Artificial intelligence for the prevention and prediction of colorectal neoplasms

Kohjiro Tokutake (jir0908guitar@gmail.com)
Nagano Red Cross Hospital: Nagano Sekijuji Byoin
https://orcid.org/0000-0002-2827-4826

Aaron Morelos-Gomez
Shinshu university

Ken-ichi Hoshi
Nagano Red Cross Hospital: Nagano Sekijuji Byoin

Michio Katouda
Research Organization for Information Science & Technology

Syogo Tejima
Research Organization for Information Science & Technology

Morinobu Endo
Shinsung University

Research Article

Keywords: Colorectal cancer, screening, blood data, machine learning, artificial intelligence

Posted Date: April 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2711316/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Journal of Translational Medicine on July 3rd, 2023. See the published version at https://doi.org/10.1186/s12967-023-04258-5.
Abstract

Background

Colonoscopy is a useful as a cancer screening test. However, in countries with limited medical resources, there are restrictions on the widespread use of endoscopy. Non-invasive screening methods to determine whether a patient requires a colonoscopy are thus desired. Here, we investigated whether artificial intelligence (AI) can predict colorectal neoplasia.

Methods

We used data from physical exams and blood analyses to determine the incidence of colorectal polyp. However, these features exhibit highly overlapping classes. The use of a kernel density estimator (KDE)-based transformation improved the separability of both classes.

Results

Along with an adequate polyp size threshold, the optimal machine learning (ML) models' performance provided 0.37 and 0.39 Matthews correlation coefficient (MCC) for the datasets of men and women, respectively. The models exhibit a higher discrimination than fecal occult blood test with 0.047 and 0.074 MCC for men and women, respectively.

Conclusion

The ML model can be chosen according to the desired polyp size discrimination threshold, may suggest further colorectal screening, and possible adenoma size. The KDE feature transformation could serve to score each biomarker and background factors (healthy lifestyles) to suggest measures to be taken against colorectal adenoma growth. All the information that the AI model provides can lower the workload for healthcare providers and be implemented in health care systems with scarce resources. Furthermore, risk stratification may help us to optimize the efficiency of resources for screening colonoscopy.

Background

Despite evidence that several colorectal cancer (CRC) screening strategies can reduce CRC mortality, screening rates remain low. CRC is one of the most common forms of cancer worldwide, causing the third highest number of cancer deaths.(1) Both the incidence and mortality of CRC have risen rapidly in Asian countries.(2) Strong evidence shows that screening for CRC can improve patients' survival, and many countries have implemented CRC screening programs.(3) Colonoscopy is considered the most accurate test for the early detection and prevention of CRC. For high-risk patients who have a family history of CRC in first-degree relatives, colonoscopy is the recommended screening tool.(4) It has been demonstrated that the mortality rate decreases due to colonoscopic polypectomy.(5)
However, the burden on the health care system is overwhelming in many countries.(6) According to studies conducted in Asia-Pacific countries, the reported prevalence of advanced colorectal neoplasms in the general population ranges from 3% to 12.5%, and thus unnecessary colonoscopies would be performed if a colonoscopy were offered to the entire population as the primary screening modality.(7)

Computed tomography colonography is an option for CRC screening in asymptomatic average-risk individuals.(8) However, since the corresponding facilities are limited, the number of screening tests is limited, and there is a challenge to reduce radiation exposure and colonic perforation.

Stool-based CRC screening tests are suitable for screening a large population. Multitarget stool DNA (mt-sDNA) testing, which includes multiple molecular assays combined with a hemoglobin immunoassay is available(9). However, genetic analysis is more expensive than other stool-based tests. The fecal immunochemical test (FIT) is preferred as the primary screening method because it is less expensive than colonoscopy and genetic analyses. The current practice in Japan uses colonoscopy as a confirmatory test for patients who test positive for fecal occult blood; it is considered as a primary screening tool in population screening programs. Unfortunately, FIT-positive tends to be disregarded, and patients do not receive a colonoscopy. For asymptomatic subjects, the most effective motivation to undergo a CRC screening test is a recommendation from a family physician.(10) A new and validated screening method would be a good motivation to receive colonoscopy for subject with a high risk of polyp. Improving awareness of CRC and promoting the physicians’ role are necessary to increase the screening participation rates.

