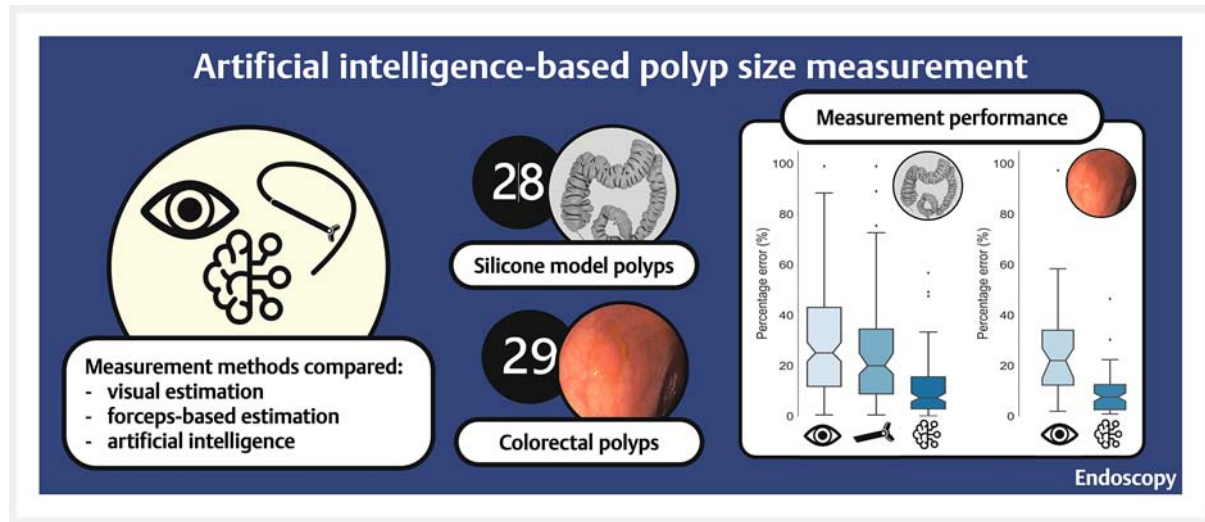


Artificial intelligence-based polyp size measurement in gastrointestinal endoscopy using the auxiliary waterjet as a reference





OPEN
ACCESS



GRAPHICAL ABSTRACT



Authors

Boban Sudarevic^{1,2}, Philipp Sodmann¹, Ioannis Kafetzis¹, Joel Troya¹ , Thomas J. Lux¹, Zita Saßmannshausen¹, Katja Herlod¹, Stefan A. Schmidt³ , Markus Brand¹ , Katrin Schöttker¹, Wolfram G. Zoller², Alexander Meining¹, Alexander Hann¹ 

Institutions


- 1 Interventional and Experimental Endoscopy (InExEn), Internal Medicine II, University Hospital Würzburg, Würzburg, Germany
- 2 Department of Internal Medicine and Gastroenterology, Katharinenhospital, Stuttgart, Germany
- 3 Department of Diagnostic and Interventional Radiology, University Hospital Ulm, Ulm, Germany

submitted 3.1.2023

accepted after revision 19.4.2023

accepted manuscript online 20.4.2023

published online 12.7.2023

 Fig. 1 s–5 s, Appendix 1 s–64s, Tables 1 s–3 s, Supplementary material is available under <https://doi.org/10.1055/a-2077-7398>

Bibliography

Endoscopy 2023; 55: 871–876

DOI 10.1055/a-2077-7398

ISSN 0013-726X

© 2023. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Corresponding author

Alexander Hann, MD, Universitätsklinikum Würzburg, Medizinische Klinik und Poliklinik II, Oberdürrbacher Str. 6, 97080 Würzburg, Germany
hann_a@ukw.de

ABSTRACT

Background Measurement of colorectal polyp size during endoscopy is mainly performed visually. In this work, we propose a novel polyp size measurement system (Poseidon) based on artificial intelligence (AI) using the auxiliary waterjet as a measurement reference.

