PLOS ONE

RESEARCH ARTICLE

Detecting microsatellite instability in colorectal cancer using Transformer-based colonoscopy image classification and retrieval

Chung-Ming Lo[1], **Jeng-Kai Jiang**[2,3☯], **Chun-Chi Lin**[2,3☯*]

1 Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taipei, Taiwan, **2** Department of Surgery, Division of Colon and Rectal Surgery, Taipei Veterans General Hospital, Taipei, Taiwan, **3** Department of Surgery, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

☯ These authors contributed equally to this work.
* cclin15@vghtpe.gov.tw

Citation: Lo C-M, Jiang J-K, Lin C-C (2024) Detecting microsatellite instability in colorectal cancer using Transformer-based colonoscopy image classification and retrieval. PLoS ONE 19(1): e0292277. https://doi.org/10.1371/journal.pone.0292277

Editor: Abel C. H. Chen, Chunghwa Telecom Co. Ltd., TAIWAN

Received: March 25, 2023

Accepted: September 15, 2023

Published: January 25, 2024

Data Availability Statement: Data is uploaded as supporting information.

Funding: The authors thank the Ministry of Science and Technology of Taiwan (MOST 111-2221-E-004-012) and VGHUST Joint Research Program (VGHUST112-G1-4-1, VGHUST112-G1-4-2) for financially supporting this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Colorectal cancer (CRC) is a major global health concern, with microsatellite instability-high (MSI-H) being a defining characteristic of hereditary nonpolyposis colorectal cancer syndrome and affecting 15% of sporadic CRCs. Tumors with MSI-H have unique features and better prognosis compared to MSI-L and microsatellite stable (MSS) tumors. This study proposed establishing a MSI prediction model using more available and low-cost colonoscopy images instead of histopathology. The experiment utilized a database of 427 MSI-H and 1590 MSS colonoscopy images and vision Transformer (ViT) with different feature training approaches to establish the MSI prediction model. The accuracy of combining pre-trained ViT features was 84% with an area under the receiver operating characteristic curve of 0.86, which was better than that of DenseNet201 (80%, 0.80) in the experiment with support vector machine. The content-based image retrieval (CBIR) approach showed that ViT features can obtain a mean average precision of 0.81 compared to 0.79 of DenseNet201. ViT reduced the issues that occur in convolutional neural networks, including limited receptive field and gradient disappearance, and may be better at interpreting diagnostic information around tumors and surrounding tissues. By using CBIR, the presentation of similar images with the same MSI status would provide more convincing deep learning suggestions for clinical use.

Introduction

Colorectal cancer (CRC) affected more than 1.9 million people in 2020 and was responsible for approximately 935,000 deaths. It is currently the second leading cause of death and the third most commonly diagnosed cancer worldwide [1]. Microsatellite instability-high (MSI-H) is a defining characteristic of hereditary nonpolyposis colorectal cancer syndrome, and around 15% of sporadic colorectal carcinomas exhibit MSI-H [2]. Tumors with MSI-H display distinct features, such as a preference for developing in the proximal colon,

Competing interests: The authors have declared that no competing interests exist.

lymphocytic infiltration, and a poorly differentiated, mucinous, or signet ring appearance [2,3]. Furthermore, MSI-H tumors are associated with specific pathological characteristics, such as a host immune response, including Crohn's-like lymphoid reaction, intratumoral lymphocytic infiltrate, and intraepithelial T cells. Compared to MSI-low (MSI-L) and microsatellite stable (MSS) tumors, CRCs with MSI-H have a more favorable prognosis [2,4].

