https://doi.org/10.33878/2073-7556-2024-23-2-28-34

The effectiveness of endoscopic diagnostics of colon tumors using artificial intelligence: prospective tandem study

Sergey I. Achkasov[1,2], Yuri A. Shelygin[1,2], Alexey A. Likutov[1,2], Dmitry G. Shakhmatov[1,2], Oleg M. Yugai[1], Ilya V. Nazarov[1], Tatiana A. Savitskaya[1], Airat F. Mingazov[1]

[1]Ryzhikh National Medical Research Center of Coloproctology (Salyama Adilya st., 2, Moscow, 123423, Russia)
[2]Russian Medical Academy of Postgraduate Education (Barrikadnaya st., 2/1, p.1, Moscow, 125993, Russia)

ABSTRACT *AIM: to estimate the effectiveness of a medical decision support system based on artificial intelligence in the endoscopic diagnosis of benign tumors during tandem study.*
PATIENTS AND METHODS: from October to December 2023, a single-center comparative tandem study of medical decision support system based on artificial intelligence "ArtInCol" was done. The first stage was a traditional colonoscopy under sedation, the second one — colonoscopy using AI. A pairwise comparison of the main indicators of the effectiveness was made.
RESULTS: in the primary endpoint, the polyp detection rate (PDR) in the traditional colonoscopy group was 40.6% vs 56.4% in the AI group, p = 0.0001 (RR = 1.39; 95% CI: 1.04–1.87). The mean number of lesions detected (MPP) was 1.63 (± 1.2) vs 2.47 (± 1.8) in the AI group (mean difference = 0.84; (95% CI: 1.07–0.61).
CONCLUSION: the study demonstrated the effectiveness of the original medical decision support system for benign colon tumors detection in real clinical practice. The further stage, a multicenter randomized trial is needed.

KEYWORDS: *colonoscopy, adenoma, polyps, artificial intelligence*

CONFLICT OF INTEREST: *financing from the Skolkovo grant*

FOR CITATION: Achkasov S.I., Shelygin Y.A., Likutov A.A., Shakhmatov D.G., Yugai O.M., Nazarov I.V., Savitskaya T.A., Mingazov A.F. The effectiveness of endoscopic diagnostics of colon tumors using artificial intelligence: prospective tandem study. *Koloproktologia.* 2024;23(2):28–34. (in Russ.). https://doi.org/10.33878/2073-7556-2024-23-2-28-34

ADDRESS FOR CORRESPONDENCE: *Mingazov Airat Fanilevich, Ryzhikh National Medical Research Center of Coloproctology, Salyama Adilya st., 2, Moscow, 123423, phone number: +7 (927) 695-17-52; e-mail: ajr.mingazof@gmail.com*

Received — 13.03.2024 *Revised — 21.03.2024* *Accepted for publication — 24.04.2024*

INTRODUCTION

Colonoscopy remains the key method of large intestine tumors diagnostics to date, which is important for the prevention of colorectal cancer. It is known that early detection and removal of polyps significantly reduces the risk of colorectal cancer, according to a vast American population study, by 53% [1]. It is important to note that colorectal cancer is one of the most frequent malignant diseases worldwide, thereby determining the high relevance of its prevention [2].

With the development of digital technologies, including artificial intelligence (AI), new prospects are opening up for improving methods of diagnosing large intestine tumors and, consequently, secondary prevention of colorectal cancer. The results of many foreign scientific papers have shown that the use of AI in colonoscopy can contribute to an increase in the number of detectable polyps, improve the quality and accuracy of the study [3,4].

In 2022, on the basis of the National Center of Coloproctology of the Ministry of Health of Russia, the 'Alnisoft' company developed an algorithm based on artificial intelligence as a component of the medical decision support system for colonoscopy 'ArtInCol'. The internal validation of the algorithm based on archival material from video recordings of colonoscopies has

demonstrated promising results in the form of an 83.2% accuracy in detecting tumors [5]. It is worth noting that the work carried out is based on a retrospective analysis of the archive of endoscopic studies and does not include known criteria for the effectiveness of colonoscopy, such as the adenoma detection rate (ADR) and polyp detection rate (PDR).

Thus, due to the need for clinical validation of the developed system, the presented comparative non-randomized study was done.

PATIENTS AND METHODS

A comparative tandem trial of the diagnostic method was done at the RNMRC of Coloproctology of the Ministry of Health of Russia within the period from October to December in 2023. Tandem trial is a method of confirming the effectiveness of a diagnostics, in which a consistent application of control and experimental visualizing methods is carried out. The study included adult patients who were assigned a screening colonoscopy and signed informed voluntary consent.