Methods Visual estimation, biopsy forceps-based estimation, and Poseidon were compared using a computed tomography colonography-based silicone model with 28 polyps of defined sizes. Four experienced gastroenterologists estimated polyp sizes visually and with biopsy forceps. Furthermore, the gastroenterologists recorded images of each polyp with the waterjet in proximity for the application of Poseidon. Additionally, Poseidon's measurements of

29 colorectal polyps during routine clinical practice were compared with visual estimates.

Results In the silicone model, visual estimation had the largest median percentage error of 25.1% (95%CI 19.1%–30.4%), followed by biopsy forceps-based estimation: median 20.0% (95%CI 14.4%–25.6%). Poseidon gave a significantly lower median percentage error of 7.4% (95%CI 5.0%–9.4%) compared with other methods. During routine colonoscopies, Poseidon presented a significantly lower median percentage error (7.7%, 95%CI 6.1%–9.3%) than visual estimation (22.1%, 95%CI 15.1%–26.9%).

Conclusion In this work, we present a novel AI-based method for measuring colorectal polyp size with significantly higher accuracy than other common sizing methods.

Introduction

Colorectal cancer (CRC) is the third most frequently diagnosed cancer and the second most common cause of cancer-related death worldwide [1]. Polyp size directly correlates to the risk of developing CRC, as larger polyps tend to have a higher risk of becoming cancerous than smaller ones [2]. Therefore, in current guidelines for preventing CRC, polyp size is one of the critical factors for managing colonoscopy surveillance intervals and selecting an adequate resection method [3, 4].

Currently, the commonest method for sizing polyps is visual estimation performed by an examiner during colonoscopy. Several studies have demonstrated that visual estimation is unreliable, as it exhibits high intraobserver variability and poor accuracy [5]. This often leads to incorrect colonoscopy surveillance intervals being determined [5, 6]. Polyp size measured by the pathologist after resection is frequently used as a gold standard in studies [7, 8]; however, post-polypectomy measurement can introduce inaccuracies, as the specimen experiences trauma and shrinkage owing to the resection and fixation processes. Therefore, developing an accurate in situ method for polyp sizing has attracted considerable research attention over the past years.

Numerous methods have been investigated to improve the accuracy of polyp sizing. Device-based methods such as polypectomy instruments [9], add-on caps [10], structured light, and virtual scales [11] have been explored. Despite improved measurement accuracy, widespread adoption into clinical practice is however lacking as such methods require specific endoscopic systems or additional devices that are not routinely used.

Artificial intelligence (AI) was introduced in clinical practice several years ago, with the main focus being polyp detection [12, 13]; however, AI-based diagnosis has been further explored for polyp characterization [14] and other gastroenterological diseases, such as eosinophilic esophagitis [15] and Crohn's disease [12]. AI has the potential to solve the issue of polyp sizing, and several AI-based concepts have recently been described [16–18]. Nonetheless, polyp size measurement sup-

ported by AI is in its early stages, and further development is necessary to explore its potential.

In our work, we developed a polyp sizing method to address the shortcomings of the current technologies. The auxiliary waterjet, commonly used for cleaning mucosal surfaces during colonoscopy, was selected as a measurement reference. Our AI-based measurement system “Poseidon” first identifies the location of the waterjet and the polyp in the image and then calculates its size. We compared Poseidon's accuracy to visual and biopsy forceps-based estimations in a computed tomography colonography (CTC)-derived colon that was 3D printed and cast using silicone. Additionally, we compared Poseidon with visual estimation during routine clinical practice.

Methods

Polyp sizing in the silicone-based colon

Four experienced board-certified gastroenterologists performed a colonoscopic withdrawal in an artificial silicone-based colon model, which replicated an anonymized 3D model generated by CTC (**Fig. 1 s**, see online-only Supplementary material). Details of the process used to manufacture the silicone-based colon are described in **Appendix 1 s**. The silicone-based model included 28 polyps (summarized in ► **Table 1**, detailed in **Table 1 s**) that differed in their size and Paris classification. The gold-standard size in millimeters was established by manually measuring the polyps with digital calipers and rounding the measurements to two decimal places.

