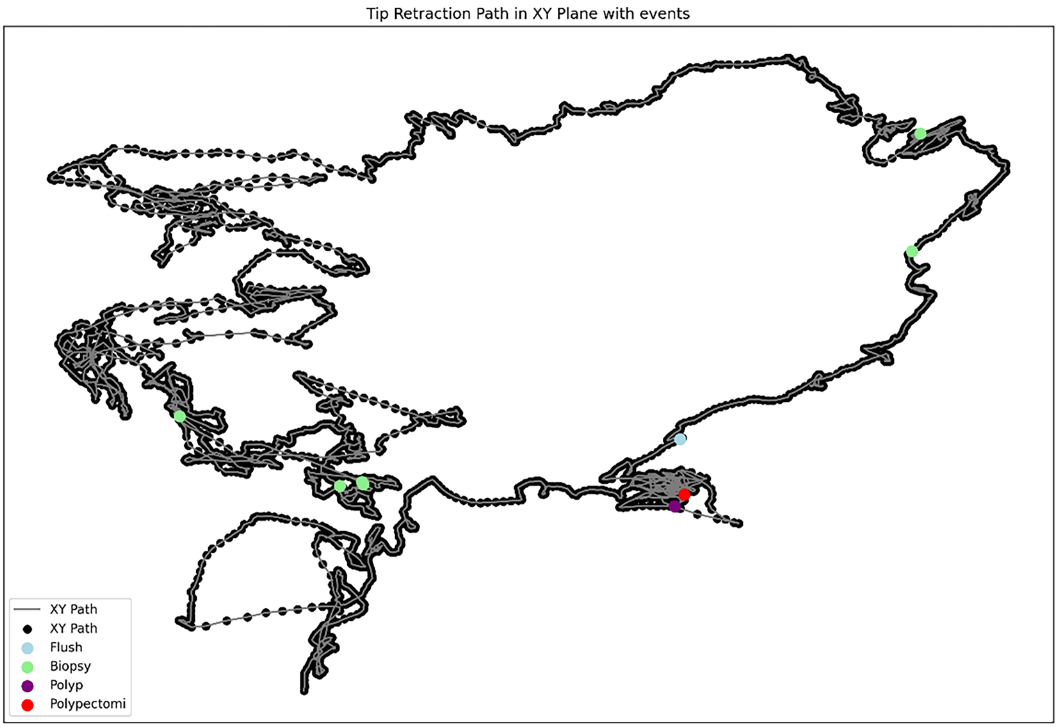

Fig. 2 Example of the retraction phase of a colonoscopy from the CDS2 dataset. The -o- line represent the path of the endoscope tip, and the coloured circle are different events logged during the procedure, e.g. polyp (visual), polypectomy, biopsy, and flushing. File name: MwrAUtS601f3_xy_tip_path.