Non-inclusion criteria:

1. Verified diagnosis of benign or malignant tumors of the large intestine.
2. Inflammatory bowel diseases.
3. History of colorectal resection.
4. Contraindications to colonoscopy under sedation.

Exclusion criteria:

1. Poor bowel cleansing (less than 6 points, according to the Boston Bowel Preparation Scale).
2. The patient's refusal to continue participating in the study at any stage.

Colonoscopy was performed using the standard method using the 'Pentax 7010' (Japan) and 'Olympus Exera-III' (Japan) devices with high definition (HD) both in white light and in narrow spectrum modes. All studies were performed under intravenous sedation.

In accordance with the developed design, the tandem trial was done sequentially by two different endoscopists with comparable previous experience in performing colonoscopies — over 1,000 each.

At the first stage, an endoscopist performed a standard procedure with a time of the device withdrawal in the range from 6 to 8 minutes. At the second stage, another endoscopist, blinded on the results of an earlier procedure, performed a repeat colonoscopy in the same time range, but at the same time used the artificial intelligence system 'ArtInCol'. When performing the study using an AI assistant, the registration of the detected tumors was carried out with a stable fixation of the digital frame in the polyp area and subsequent verification of the tumor by the endoscopist who performed the diagnostic procedure (Fig. 1).

According to the results of the performed studies, all detected tumors were fixed, grouping them depending on the type (hyperplastic polyp, adenoma, serrated adenoma) and size (⩽ 1 cm and > 1 cm).

False positives of the AI assistant were recorded, for which machine vision was understood to highlight an area free of tumors for 3 seconds or more.

Primary endpoint:

– the detection rate of tumors of all types (PDR indicator).

Secondary endpoints:

– the adenoma detection rate (ADR indicator);
– the mean number of adenomas in one patient with tumors (MAP — mean adenomas per patient);
– the mean number of detected tumors of all types (MPP — mean polyps per patient);
– the rate and mean number of detected tumors, depending on the different types.

In the analysis of information, at the first stage the descriptive statistics was performed with the establishment of percentages for categorical data and the calculation of the mean and standard deviation for numerical variables with a preestablished normal distribution. A comparative analysis of the primary and secondary endpoints was performed using the MacNemar test for paired

categorical variables. Numerical variables were compared using the paired *t*-test or the Wilcoxon test. Statistical significance was assumed at $p < 0.05$. In order to demonstrate the effect value (the difference between the groups), the indicators of relative risk and the difference in averages were additionally calculated, indicating a 95% coincidence interval.

RESULTS

A total of 100 patients were included in the study. The descriptive characteristics of the examined patients are presented in Table 1.

Before comparing the main performance indicators, it was found that the groups did not differ in the time of colonoscope extraction from the large intestine ($p = 0.1$). In a comparative analysis of the primary endpoint parameter, the detection rate of large intestine tumors (PDR) in the traditional colonoscopy group was 40.6%, which is significantly less than in the group of patients studied with AI assistance — 56.4%, $p = 0.0001$ (RR = 1.39; 95% CI: 1.04–1.87). At the same time, the mean value of the number of detected tumors in the group of traditional colonoscopy was also significantly less: 1.63 (± 1.2) versus 2.47 (± 1.8) than in the group with AI (mean difference = 0.84; (95% CI: 1.07–0.61) (Table 2).

In terms of adenoma detection rate (ADR), the groups differed in favor of colonoscopy using AI: 34.7%, versus 40.6%, $p = 0.031$. Also, the colonoscopy group with the AI system exceeded traditional colonoscopy in the number of detected tumors of all types ⩽ 1 cm. It is worth noting separately that tumors over 1 cm in diameter were detected in both groups with approximately equal rate.

Table 1. *Descriptive statistics*

Variable	Value
Male Female	32 (32%) 68 (68%)
Age (years)	54.8 (± 12.3)
Direct indications for colonoscopy	80 (80%)
Quality of preparation (points)	9 (7-9)

Among patients with detected polyps of various types ($n = 55$), simultaneous removal of polyps with morphological verification was performed in 32 (58%) patients. The remaining patients were referred for elective endoscopic polypectomy.

The mean value of false positives was 2.5 (± 1) per study.

Taking into account the tandem design of the study, the indicator of missed tumors of all types was calculated. A total of 138 tumors using AI were detected, while 48 missed polyps were registered. Thus, the rate of missed tumors of all types (PMR) was 35%.