The four endoscopists inspected each polyp, estimating its size first visually and then with opened biopsy forceps (Radial Jaw 4, standard capacity; Boston Scientific, Marlborough, Massachusetts, USA). They then aimed the waterjet at the area adjacent to the polyp and obtained three images per polyp (**Fig. 2 s**). As a new feature of our polyp detection system, “EndoMind”, Poseidon analyzed the images containing a waterjet to determine the polyp size [19]. The final size of each polyp was calculated as the average of the determined sizes, but this was not presented to the endoscopists during the experiments.

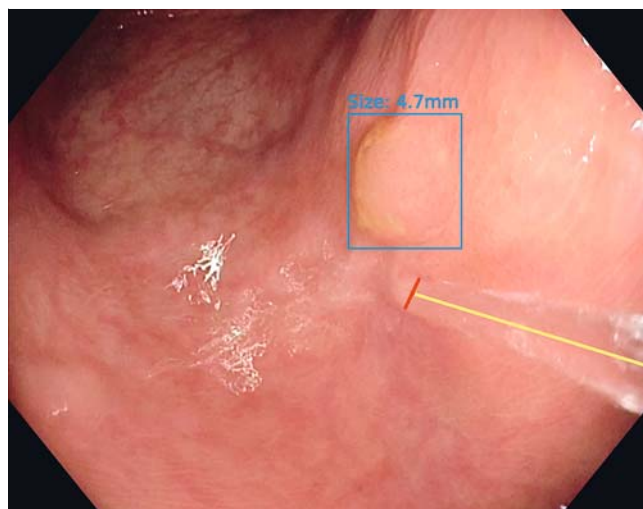
► **Table 1** Baseline characteristics of the 28 polyps in the silicone colon model.

Characteristic	n (%)
Size, mm	
▪ <5	14 (50.0)
▪ 5–10	7 (25.0)
▪ >10	7 (25.0)
Paris classification	
▪ 0-Ip	4 (14.3)
▪ 0-Is	12 (42.9)
▪ 0-IIa	12 (42.9)

An Olympus CF-HQ190L colonoscope and EVIS EXERA III endoscopic system (Olympus Europa SE&Co. KG, Hamburg, Germany) coupled with Wieser Jet-Cleaner III pump (Wieser Medizintechnik GmbH, Egenhofen, Germany) were used. Beforehand, several experiments were conducted with the particular endoscopic system to evaluate the effect of gravity and the stability of the waterjet. A description of the experimental setup and a summary of the results is given in **Appendix 2s** (**Figs. 3s** and **4s**).

Polyp sizing during routine clinical practice

We further evaluated the accuracy of Poseidon during routine clinical practice between October 13 and November 24, 2022 (► **Video 1**). Poseidon analyzed images containing the waterjet adjacent to the polyp to determine its size. Visual estimations of polyp size were collected from examination reports. To establish a reference standard, we manually measured and com-



► **Fig. 1** Endoscopic image of a polyp automatically outlined by a bounding box with a waterjet adjacent. The yellow line represents the length of the waterjet, while the red one represents its diameter, which is used as the measurement reference. The output of Poseidon, with the size estimation displayed above the bounding box, was not presented to the examiner during the study.

pared polyps with polypectomy instruments proximal to each other on the endoscopic images (**Appendix 3s**).

Polyp sizing system “Poseidon”

Poseidon is based on a combination of two AI algorithms (► **Fig. 1**). The first one is the previously described real-time polyp detection system “EndoMind”, which outlines a polyp with a bounding box [19]. The second AI algorithm localizes the waterjet in the image and determines its diameter. The polyp size is then calculated by comparing the diameter of the waterjet to the longest bounding box side around the polyp. A detailed description of the algorithm is provided in **Appendix 4s**.

Statistical analysis

Data analysis was performed using Python (Python Software Foundation, Wilmington, Delaware, USA) combined with the NumPy, Pandas, SciPy, and moepy libraries. As a measure of agreement between endoscopists, the two-way mixed intra-class correlation coefficient (ICC3) was calculated for each sizing method.

The measurement error was calculated for every result of the sizing methods by subtracting the corresponding gold standard from the result. The percentage error was calculated as the absolute value of measurement error/gold standard × 100% to represent the magnitude of mis-sizing. Additionally, a 95%CI for medians was estimated using bootstrapping with 10 000 samples, as a measure of precision.