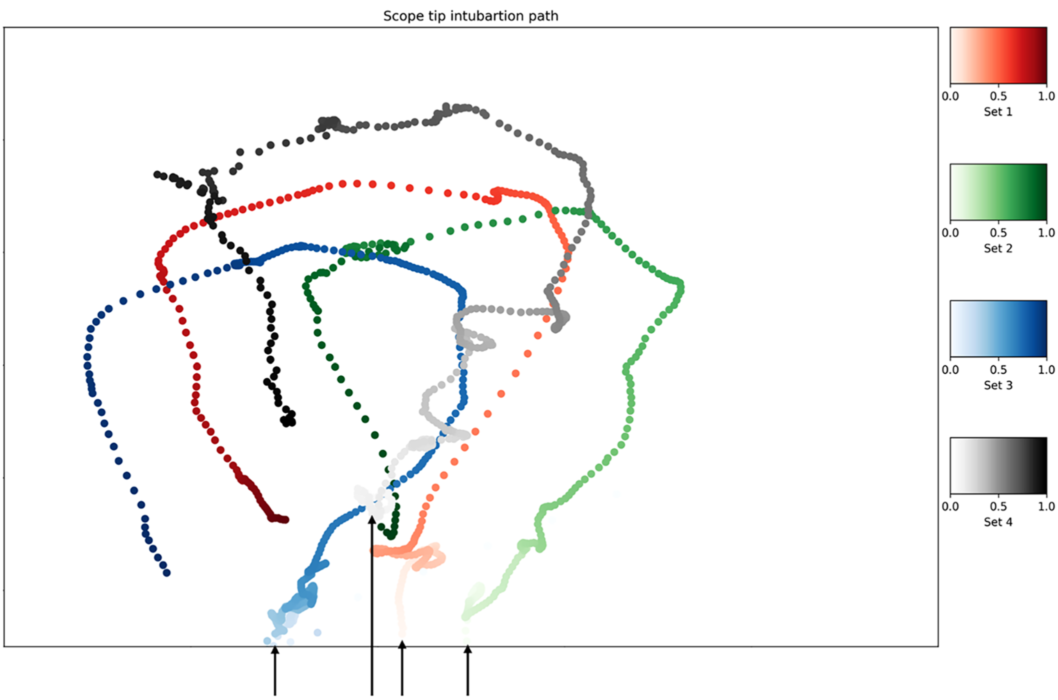

Fig. 3 Examples of the four endoscopists tip progression path recorded on case #2 of the Kyoto Colonoscopy Training Model. The data is available through the Simulation Data Set (SDS). Note the difference in the beginning position, a consequence of model or antenna displacement between persons (arrows). (File name: 2_4_tip_path).

Technical Validation

To demonstrate the technical quality of the dataset, we have demonstrated some of its potential uses and applications in previous studies[12,13,15,17–20]. Similarly, we have attempted to make some descriptive experiments on the data.

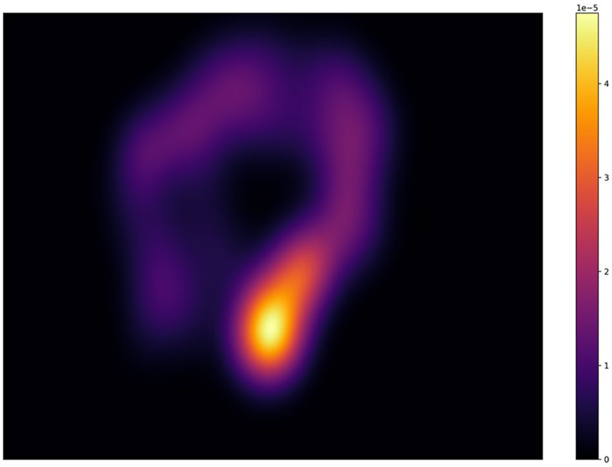

Fig. 4 Two-dimensional heatmap of SDS dataset case 2. Positions normalized to the tip at the first frame. Brighter colours indicate more frequent observations, and thus less temporal variation in scope tip presence. The data is available through the Simulation Data Set (SDS), file name: 1_2_kde.

Event	Total count	Procedures containing event	Mean event count ± sd	Percentage of files containing event
Flush	3661	506	5.7201 ± 7.7013	79.0625
Biopsy	284	61	0.4438 ± 1.8137	9.5312
Polyp	364	164	0.5688 ± 1.6582	25.625
Polypectomy	836	331	1.3062 ± 2.1399	51.7188
Retroflexion	516	493	0.8061 ± 0.4845	77.0312

Table 2. Cumulative events registered through the logfile in CDS2) every time the endoscopists perform a flush, biopsy, spot a polyp, perform a polypectomy or retroflex.

Recording and Post-processing of data. *Synchronisation to logfiles.* The coordinate data files were synchronised to logfiles, by the end time in the logfile (Figs. 1, 2). This time point indicates when the user has pressed the end recording, and the script-based control of the third-party software did not have any variation in the end time. As opposed to the start-time, which could vary with 60 s due to third-party software responsiveness. The 60 s arise from our own software, detecting whether recording was successful every 60 s, and restarting third-party software if not successful. It should be noted that the recording process was made so that the software logging of the coordinates started before the actual endoscopy.