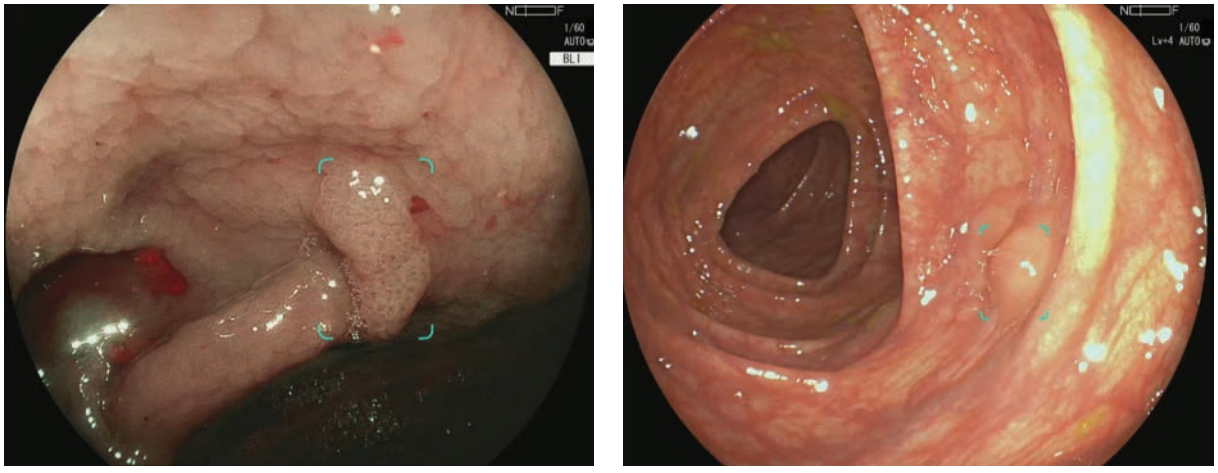

Figure 1. *Endophoto. Detection of the AI on the adenoma on the left, on the polyp on the right*

Table 2. *Results of comparative analysis*

Variable	Standard colonoscopy	Colonoscopy + AI	The value of P	The effect value
ADR (Adenoma detection rate)	34 (34%)	40 (40%)	P = 0.03*	–
PDR (Polyp detection rate)	40 (40%)	56 (56%)	P = 0.0001*	OP = 1.39 (1.04–1.87)
MPP (Mean polyps per patient) (Mean number of all detected tumors)	1.63 (± 1.41)	2.47 (± 2.07)	P = 0.0001*	MD = 0.84 (1.07–0.61)
The neoplasm is ≤ 1 cm in diameter				
MPP (Mean polyps per patient) (Mean number of all detected tumors)	1.35 (± 1.13)	2.18 (± 1.56)	P = 0.0001*	MD = 0.83 (1.06–0.59)
MAP (Mean adenomas per patient)	1.47 (± 1.13)	2.19 (± 1.68)	P = 0.0001*	MD = 0.72 (1.05–0.39)
Mean number of serrated tumors	0.88 (± 1.03)	1.61 (± 1.19)	P = 0,0001* P = 0.0001*	MD = 0.72 (1.007–0.43)
Mean number of hyperplastic polyps	0.42 (± 0.38)	1.42 (± 0.53)	P = 0.018*	MD = 1.02 (1.75–0.24)
The neoplasm is over 1 cm in diameter				
MPP (Mean polyps per patient) (Mean number of all detected tumors)	1.5 (± 1.1)	1.64 (± 1.2)	P = 0.16	MD = 0.14 (0.35–0.06)
MAP (Mean adenomas per patient) Mean number of adenomas	1.18 (± 0.6)	1.36 (± 0.67)	P = 0.34	MD = 0.18 (0.58–0.22)
Mean number of serrated adenomas	2 (± 2.2)	2.5 (± 3)	P = 0.18	MD = 0.5 (1.42–0.22)

*Note: * Statistical significance — p < 0.05*

DISCUSSION

In the course of planning scientific work to evaluate the effectiveness of AI-based MDSS, one of the non-trivial tasks is to choose the design of the study. The difficulty lies in the selection of clinically significant endpoints, the assessment of late results, as well as verification methods. In the case of medical decision support system (MDSS) in colonoscopy, one of the most common designs found in the literature is a randomized trial, and the authors choose ADR as the primary endpoint [6]. At the same time, even when a statistical significance in the rate of detected adenomas is achieved, cohort studies with a long follow-up period are required to assess late results, which, in turn, is unethical due to the development of interval cancer in missed adenomas. In addition, an important trial limitation in is the continuous improvement of the AI algorithm, which causes the loss of relevance of the results obtained due to the further development of AI.