The identification of MSI in CRC has revealed the heterogeneity of this disease, and the use of neoadjuvant therapy with immune checkpoint blockade in dMMR/MSI-H tumors has resulted in favorable response rates. This has significant clinical importance for organ-preserving approaches [5–7] and has implications for the treatment strategy in the management of CRCs. As a result, MSI testing has become a crucial component of CRC management. The Bethesda guidelines [8] have been widely accepted as the criteria for MSI testing. Currently, the NCCN Guidelines recommend universal MMR or MSI testing for all patients with a personal history of colon or rectal cancer. Besides serving as a predictive marker for immunotherapy in advanced CRC, MMR/MSI status can also aid in identifying individuals with Lynch syndrome [9].

Various methods for detecting microsatellite instability include fluorescent multiplex polymerase chain reaction (PCR) and capillary electrophoresis (CE) [10,11], immunohistochemistry (IHC) [12], and next-generation sequencing [13]. However, these techniques require a considerable amount of resources and labor. Some studies have examined the usefulness of histopathology in identifying MSI-H cancers by evaluating the pathologic features. While histopathological evaluation can be used to prioritize sporadic colon cancers for MSI studies, the morphological prediction of MSI-H has low sensitivity, necessitating molecular analysis for therapeutic decisions [14].

Colonoscopy is a valuable tool for diagnosing CRC, providing important information about the appearance, location, and depth of invasion of tumors in the colon wall. Tumors can vary in their appearance, with irregular, depressed, or ulcerating surfaces, and surrounding tissues can also provide important information. Statistical analysis can be used to summarize image characteristics that differentiate between different tumor statuses, but quantifying these image findings can be challenging. The features used to describe CRC status can be complex and difficult to quantify, and changes in illumination can result in changes in the findings. Additionally, interpreting the relationship between tumors and adjacent tissues can be difficult. Machine learning classifiers have been developed that use various methods to combine image features to classify tumors, providing an overall evaluation by probability and solving the problem of considering numerous findings simultaneously.

Deep learning offers an improved approach for extracting image features, as it can map image pixels into a high-dimensional feature space and automatically interpret the relevant image characteristics for a specific classification task without the need for human intervention. In the field of computer vision, deep learning architectures such as convolutional neural networks (CNNs) and vision Transformers (ViTs) have been developed. While CNNs have been used for pattern recognition in colonoscopy, they have limitations in scaling up the receptive field. This study proposed the use of ViT, which considers the global relationships between tumors and adjacent tissues in colonoscopy, as a more promising approach for feature extraction.

Echle et al. [15] used haematoxylin and eosin (H&E)-stained slides and molecular analysis findings to validate CNN approaches for predicting MSI in colorectal tumors across all stages. They achieved a mean area under the receiver operating characteristic curve (AUC) of 0.92. Yamashita et al. [16] proposed another CNN approach, based on a modified MobileNetV2 architecture pre-trained on ImageNet, and fine-tuned it to detect MSI from H&E-stained whole-slide images. Chang et al. [17] further improved on this by adding an attention

mechanism to the CNN, achieving an AUC of over 0.95 in predicting MSI in H&E-stained images. Peng et al. [18] classified different tissues in colorectal cancer histology slides using a CNN-based image retrieval approach, which provided more transparency and generalizability, and achieved higher precision than a classification network. Finally, Komura et al. [19] proposed a CNN-based content-based image retrieval (CBIR) method that successfully predicted 309 combinations of genomic features and cancer types by retrieving histologically similar images.

Previous studies have shown the diagnostic potential of H&E-stained slides in detecting MSI, with researchers from various countries using CNN-based methods to establish prediction models and investigate image patterns. In contrast, our study utilized the ViT architecture instead of CNNs to predict MSI, benefiting from multi-head self-attention to address the limitations of CNNs in scaling up the receptive field and avoiding gradient vanishing [20]. Furthermore, this study serves as a proof of concept, highlighting the distinctive aspect of our research, which utilizes colonoscopy instead of H&E-stained slides for deep learning-based MSI prediction. We also implemented CBIR to demonstrate its precision in identifying similarities and differences among tissue types, which can aid decision-making and establish correlations between classification bases and tissue types.