Post-processing of recorded data. Utilizing processed data instead of raw sensor data is a well-established practice in many technical fields, including medical imaging and signal processing. Raw sensor data, while rich in detail, often contain noise, artifacts, and irrelevant variations that can obscure the true signal and complicate the analysis. In our case, reading out raw data into a graph does not provide the same visualisation as shown to the endoscopists via the manufacturer's user interface. By applying signal processing techniques, such as smoothing and interpolation, we can significantly enhance the quality and usability of the data.

Smoothing, as implemented using a moving average filter, is used to reduce motion artefacts and high-frequency noise introduced by sensor imperfections or environmental conditions. The smoothing algorithm applies a running average to the data using a specified window size to compute the average across the data points. In this case, a window size of 10 was chosen based on empirical testing (see running_average() in the script), balancing the need to remove noise without distorting the signal.

Interpolation of the 3D coordinates was employed using cubic-spline interpolation. This method ensures that we maintain a smooth and high-resolution representation of the endoscope's trajectory. The interpolate_3d_coords() function generates 100 interpolated points along the XYZ axes within the first metre of the endoscope. This provides a detailed and consistent spatial representation of the endoscope, closely matching the data visualised on the endoscopist's interface. Importantly, the interpolation not only enhances the data's spatial resolution but also obscures exact sensor locations, adding a layer of protection for proprietary information.

The processed data undergoes the following steps:

- **Smoothing** to reduce jitter and irrelevant movements from the sensor data (as captured in averager_coils()).
- **Interpolation** of the coordinates using cubic-spline fitting to create a consistent, smooth path representing the endoscope's position (interpolate_3d_coords()).
- **Timestamp** conversion to maintain privacy and security by converting absolute timestamps to relative time.