The verification problem is also very important, since there is no suitable reliable confirmatory method for detecting large intestine polyps [7]. Technically, the model is based on the fact that in the course of real-time research, AI calculates the statistical probability of the presence of a benign neoplasm in a particular frame. In this case, false detections may be observed associated with the presence of gas bubbles, feces or folds of the mucous layer in the frame. An unambiguous verification method is the pathomorphological study of a removed tumor. However, this method is also

limited by the need to make a much more time-consuming study in this case. In addition, this verification method requires detailed labeling of tumors for the purpose of morphological examination, as well as the impossibility of verifying false positives. The expert level of endoscopy with the possibility of examination in a narrow light spectrum is close to a pathomorphology. However, it also does not guarantee absolute verification, while it is inaccessible during screening in real clinical practice [7].

Having at our disposal a working version of the AI assistant, which has already demonstrated diagnostic accuracy above 82% at the stage of internal validation of the system, we chose the design of a 'tandem' endoscopic examination. In this case, the analysis is performed by paired comparison, which allows us to identify the technical advantages of detecting 'additional' tumors at a deliberately lower cost of research [6]. The chosen design allowed us to obtain sufficient data on the effectiveness of the system in a smaller number of patients by using the paired comparison method. It is worth noting that the global goal of the work carried out was the external validation of the AI system in real clinical practice before conducting a large multicenter randomized trial, which will allow it to be planned in more detail and strictly. The results of the study highlight a significant improvement in the effectiveness of detecting large intestine tumors when using colonoscopy with the AI system (ArtInCol) compared with the traditional method. A statistically significant difference was obtained in the main parameters of the effectiveness of screening colonoscopy — ADR (difference — 6%) and PDR (difference — 16%). At the same time, the results obtained are consistent with data from other studies, which also note an improvement in diagnostic accuracy due to the use of an AI system. According to the largest systematic review, which included 33 randomized trials, the difference in ADR was 7.2% [3]. At the same time, it is important to take into account the effect of an increase in the ADR index on the risk of developing colorectal cancer. Thus, an increase in the detection rate of adenomas for every 1% reduces the risk of developing colorectal cancer by 5% [8].

Special attention should be paid to the fact that, according to the trial, the main share in the difference in the diagnosis of polyps (PDR) was achieved due to the detection by the AI system of tumors of small diameter (less than 1 cm in diameter). On the one hand, recent advances in the field of endoscopy, such as white light examination or narrow-spectrum imaging, have led to improved diagnosis of tumors and even optical biopsy with high diagnostic value [7]. On the other hand, according to various observational studies, the rate of missing polyps and adenomas varies from 20% to 50% after primary screening colonoscopy [9, 10].

At the same time, in a multivariate model, it was demonstrated that the omission of tumors is directly related to variable characteristics of patients, intestinal preparation, as well as location in the right sections, small size (less than 1 cm) and the multiple nature of the detected polyps [11]. In a randomized study by Zhao et al., the authors demonstrated a statistically significant decrease in the adenoma miss rate (AMR) from 36.6% to 14.6% with an increase in the colonoscope withdrawal time from 6 to 9 minutes [12]. The data of the presented studies confirm the operator dependence of screening colonoscopy, which is due to the limited possibilities of optical diagnosis by an endoscopist. It is worth noting that the use of the AI system in colonoscopy in our study made it possible to identify missed tumors of all types (PMR — polyp miss rate) — 35%. Performing a comparative assessment based on the indicators of missed tumors (AMR and PMR) was not possible with this study design.

An additional parameter for evaluating the effectiveness of the diagnosis of tumors is the number of detected polyps per 1 colonoscopy. In accordance with the results of our research, the developed AI system allows us to detect statistically significantly more tumors of all types. Thus, the mean detectable number of polyps (MPP) was

1.63 (± 1.2) in the traditional colonoscopy group, versus 2.47 (± 1.8) in the diagnosis using an AI system. At the same time, the statistical difference was achieved mainly due to the detection of tumors of small diameter (less than 1 cm) (mean difference = 0.84; 95% CI: 1.06–0.59).

An important point is that statistical calculations of this group of indicators were performed among patients with detected polyps, and not from among all colonoscopies. According to a meta-analysis by Lou et al., which summarized data from randomized studies of similar AI systems from different countries, the value of the number of detected tumors also correlates with the global ones when calculating the total number of colonoscopies [3].