Stratifying the population by risk offers the potential to improve the efficiency of screening. CRC risk models can serve for pre-screening to suggest further actions if a patient has a high probability for developing CRC. Diverse data types have been used to create CRC risk models: routine data, self-completed questionnaire, non-genetic biomarkers and genetic biomarkers.(11,12)

Artificial intelligence (AI) is a tool that can be applied in gastroenterology. In particular, AI has been used for the identification of polyps based on colonoscopy images.(13,14) AI models may also use datasets other than colonoscopy images, such as blood cell counts, urine analysis, gene expression, and other biomarkers.(15–18) Among the diversity of medical data, blood biomarkers have been used for predicting ovarian cancer,(19) human aging,(20) renal injury, (21) breast cancer,(22) atrial fibrillation(23) and other disorders. The use of blood biomarkers for predicting colorectal adenoma is thus promising, possibly reducing the risk for CRC if done in a timely manner. In particular, blood biomarkers of inflammation have been used to predict the presence of colorectal adenoma(24).

We investigated whether artificial intelligence could predict colorectal neoplasia, toward the goals of predicting and preventing CRC. In the present study, we developed an AI pipeline to obtain models that identify patients with high likelihood to exhibit colorectal neoplasia, using accumulated data from medical checkup. The insights and developed models may serve as a highly efficient guide to prevent and diagnose colorectal adenomas.
A numerical feature mapping algorithm was developed based on kernel density estimation (KDE) to transform the initial features into probabilistic features; along with an adequate polyp size threshold, the optimal models exhibit an MCC of 0.303 and 0.358 for men and women, respectively, which exhibit higher discrimination performance than without a KDE transformation and higher performance than FIT.

Methods

We retrospectively analysed the data of consecutive patients who had undergone a medical checkup and colonoscopy within one year after a medical checkup at Nagano Red Cross Hospital in 2015-2019. If the patients received a positive result on the FIT at a medical checkup, they were advised to undergo a colonoscopy. Even the patients with a negative FIT result could undergo a colonoscopy if they desire to receive screening test. Other patients underwent a colonoscopy several months after their medical checkup, for other reasons. For example, some patients were treated due to abdominal symptoms a few months after the medical checkup and underwent a colonoscopy within 1 year.

The inclusion criteria was men or women > 20 years old. The exclusion criteria included the following: history of colorectal surgery or colorectal cancer, familial adenomatous polyposis, Lynch syndrome, and any contraindication for colonoscopy. Subjects who have received the second colonoscopy within the research period also have been excluded.

Colonoscopy and colorectal neoplasms

The hospital gastroenterology department has experienced gastroenterologists who perform colonoscopies. All the subjects received same-day bowel preparation. The colonoscopist spent > 6 min for each scope withdrawal. All detailed findings, including neoplastic and non-neoplastic lesions, were recorded in a standard case report form. All the patients for this study had a colonoscopy and the results gave positive or negative for polyp occurrence. Colonoscopists were required to describe the extent of the examination, document cecal visualization, rate the quality of preparation on the Boston Bowel Preparation Scale (BBPS, 0–3 points for each of 3 colon segments), and record the size and location of lesions. The size of polyps was recorded as the largest observation starting from 1 mm.

Results

A total of 1003 subjects from medical checkup exams at the Nagano Red Cross Hospital were recruited in this study. The mean age of the study subjects is 60.4 years, 611 subjects are men (60.9%) and 392 are women (39.1%). The mean time from medical checkup to colonoscopy was 74 days, and the mean withdrawal time was 555 seconds. The bowel preparation was evaluated based on BBPS and mean score was 7.54. Overall, there are 13 cases (1.3%) of invasive cancer, 93 cases >=10mm adenoma (8.0%), 113 cases of >=8mm (11.3%), 210 cases of >= 6 mm (20.9%), and 547 cases of >= 1 mm (54.5%) found in this series. Both malignant and benign lesions were treated as polyps, and the maximum diameter among the observed polyps was measured and recorded.
The overall pipeline consisted of data separation by sex, regrouping patients according to polyp size, transforming the initial data into probabilistic values, optimization of the data transformation parameters, feature selection and selection of the optimal machine learning (ML) method (Figure 1). This was done to explore a wide variety of models and select the best for the men and the women datasets. A complete description of the method can be found in the supporting information.