Materials and methods

Study population

In this research, we conducted an analysis of two cohorts consisting of CRC patients. The first cohort was enrolled between May 2014 and December 2017 and comprised of 441 patients, among whom 407 had MSS CRC and 34 had MSI-H CRC. To increase the sample size of MSI-H CRC, we enrolled an additional 89 patients who underwent surgery between January 2018 and May 2021. All patients underwent primary tumor resection at our hospital, and preoperative colonoscopy imaging was conducted for analysis. The primary tumor colonoscopy images were randomly captured during tumor diagnosis or preoperative localization. Patients with synchronous or metachronous CRC and those who received neoadjuvant therapy were excluded from the study. Patients who didn't have data about MSI status or pre-operative colonoscopy images were also excluded. The institutional review board approved the study protocol (2023-01-001CC), and the requirement for written informed consent was waived. The data were collected and analysis was conducted since Feb 2023 after approval of institutional review board. Because the correlation of MSI and the colonoscopy is necessary in the current study, the authors (J-K Jiang and C-C Lin) could access to information that could identify individual participants during or after data collection.

MSI testing

MSI testing was performed at our hospital since 2014. Immunohistochemistry (IHC) staining of tumor tissue was used to detect the expression of the four mismatch repair (MMR) genes, namely MLH1, MSH2, MSH6, and PMS2. A normal IHC test indicated that all four MMR proteins were expressed normally, and the tumor was considered MSS. Conversely, an abnormal IHC test suggested that at least one of the MMR proteins was not expressed, indicating a possible inherited mutation in the related gene. Loss of protein expression by IHC in any of the MMR genes was confirmed by specialized gastrointestinal pathologists with expertise in CRC pathology.

Vision Transformer

Deep learning approaches including CNNs and ViT have been suggested to recognize patterns in medical images [21–23]. Especially, ViT has shown improved generalization compared to

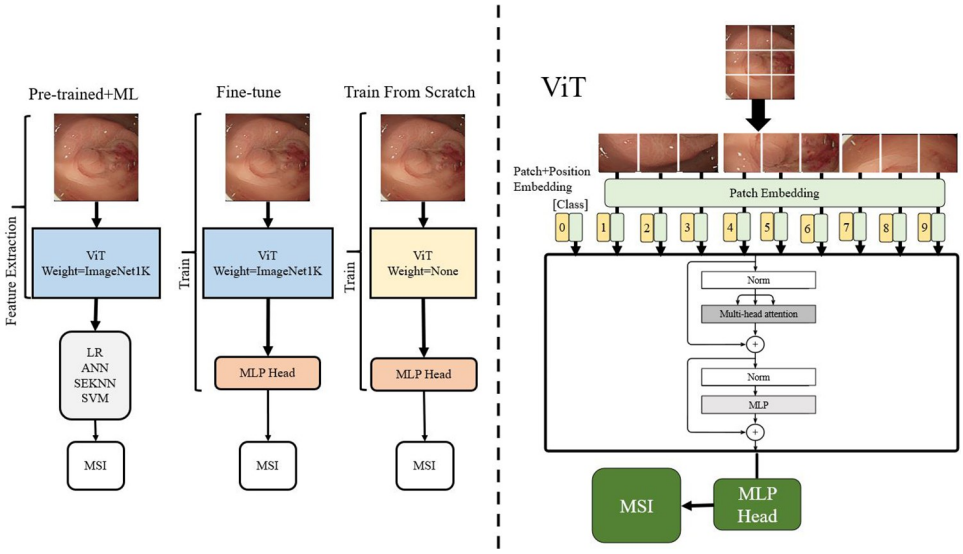

Fig 1. Vision Transformer modeling.

https://doi.org/10.1371/journal.pone.0292277.g001

CNNs, meaning that it can perform well on images outside of the training dataset [24,25]. ViT's architecture, as depicted in Fig 1, begins by flattening the split patches and projecting them into patch embeddings.