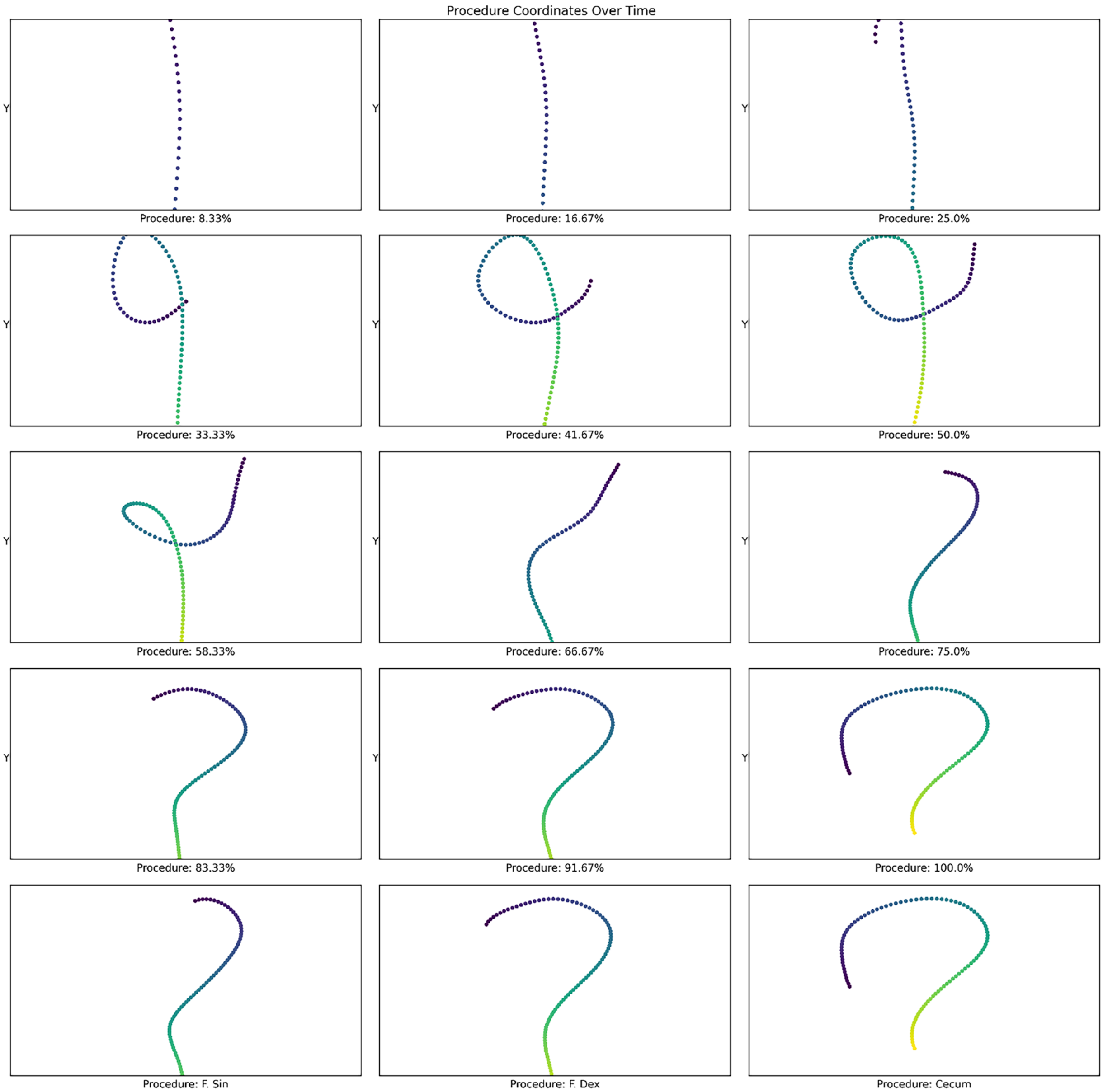

Fig. 5 Visualization of the entire scope throughout one procedure and at the three anatomical landmarks (right flexure, left flexure and cecum, using the logfile) through the progression of an experienced endoscopist on case #3 on the Kyoto Colonoscopy Model. The progression shows the endoscopist straightening an alpha-helix (33% to 65%). The data is available through the Simulation Data Set (SDS), File name: kUzqgqjDFnQd_time_plot.

Together, these steps improve the interpretability of the data and make them suitable for further computational analysis and research, while protecting sensitive and proprietary information.

Timestamp conversion to relative time. To maintain privacy and protect proprietary information, absolute timestamps in the recorded data are converted to relative time. This process begins by identifying the first timestamp in the data (t_0) and using it as a reference point. Each subsequent timestamp is then converted to a relative interval based on the time difference from t_0. This conversion is handled by the calculate_relative_time() function, which parses the time from strings, calculates the time differences, and returns them in milliseconds.

Smoothing time to remove sensor artifacts. Recorded sensor data often includes high-frequency noise and motion artefacts caused by factors such as environmental disturbances or sensor imperfections. To address this, we applied a moving average filter using the running_average() function. This filter smooths out the data by averaging values within a specific window, reducing jitter and irrelevant movements that appear in the sensor data but are not part of the endoscope's actual motion.

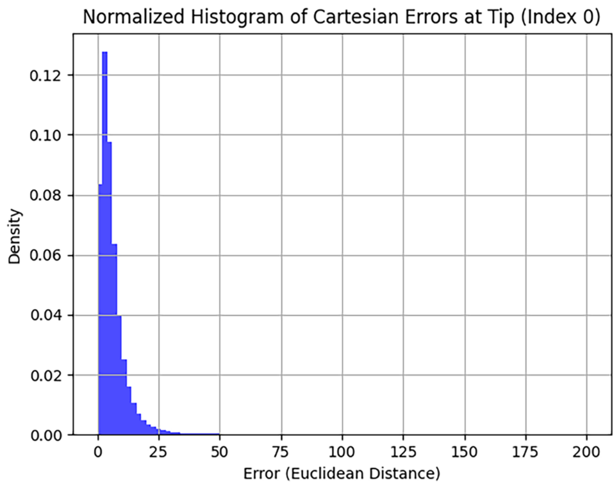

Fig. 6 Post processing deviation of tip coordinates (Praw – Pprocessed), of 400899 time points, from 50 endoscopies.