CONCLUSION

Taking into account the described limitations, the study demonstrated the effectiveness of the domestic ArtInCol medical decision support system for the diagnosis of benign tumors of the large intestine in real clinical practice. The results of the work done will serve as the basis for further multicenter randomized research in order to compare the effectiveness of the AI system in real clinical practice.

AUTHORS CONTRIBUTION

Concept and design of the study: *Sergey I. Achkasov, Yuri A. Shelygin, Dmitry G. Shakhmatov, Airat F. Mingazov*

Collection and processing of materials: *Alexey A. Likutov, Ilya V. Nazarov, Oleg M. Yugai, Tatiana A. Savitskaya*

Statistical processing: *Airat F. Mingazov*

Text writing: *Airat F. Mingazov, Dmitriy G. Shakhmatov*

Editing: Sergey I. Achkasov

INFORMATION ABOUT THE AUTHORS (ORCID)

Sergey I. Achkasov — 0000-0001-9294-5447
Yuri A. Shelygin — 0000-0002-8480-9362
Alexey A. Likutov –0000-0001-5848-4050
Dmitry G. Shakhmatov — 0000-0001-7964-2126
Oleg M. Yugai — 0000-0003-4679-5497
Ilya V. Nazarov — 0000-0002-6876-4272
Tatiana A. Savitskaya — 0009-0008-2420-2770
Airat F. Mingazov — 0000-0002-4558-560X

REFERENCES

1. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med*. 2012 Feb 23;366(8):687–96. doi: 10.1056/NEJMoa1100370

2. 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 May;71(3):209–249. doi: 10.3322/caac.21660

3. Lou S, Du F, Song W, et al. Artificial intelligence for colorectal neoplasia detection during colonoscopy: a systematic review and meta-analysis of randomized clinical trials. *E Clinical Medicine*. 2023 Nov 30;66:102341. doi: 10.1016/j.eclinm.2023.102341

4. Wei MT, Fay S, Yung D, et al. Artificial intelligence-assisted colonoscopy in real-world clinical practice: A systematic review and meta-analysis. *Clin Transl Gastroenterol*. 2023 Dec 26. doi: 10.14309/ctg.0000000000000671

5. Mtvralashvili D.A., Shakhmatov D.G., Likutov A.A., et al. AI-based algorithm for clinical decision support system in colonoscopy. *Koloproktologia*. 2023;22(2):92–102. (in Russ.). doi: 10.33878/2073-7556-2023-22-2-92-102

6. Lee MC, Parker CH, Liu LW, et al. Impact of study design on adenoma detection in the evaluation of artificial intelligence-aided colonoscopy: a systematic review and meta-analysis. *Gastrointest Endosc*. 2024 Jan 23:S0016–5107(24)00040-3. doi: 10.1016/j.gie.2024.01.021

7. Vadhwana B, Tarazi M, Patel V. The Role of Artificial Intelligence in Prospective Real-Time Histological Prediction of Colorectal Lesions during Colonoscopy: A Systematic Review and Meta-Analysis. *Diagnostics (Basel)*. 2023 Oct 20;13(20):3267. doi: 10.3390/diagnostics13203267

8. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and

death. *N Engl J Med*. 2014 Apr 3;370(14):1298–306. doi: 10.1056/NEJMoa1309086

9. Seo JH, Lee BI, Lee K, et al. Adenoma miss rate of polypectomy-referring hospitals is high in Korea. *Korean J Intern Med*. 2020 Jul;35(4):881–888. doi: 10.3904/kjim.2018.099 Epub 2019 Oct 17. PMID: 31610632

10. Jiang W, Xin L, Zhu S, et al. Risk Factors Related to Polyp Miss Rate of Short-Term Repeated Colonoscopy. *Dig Dis Sci*. 2023 May;68(5):2040–2049. doi: 10.1007/s10620-023-07848-x Epub 2023 Apr 5.

11. Shin JG, Kim HW, Park SB, et al. Polyp missing rate and its associated risk factors of referring hospitals for endoscopic resection of advanced colorectal neoplasia. *Medicine (Baltimore)*. 2017 May;96(19):e6742. doi: 10.1097/MD.0000000000006742

12. Zhao S, Song Y, Wang S, et al. Reduced adenoma miss rate with 9-minute vs 6-minute withdrawal time for screening colonoscopy: a multicenter randomized tandem trial. *Am J Gastroenterol*. 2022. doi: 10.14309/ajg.0000000000002055