The sizing methods were further evaluated in the subgroup analysis, where polyps <5 mm were regarded as diminutive, 5–10 mm as small, and >10 mm as large. The Wilcoxon signed rank test was used with a *P* value of 0.05 as a threshold to determine if there were significant differences between the sizing methods. A significant difference and a lower median percen-

The image shows a polypoid lesion on the colonic mucosa. A blue rectangular bounding box is drawn around the polyp. A QR code is located in the top right corner of the image area.

► **Video 1** Video clip showing the performance of Poseidon in routine clinical practice.
 Online content viewable at:
<https://doi.org/10.1055/a-2077-7398>

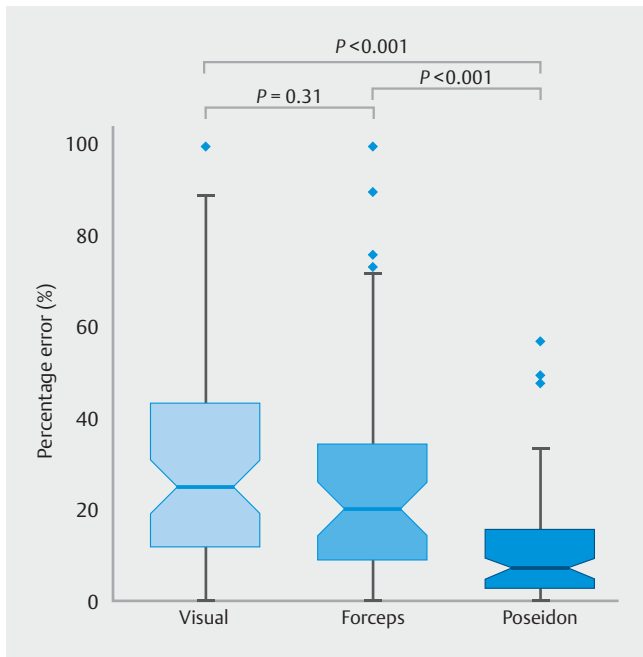


Fig. 2 Box plots of percentage errors for each sizing method. The notch around the median value represents the 95%CI derived using bootstrapping (n = 10 000).

age error were used to determine that one method was more accurate than the other.

Ethics

The recording of videos during routine clinical examinations to test AI algorithms was approved by the ethical committee of the University Hospital Würzburg (12/20, 20200114–04). Signed informed consent was obtained from each patient before the examination was recorded.

Availability of the algorithm

The Poseidon algorithm and installation instructions will be freely available for research purposes at: <https://www.ukw.de/research/inexen/applied-ai/poseidon-ai-based-polyp-size-estimation/>.

Results

Performance of polyp sizing in the colon model

Each of the four endoscopists detected all 28 polyps in the silicone colon model. They assessed the size of each polyp visually and with the biopsy forceps, and then captured three images including the waterjet. In one case, one of the endoscopists could not fully capture the given polyp within the endoscopic field of view. This case was not included in the analysis, leaving 111 measurements by each sizing method in total. The agreement between endoscopists was 0.82 (95%CI 0.71–0.90) for visual estimation, 0.84 (95%CI 0.75–0.92) for forceps-based measurement, and 0.94 (95%CI 0.90–0.97) when using Poseidon.

As presented in **Fig. 2**, the biopsy forceps-based estimation was significantly more accurate (median percentage error 20.0%, 95%CI 14.4%–25.6%) than the visual estimation (median percentage error 25.1%, 95%CI 19.1%–30.4%; $P=0.03$), while Poseidon was significantly more accurate than both, with a median percentage error of 7.4% (95%CI 5.0%–9.4%). Furthermore, Poseidon gave a significantly lower percentage error ($P<0.001$) than the other two methods for each of the polyp size subgroups (**Table 2**).