The polyp size is widely used to determine clinical importance where the use of a polyp size threshold is commonly assigned between 5 and 10 mm, e.g. in the European society of gastrointestinal endoscopy guidelines for CRC screening.(25) The choice of a polyp size threshold could be due to (1) the risk of developing CRC decreases with polyp size when the polyp(s) are \( \leq 10 \) mm,(26) and (2) the human error for detecting polyps increases when the polyp size is smaller.(27) Because of these considerations, the polyp size was used as a threshold; patients with polyp sizes below the threshold are kept in the same group as the patients without polyps (no) and the rest in the group with polyps (yes).

The input data were treated since the initial numerical features exhibited distributions that are highly overlapping for both incidences of polyps (Figure S1). To convert the numerical data into probabilistic values, a KDE-based transformation was implemented. The use of KDE for feature transformation has shown better performance than without KDE-based transformation.(28) Here, for each feature, a KDE was generated using the group of patients with polyps, then the KDE was normalized to 1, and an exponent between 1 and 4 was applied (Figure 2a).

Sigmoid functions have been widely used to correlate numerical variables with binary classifications, it is used in logistic regression and has been applied in gastroenterology.(29,30) The KDE was transformed into a sigmoid-like function where the KDE has a value of one before or after the maximum KDE value, according to the Table S1, and an exponent between 1 and 4 was applied to the sigmoid-like KDE (Figure 2b).

An optimization algorithm was implemented to obtain the best ML models for men and women (Figure 3), using a specified polyp size threshold (0, 6, 8, or 10 mm), selection between KDE or sigmoid-like KDE transformations and their respective exponents between 1 and 4. For the present ML models, the sensitivity, specificity, AUC, accuracy, and MCC were calculated with the test data for understanding the performance of the models (Tables S2, S3, and Figure 3), and MCC was used as the target metric for the optimization process. For the men's dataset, the best model is at a polyp threshold size of 0 mm, exponent 4, sigmoid-like KDE-based feature transformation using a Gaussian process classifier, and 0.37 MCC (Figure 3a-c). For the women's data, the optimal performance is chosen at 0 mm, exponent 3, sigmoid-like KDE-based feature transformation using A Gaussian process classifier, yielding 0.39 MCC (Figure 3d-f).

The optimal models with a KDE-based feature transformation have a better discrimination performance than the models without a KDE-based feature transformation considering 0 mm as the polyp size threshold: 0.24 MCC for men, and 0.14 MCC for women (Figure 3g,h). The higher discrimination performance at 0 mm threshold could be due to the lower probability in patients with smaller polyps to
exhibit a false positive given the lack of a previous polyp resection or improvements to their lifestyle. These results demonstrate that the sigmoid-like KDE-based feature transformations enhance the model performance by increasing the separation between the maximum peaks in the feature distributions for each polyp incidence (“yes” and “no”, Figure 4a-d).

The AI performance metrics are distinct between both sex datasets possibly due to the following reasons. (a) In this study group, the drinking habits and polyp incidence vary among men and women (Table S6). (b) Additionally, the number of patients with polyps among the women (n = 54) is lower than that among the men (n = 116). (c) The datasets by sex exhibit several features with P-values < 0.05 using the Kruskal-Wallis H-test (Table S5). The optimization algorithm applied on both men and women datasets yielded the selection of 34 and 16 features, respectively, from the initial 39 features. The feature importance was obtained via the Shapley additive explanations (SHAP) method, and for each feature the polyp likelihood can be observed as a higher SHAP value suggests a higher probability to exhibit a polyp (Figures 5a,b). (31,32) For both sexes, the features are mainly related to obesity and alcohol consumption habits, and may be linked to polyp incidence.(33–35) In particular, age and obesity appear as the most important features for both sexes. Features related to obesity exhibit high feature importance, and have been linked with polyp incidence. (33,36) Both high-density lipoprotein (HDL) and low-density lipoprotein (LDL) are good indicators and they are found within the selected features for each sex, suggesting that lipid profiles can affect the development of colorectal adenomas.(37,38)