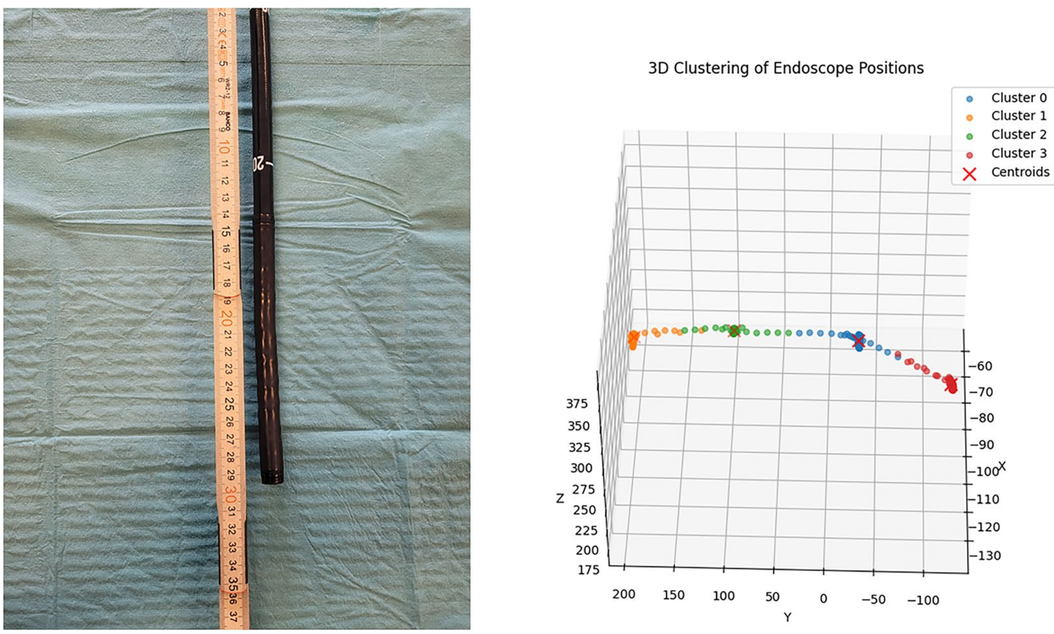

Fig. 7 Setup and determination of approximate coordinate to mm.

For this purpose, a window size of 10 was selected, which provides a balance between noise reduction and signal preservation. Smoothed data produce a more accurate representation of the position of the endoscope, making it more suitable for downstream analysis and matching the display shown to the endoscopist.

Interpolation of Endoscope Coordinates. To generate a detailed and consistent model of the path of the endoscope, cubic spline interpolation is applied to the raw 3D coordinates using the interpolate_3d_coords() function. This function takes the XYZ coordinates of the endoscope and generates a spline with a specified number of points (100 in this case) that provides a smooth representation of the endoscope's path Fig. 5. The interpolation results in a high-resolution dataset that closely mirrors the display seen by the endoscopist on their interface.

Furthermore, this interpolation process removes the exact locations of the sensors along the endoscope, protecting proprietary information related to the sensor configuration. Since the published data is based on interpolated points rather than raw sensor data, it is impossible to reverse-engineer the exact sensor positions, ensuring that the sensitive details of the system remain confidential while maintaining the accuracy and utility of the data for research purposes Fig. 5.

Technical validation, metric. Goal: Representing reality after post-processing. The deviation from raw to processed coordinates for 400000 + time points, and 50 colonoscopies, can be seen in Fig. 6, the peak