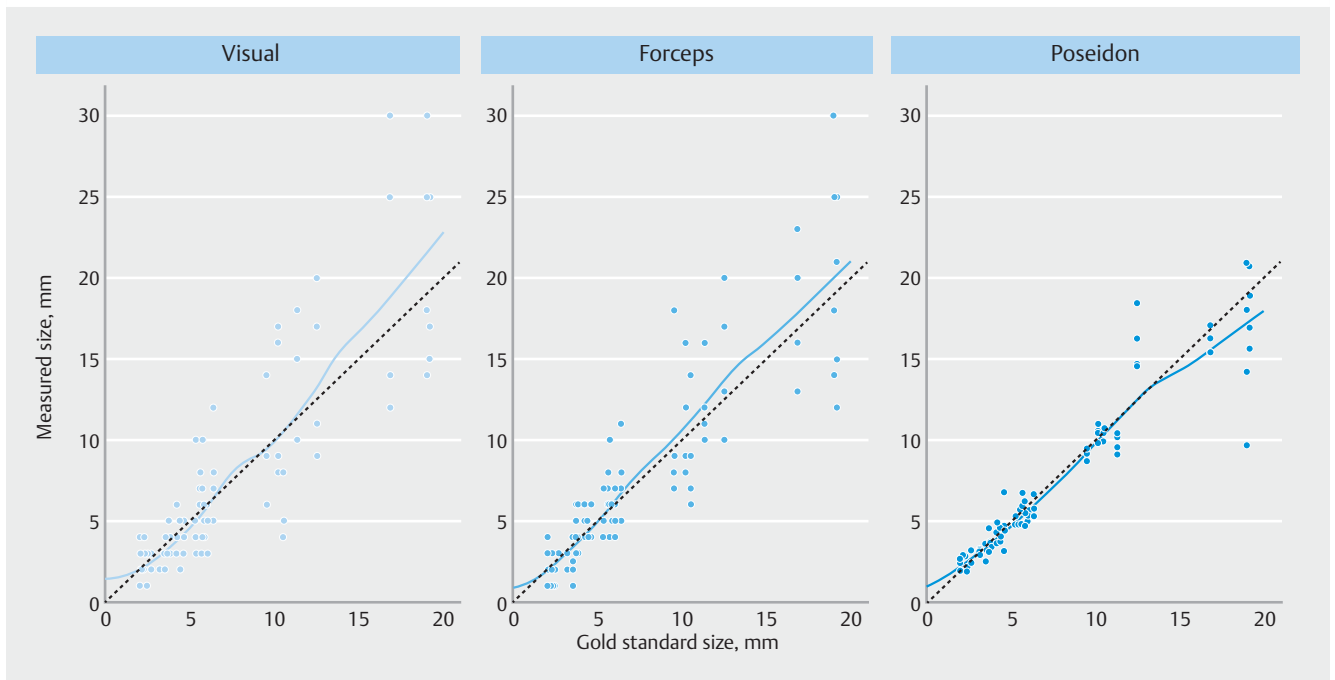
Visual and biopsy forceps-based estimation tended to overestimate polyp size, while Poseidon showed a tendency to underestimate for large polyps (**Fig. 3**). The results of a binary classification of the polyps into size classes using each sizing method are shown in **Fig. 5s**.

Performance of polyp sizing during routine clinical practice

A total of 29 polyps were included from 17 examinations performed by three experienced endoscopists. Most of the polyps were adenomas (69%) ranging in size from 2.5 mm to 13.7 mm (**Tables 2s** and **3s**). Poseidon was significantly more accurate (median percentage error 7.7%, 95%CI 6.1%–9.3%) than visual estimation of polyp size (median percentage error 22.1%, 95%CI 15.1%–26.9%), as shown in **Table 2**.

Table 2 Comparison of the three sizing methods for all polyps and for the different polyp size subgroups (all results given as median percentage error [95%CI]).