AI models can provide a prediction probability or a decision function, here, we consider it as a score value, labelled as polyp score. The KDE of the polyp scores was classified by the real polyp incidence (yes/no), and showed that the optimal model for men exhibits some overlap between both polyp incidence distributions, fortunately, the opposite edges exhibit a lower overlap (Figure 6a). Surprisingly, the score distributions for true positive (TP), true negative (TN), false positive (FP), and false negative (FN) show that the FN and FP values exhibit a narrower distribution than the TN and TP (Figure 6b). The confusion matrix shows a higher accurate prediction for “yes” than for the “no” polyp incidence (Figure 6c). Regarding the women’s dataset, the KDE polyp scores exhibit a larger gap between the maximum distributions than those for the men (Figure 6d). In addition, the results reveal similar polyp score distributions among TP-FP and TN-FN (Figure 6e). From the confusion matrix, the accuracies for the true predictions are higher than the false predictions. In addition, the true predictions exhibit similar accuracies for both polyp incidences (Figure 6f). The models for the men and women had a low FN, meaning that few patients could be mis-predicted as not having any polyps when a colonoscopy shows a polyp.

Discussion

The FIT is widely used due to its low cost, non-invasive nature, and it has proven efficiency in the reduction of the CRC incidence given its high detection rate of adenomas.(39) The present models
outperform the FIT by sensitivity, specificity, and MCC for both sexes (Figure 7a,b). In addition, the present technique could be better than genetic analysis of stool samples which can identify colorectal adenomas with an average 35% sensitivity (21% standard deviation). In addition, gene expression analysis can be more expensive and require more specialized equipment than the present blood marker analysis. Therefore, the present models can provide an easier and cheaper screening method.

Unfortunately, some patients convince themselves that the FIT-positive may be caused by hemorrhoidal bleeding, and underestimate the FIT results. Since this model is based on biomarkers and background factors, it could be more persuasive for patients. In addition, the sigmoid-like KDE for each feature could be used as a recommendation for patients to maintain or improve their health lifestyle (e.g., nutrition, alcohol consumption, exercise), and to take care when reaching a particular age. When the KDE transformed values are close to 1 it could be advisable to take measures to avoid the development of colorectal adenoma or carcinoma. The results can be explained from the health check doctor on the day of medical checkup and lead to effective motivation.

The best models outperform the predictions by FIT, however, there could still be room for improvement. Here, the population was considered for a single hospital, with data from other clinics and hospitals it could be possible to obtain better representative data. In addition, clinical history could also help by knowing if a patient is currently taking pharmaceuticals that may reduce polyps such as aspirin. This study includes few cases with a history of colon polypectomy. Some cases which had polyps removed few years ago may be judged to be no polyp, therefore, clinical history could help reduce false negatives. The models may improve with a larger dataset with patients who are subjected to a colonoscopy for the first time. Furthermore, based on the present results, it is worthwhile to note that further study by more number of patient data and also other various algorithms for optimal prediction of the colorectal neoplasms are very promising.

Limitations

The data for this study were collected in a single hospital, where 22 cases with CRC were identified (14 invasive cancers and 8 mucosal cancers), making it difficult to estimate the sensitivity for invasive cancers precisely. Our data includes few cases with a history of colon polypectomy. We believe that the present models could be enhanced by using a larger dataset of patients who had colonoscopy for the first time as well as medical history such as diabetes, colorectal adenoma, and data from previous colonoscopies.

Conclusions

The results of the analyses demonstrate that the use of a KDE-based feature transformation can enhance the model performance by increasing the separability among polyp incidence groups. In particular, the sigmoid-like KDE transformation is more beneficial and could provide insights into which biomarker should be monitored to avoid further colorectal adenoma development.
This study demonstrates that risk scoring supported by AI is an effective model to stratify the risk of advanced colorectal neoplasm in asymptomatic subjects undergoing screening tests, which could be further improved with a larger dataset. Each country has its guidelines to determine which treatment or procedure are suggested according to the size of the polyp. The pipeline has the option to select the best model according to a polyp size discrimination threshold. With the given polyp score, the present models could identify patients at a high likelihood of detecting colorectal neoplasm. The procedure is simple enough, requiring only the medical data to be used on the trained models. The models can be used by family physicians and community health care providers to improve screening efficiency, requiring only the input of a physical exam, lifestyle habits, and blood analysis. No additional techniques or costs such as genetic testing are required. We hope that the present findings could aid in the development of AI assisted colorectal adenoma screening.