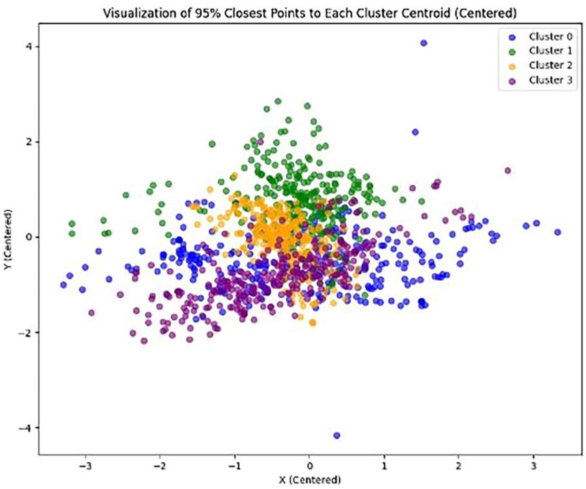

Fig. 8 Variation in time of the tip position for an endoscope laying on a table, after post-processing. Each cluster contains the 95% closest points the the cluster center to avoid including data from when moving the endoscope tip to each position.

```
Coordinate_data
├── Colo_Plot_and_scripts
│         ├── Multi_tip_case_plots
│         ├── Procedure_over_time_plots
│         ├── Tip_path_event_plot
│         ├── plot_scripts.py
│         ├── multi_case_tip_plot.py
│         ├── procedure_over_time.py
│         ├── tip_path_w_event.py
│         └── event_table.png
├── Processed_delivery
│         ├── ClinBatch1
│         │         ├── Procedure_ID_1
│         │         │         ├── animation.mp4
│         │         │         ├── LogFile_P.txt
│         │         │         ├── LogFile_T_sync.txt
│         │         │         ├── T-,X-,Y-,Z.txt
│         │         │         ├── T-,X-,Y-,Z.json
│         │         │         └── T-,X-,Y-,Z.npy
│         │         ├── ...
│         │         └── Procedure_ID_n
│         ├── ClinBatch2
│         │         ├── Procedure_ID_1
│         │         │         ├── animation.mp4
│         │         │         ├── LogFile_P.txt
│         │         │         ├── LogFile_T_sync.txt
│         │         │         ├── T-,X-,Y-,Z.txt
│         │         │         ├── T-,X-,Y-,Z.json
│         │         │         └── T-,X-,Y-,Z.npy
│         │         ├── ...
│         │         └── Procedure_ID_n
│         └── SimBatch
│                   ├── Procedure_ID_1
│                   ├── ...
│                   └── Procedure_ID_n
├── CDS1_meta.txt
├── CDS2_meta.txt
├── SDS_meta.txt
```

Fig. 9 Calibrated endoscopy procedures. File and directory tree of the zipped dataset, version 1, including scripts. Plots and scripts folder, contains workable examples, "Processed delivery" contains the actual datasets (CDS1, CDS2, and SDS), under each is a folder per procedure. Metadata for each dataset is found at the root level.

resides at 2, roughly corresponding to a deviation of 1.1 mm (see below for coordinate-to-metric conversion). To assess how much variation there was over time and to get a conversion ratio of pixels to mm (which is not available to us in documentation). We constructed a small experiment (Figs. 7, 8) where we placed an endoscope on a table and moved it 10 cm at a time, having it lie at each position for 60 seconds.

The coordinates variations in time, for clusters centred to origo, the Overall 95% confidence intervals were for **X-axes**: 95% CI(−0.325, −0.220), Mean = −0.272, SD = 0.898, **Y axes**: 95% CI (−0.193, −0.093), Mean = −0.143, SD = 0.864 and **Z axes** 95% CI (0.667, 0.791), Mean = 0.729, SD = 1.062.

The average euclidean distance between cluster centres in 3D were 176.21 pixels, giving a conversion factor of coordinates to millimetres of 1.7621 coordinates/mm. The temporal variation of data, electromagnetic noise, and the difference of these for the different sensor location throughout the endoscope may not be applicable for a global conversion of coordinates to millimetres. The ScopeGuide system used is CE-marked and designed for intraintestinal use, where fluids are abundant; we have not observed signal artefacts attributable to intestinal fluids. Metallic implants such as hip prostheses are typically non-magnetic and have not presented interference in our experience. The most relevant interference risk arises when the external antenna is physically obstructed or displaced during a procedure, which unfortunately was not systematically logged."

We would like to clarify that only SDS was obtained in a non-clinical setting, while both CDS and CDS2 were collected during real-world colonoscopy procedures on patients. The table-based experiment was not intended as a substitute for *in vivo* validation, but rather to quantify sensor jitter and demonstrate the data's temporal fidelity. Regarding the effects of dynamic clinical factors such as peristalsis and operator technique, these are considered integral parts of the actual signal and not sources of measurement error. Therefore, further *in vivo* validation would not only be ethically and logistically complex, but may also be of limited additional value, given a CE-marked medical device was used, which reassures us about its accuracy and reliability in clinical practice

Usage Notes

The most present database and code for processing is available via the Github platform (https://github.com/CAMES-Engineering/Colonoscopy-Coordinates, with a specific release of the code available to the reviewers… CAMES-Engineering/OSABPS/releases/tag/review) and arranged as per Fig. 9.