	Diminutive	Small	Large	Overall
Silicone colon				
Visual	19.1 (14.0–30.4)	25.0 (14.4–37.0)	30.3 (22.8–50.4)	25.1 (19.1–30.4)
Forceps	16.7 (13.0–29.9)	19.3 (12.1–26.0)	22.3 (16.0–33.3)	20.0 (14.4–25.6)
Poseidon	7.3 (4.7–9.7)	5.3 (3.6–11.0)	8.2 (3.8–13.8)	7.4 (5.0–9.4)
Routine clinical practice				
Visual	20.8 (11.5–24.2)	38.2 (15.1–45.6)	23.4 (19.8–26.9)	22.1 (15.1–26.9)
Poseidon	8.1 (2.7–13.2)	6.3 (3.5–10.0)	9.0 (2.2–15.7)	7.7 (6.1–9.3)



► **Fig. 3** Scatter plots of measurement results plotted against the corresponding gold standard sizes for each sizing method. The dashed lines represent exact values, with the points closer to the lines being more accurate measurements. Points above the dashed line are overestimates, while those below underestimate the polyp size. The fitted curves show the over- or underestimation tendency of each sizing method.

Discussion

Knowing the accurate size of a polyp is essential for determining the appropriate follow-up interval after resection [3, 4]. Additionally, the resection technique used depends on this estimation [20]. An accurate and easy-to-use method is needed in routine clinical practice. Although visual estimation is currently the most practical method for addressing polyp size, it has been shown to lack accuracy [5].

Several previous studies have reported the percentage error as a metric of accuracy because it indicates the magnitude of mis-sizing, an equal measurement error having a greater impact on smaller polyps than on larger ones. In a work by Chapini et al. polyp size was visually overestimated in 32% of cases and underestimated in 20% by more than 20% [5]. In another study by Eichenseer et al. [6], 62.6% of polyps were mis-sized by more than 33%. Our findings are comparable, as visual estimation had a median percentage error of 25.1% in the colon model and 22.1% during regular clinical practice.

Biopsy forceps-based estimation was proposed as a way to improve the accuracy of visual estimation [9]. This is supported by our study, as the biopsy forceps-based estimation significantly reduced the median percentage error from 25.1% to 20.0% ($P=0.03$). Although forceps-based estimation does present an improvement over visual estimation, its accuracy is still suboptimal. Furthermore, a large proportion of polyps are currently resected using cold snares without the need for biopsy forceps [20], and instrument exchange may be time-consuming.

In recent years, interest in AI has been rapidly expanding in gastroenterology to improve the diagnosis, treatment, and management of various diseases. Consequently, several AI-based systems have been developed to measure polyp size. Some systems concentrate on binary size classification, where a threshold size is defined as 5 mm or 10 mm, and polyps are labeled as smaller or bigger than the threshold [16, 17]. Another AI-based system described by Kwak et al. relies on AI for a part of the measurement process and requires additional manual labor to define a linear segment on the image to be measured [18]. This method significantly improved accuracy compared with visual and biopsy forceps-based estimation. Even so, our approach is automated and does not require manual labor to annotate the polyp segments.

Recently, a study by Shimoda et al. evaluated a virtual scale that uses a laser beam to estimate the distance between the endoscope and the mucous surface [11]. Rather than measuring the polyp size, this system generates an on-screen scale and supports endoscopists in estimating the polyp size. The study reported a significant improvement in the accuracy of estimation from 62.5% without the virtual scale to 84% with. Regardless, this method requires a specific endoscopy system and colonoscope with an integrated laser beam. In contrast, our system is not bound to a specific endoscope manufacturer or type, as long as it has an auxiliary water channel. Implementing the waterjet as a measurement reference also removes the need to introduce additional devices, allowing its potential compatibility with a wide range of endoscopy systems.

We do however recognize that our system has several limitations as the accuracy of Poseidon depends on the proper posi-

tioning of polyps in the endoscopic image. The waterjet should be aimed adjacent to a polyp to be equally distanced away from the endoscope, and the polyp needs to be recognized by our polyp detection system. Although some tolerance exists, failure to adhere to this can affect the accuracy of the measurement. Additionally, the reference size used for the polyps from routine clinical practice was established through pixel-wise comparison of polyps and resection instruments, which can itself introduce some inaccuracies.

In conclusion, we have developed a freely available AI system for measuring polyp size in gastrointestinal endoscopy. It uses the auxiliary waterjet as a reference and does not require additional devices or instruments to make the measurement. The system demonstrated a significantly higher accuracy than other commonly used polyp sizing methods. In future, the system should be expanded to operate with additional endoscopy systems, while a larger study is required to further evaluate its performance when challenged with a wide range of polyps that are found in daily clinical practice, in particular larger polyps.