**Abbreviations**

AI: artificial intelligence; KDE: kernel density estimator; ML: machine learning; MCC: matthews correlation coefficient; CRC: colorectal cancer; FIT: fecal immunochemical test; BBPS: Boston Bowel Preparation Scale; SHAP: Shapley additive explanations; HDL: high-density lipoprotein; LDL: low-density lipoprotein; TP: true positive; TN: true negative; FP: false positive; FN: false negative.

**Declarations**

**Ethics approval and consent to participate**

This study was approved by the Institutional Review Board of Nagano Red Cross Hospital (approved number: 2018-155). Informed consent was not required for this study as the data was de-identified when outputting data from electronic medical records.

**Consent for publication**

Not applicable.

**Availability of data and materials**

The Nagano Red Cross Hospital owns the data and is unsuitable for public availability due to hospital restrictions. The code used in this study to develop and evaluate the models are proprietary, therefore, the code is not publicly available.

**Competing interests**

The authors declare no competing interests.

**Funding**
Not applicable.

Authors' contributions

KT and KH designed outline of the study and collected clinical data. AMG designed the optimization and prediction models. AMG, KT, MK, ST and ME participated in discussions and the manuscript writing. All authors read and approved the final manuscript.

Acknowledgements

Thanks to all participants involved in this research. We would also like to thank Nanako Kamisawa, Junichi Kobayashi, Keiko Shibata, Soichiro Shibata, Tetsuya Ito, Masafumi Maruyama, Toru Fujisawa, Hiromitsu Mori, Yoshiaki Matsuda and Shuichi Wada, Nagano Red Cross Hospital for collaboration in the collecting clinical data.

References


Figures
Figure 1

Pipeline to obtain the best optimal after selecting gender, polyp size threshold, and feature transformation. The training data was used for the optimization process. First, the KDE bandwidth was optimized for each feature, then the feature that decreased the most the discrimination performance was eliminated, these steps were iterated until 3 features remained. The best 6 models from all the iterations were selected and evaluated on the test data to select the best of all. All the model selections were based on the MCC value.
Figure 2

(a) KDE and (b) sigmoid-like KDE functions with different exponents were generated.
Figure 3

Classification metrics for the (a-c) men and (d-f) women performed on the test datasets. The Matthews correlation data for (a and d) KDE, and (b and e) sigmoid-like KDE feature transformations at different polyp size thresholds and KDE exponents are shown. Receiver operating characteristic (ROC) curves on the test data for (c) the men and (f) the women applying the optimal models, the optimal models were those with the highest Matthews correlation: 0 mm polyp size threshold, sigmoid-like KDE-based feature transformation. The Matthews correlation obtained at distinct polyp size threshold without KDE transformation (No - KDE) for the datasets of (g) the men and (h) the women are shown.
Figure 4

KDE plots of the features with the highest feature importance for (a) men and (c) women and after (b and c) sigmoid-like KDE transformations.
Concerning the feature importance for the (e) men’s and (f) women’s datasets, the shown feature importance were obtained via the SHAP method. The plotted results are those obtained for the optimal models: polyp size threshold 0 mm, sigmoid-like KDE transformation, and KDE exponent 4 for the men; polyp size threshold 0 mm, sigmoid-like KDE transformation and KDE exponent 3 for the women.
Prediction model performance for the best models for men and women: 0S-SKDE-4E-GP and 0S-SKDE-3E-GP, respectively. S stands for the polyp threshold size, SKDE is the sigmoid-like KDE transformation, E is the exponent applied on the KDE transformation and GP is Gaussian process classifier. (a,d) The KDE of the predicted scores for polyp incidence. Each distribution of yes/no is for the true incidence. (b,e) Violin plots of the predicted scores for true positive (TP), true negative (TN), false positive (FP), and false negative (FN). (c,f) Normalized confusion matrix.
Figure 7

(a,b) Sensitivity (magenta), specificity (red), and Matthews correlation (orange). The present models outperform the FIT results for both genders. All the data are presented for the optimal selected models.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPORTINGINFORMATION.docx