Code availability

The dataset can be used without any further code, but example usages, and code for the examples in this publication is provided at the Github repository.

Received: 27 November 2024; Accepted: 3 July 2025;
Published online: 10 July 2025

References

1. Bray, F. *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* **74**(3), 229–63 (2024).
2. Bretthauer, M. *et al*. Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death. *New England Journal of Medicine.* **387**(17), 1547–56 (2022).
3. Corley, D. A. *et al*. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med.* **370**(14), 1298–306 (2014).
4. Hassan, C. *et al*. Variability in Adenoma Detection Rate in Control Groups of Randomized Colonoscopy Trials. *Gastrointestinal Endoscopy*. (2022).
5. Zhao, S. *et al*. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. *Gastroenterology.* **156**(6), 1661–74.e11 (2019).
6. Rex, D. K. *et al*. Quality indicators for colonoscopy. *Gastrointest Endosc.* **100**(3), 352–81 (2024).
7. Hassan, C. *et al*. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc.* **93**(1), 77–85.e6 (2021).
8. Cold, K. M. *et al*. Computer-aided quality assessment of endoscopist competence during colonoscopy: A systematic review. *Gastrointestinal Endoscopy*. (2024).
9. Borgli, H. *et al*. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data.* **7**(1), 283 (2020).
10. Azagra, P. *et al*. Endomapper dataset of complete calibrated endoscopy procedures. *Scientific Data.* **10**(1), 671 (2023).
11. Wang, Z. *et al*. Improving Foundation Model for Endoscopy Video Analysis via Representation Learning on Long Sequences. *IEEE J Biomed Health Inform* **29**(5), 3526-3536, https://doi.org/10.1109/JBHI.2025.3532311 (2025).
12. Vilmann, A. S. *et al*. Colonoscope retraction technique and predicting adenoma detection rate: a multicenter study. *Gastrointestinal endoscopy*. (2022).
13. Preisler, L., Bulut, M., Svendsen, M. S., Svendsen, L. B. & Konge, L. An automatic measure of progression during colonoscopy correlates to patient experienced pain. *Scand J Gastroenterol.* **53**(3), 345–9 (2018).
14. Biffi, C. *et al*. REAL-Colon: A dataset for developing real-world AI applications in colonoscopy. Scientific. *Data.* **11**(1), 539 (2024).
15. Vilmann, A. S. *et al*. Using computerized assessment in simulated colonoscopy: a validation study. *Endosc Int Open.* **8**(6), E783–e91 (2020).
16. Cold, K. M. Colonoscopy-Coordinates. V1. https://doi.org/10.7910/DVN/F9B4SG. ed: Harvard Dataverse; Original Release Date: 2025-05-25 23:45:29.541.
17. Cold, K. M. *et al*. Is the Transverse Colon Overlooked? Establishing a Comprehensive Colonoscopy Database from a Multicenter Cluster-Randomized Controlled Trial. *Diagnostics.* **15**(5), 591 (2025).
18. Vilmann, A. S. *et al*. Expertise in colonoscopy intubation does not predict diagnostic accuracy: a simulation-based study. *Endosc Int Open.* **10**(1), E30–e6 (2022).
19. Nerup, N., Preisler, L., Svendsen, M. B., Svendsen, L. B. & Konge, L. Assessment of colonoscopy by use of magnetic endoscopic imaging: design and validation of an automated tool. *Gastrointest Endosc.* **81**(3), 548–54 (2015).
20. Cold, K. M. *et al*. Development and validation of the Open-Source Automatic Bowel Preparation Scale. *Gastrointestinal Endoscopy*. (2024).

Acknowledgements

The data collection was supported by the Danish Cancer Society, The Danish Cancer Research Foundation, The Vissing Foundation, Aase and Ejnar Danielsens Foundation and Intelligent Robotic Endoscopy (IRE) project, financed by the European Union Horizon Grant (IRE - 101135082 - GAP-101135082). Funded by the European Union.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.M.C.