Competing Interests

The authors declare that they have no conflict of interest.

Funding

Eva Mayr-Stihl Stiftung
Fischerwerke GmbH & Co. KG
Forum Gesundheitsstandort Baden-Württemberg
Dieter von Holtzbrinck Foundation GmbH
Bavarian Center for Cancer Research
<http://dx.doi.org/10.13039/501100022804>

References

- [1] Sung H, Ferlay J, Siegel RL et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021; 71: 209–249
- [2] Turner KO, Genta RM, Sonnenberg A. Lesions of all types exist in colon polyps of all sizes. *Am J Gastroenterol* 2018; 113: 303–306
- [3] Hassan C, Quintero E, Dumonceau JM et al. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2013; 45: 842–864
- [4] Gupta S, Lieberman D, Anderson JC et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastrointest Endosc* 2020; 91: 463–485.e5
- [5] Chaptini L, Chaaya A, Depalma F et al. Variation in polyp size estimation among endoscopists and impact on surveillance intervals. *Gastrointest Endosc* 2014; 80: 652–659
- [6] Eichenseer PJ, Dhanekula R, Jakate S et al. Endoscopic mis-sizing of polyps changes colorectal cancer surveillance recommendations. *Dis Colon Rectum* 2013; 56: 315–321
- [7] Rubio CA, Grimelius L, Lindholm J et al. Reliability of the reported size of removed colorectal polyps. *Anticancer Res* 2006; 26: 4895–4899
- [8] Gourevitch RA, Rose S, Crockett SD et al. Variation in pathologist classification of colorectal adenomas and serrated polyps. *Am J Gastroenterol* 2018; 113: 431–439
- [9] Jin HY. Use of disposable graduated biopsy forceps improves accuracy of polyp size measurements during endoscopy. *World J Gastroenterol* 2015; 21: 623
- [10] Kume K, Watanabe T, Yoshikawa I et al. Endoscopic measurement of polyp size using a novel calibrated hood. *Gastroenterol Res Pract* 2014; 2014: 1–4
- [11] Shimoda R, Akutagawa T, Tomonaga M et al. Estimating colorectal polyp size with a virtual scale endoscope and visual estimation during colonoscopy: Prospective, preliminary comparison of accuracy. *Dig Endosc* 2022; 34: 1471–1477
- [12] Konikoff T, Goren I, Yalon M et al. Machine learning for selecting patients with Crohn's disease for abdominopelvic computed tomography in the emergency department. *Dig Liver Dis* 2021; 53: 1559–1564
- [13] Levy I, Bruckmayer L, Klang E et al. Artificial intelligence-aided colonoscopy does not increase adenoma detection rate in routine clinical practice. *Am J Gastroenterol* 2022; 117: 1871–1873
- [14] Kader R, Cid-Mejias A, Brandao P et al. Polyp characterisation using deep learning and a publicly accessible polyp video database. *Dig Endosc* 2022; doi:10.1111/den.14500
- [15] Römmele C, Mendel R, Barrett C et al. An artificial intelligence algorithm is highly accurate for detecting endoscopic features of eosinophilic esophagitis. *Sci Rep* 2022; 12: 11115
- [16] Itoh H, Oda M, Jiang K et al. Binary polyp-size classification based on deep-learned spatial information. *Int J Comput Assist Radiol Surg* 2021; 16: 1817–1828
- [17] Abdelrahim M, Saiga H, Maeda N et al. Automated sizing of colorectal polyps using computer vision. *Gut* 2022; 71: 7–9
- [18] Kwak MS, Cha JM, Jeon JW et al. Artificial intelligence-based measurement outperforms current methods for colorectal polyp size measurement. *Dig Endosc* 2022; 34: 1188–1195
- [19] Lux TJ, Banck M, Saßmannshausen Z et al. Pilot study of a new freely available computer-aided polyp detection system in clinical practice. *Int J Colorectal Dis* 2022; 37: 1349–1354
- [20] Kaltenbach T, Anderson JC, Burke CA et al. Endoscopic removal of colorectal lesions: recommendations by the US Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol* 2020; 115: 435–464