This resulting sequence is then preprocessed with a prepend class token (x_{class}) and position embeddings (E_{pos}) to preserve the positional information of the original image, as expressed in Formula 1:

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \ldots; x_p^N E] + E_{pos} \quad (1)$$

where x_1, x_2, . . ., and x_p are patches.

Subsequently, the encoder is constructed using multiple rounds of concatenating multi-headed self-attention, and multilayer perceptron blocks. The blocks are layered in the usual way and concluded with a residual connection. The multihead self-attention mechanism employed by ViT involves converting the input x^i into a^i, which is then processed through the self-attention layer. At this stage, it is multiplied by three different matrices to generate query (q), key (k), and value (v) vectors. These vectors can be further expanded into a multihead structure by repeating the operation. The q vector is then used to perform inner products with different k vectors, producing similarity measurements that consider long-term dependence. The resulting weights are then multiplied by the v vector to produce the final output.

Model training

Training deep neural networks involves optimizing the model parameters to minimize the loss function for a specific task. There are three different approaches as shown in Fig 1 to model training, each with trade-offs between accuracy and efficiency: training from scratch, pre-training, and fine-tuning. Training from scratch initializes the training process with random weights and biases, and the model is trained on a large labeled dataset for a specific task. While computationally expensive, this approach can achieve good performance with enough labeled data. Pre-training involves training a neural network on a large and diverse dataset, such as ImageNet, to generate substantial weights and biases that represent characteristics of the data. If the labeled dataset for the specific task is limited, pre-training can provide general features

that help distinguish data. Fine-tuning continues training on a smaller, task-specific dataset to learn more specific features related to the target task, leading to faster convergence and better performance. Fine-tuning requires fewer computational resources and is faster than training from scratch. In the experiment, using ViT as the base deep learning architecture, the three approaches were implemented for comparisons for the MSI prediction.

In the literature, when trained on large-scale datasets, ViT outperformed the ResNet architecture [20]. Thus, the pre-trained parameters used in the experiment was trained from ImageNet ILSVRC 2012 dataset [26], containing 1.3M images and 1k classes. By splitting the input image with patch size = 16×16, the multi-head self-attention mechanism trained 768 feature vectors to represent image characteristics.

Performance evaluation

The approaches of train from scratch and fin-tune are complete neural networks which have the end layer to generate classification labels for each image. For a specific task, the output vector represents the probabilities of the input belonging to each class. Softmax function can normalize each element in this vector to a range of 0 to 1, and the sum of probabilities is 1. Consequently, the softmax function can interpret the probability vector output as the probabilities for each corresponding class. During training, the cross-entropy loss function calculates the difference between the predicted probabilities and the true labels, and then the backpropagation mechanism is used to update the weights in the network.

Pre-training approach used an alternative way for classification since it is trained based on the prior data and labels which can't used in other tasks. Extracting features from the pre-train model and combining them in machine learning classifiers would be more practical. Additionally, the machine learning classifiers can perform better in generating nonlinear decision boundaries than softmax. Also, the minority class can be effectively handled in the classification.

In the experiment, four classifiers were used for training and testing the MSI classification using pre-train features, including logistic regression (LR) [27], artificial neural network (ANN) [28], subspace ensemble k nearest neighbor (SEKNN) [29], and support vector machine (SVM) [30]. The performances of different classifiers were calculated and compared in the experiment.

LR uses a logistic function, also known as the sigmoid function, as the cost function. The sigmoid function is an S-shaped curve that maps any number to a score between 0.0 and 1.0. In LR, the goal is to predict the probability of a binary outcome based on one or more input features. The model assigns weights and intercepts to each feature, and then uses these weights and intercepts to compute a score for each data point. The score is transformed into a probability value using the logistic function, which maps the score to a value between 0.0 and 1.0.

In ANN, the network is composed of layers of interconnected nodes, or neurons. The input features are fed into the network, and each feature is individually connected to the neurons in the middle hidden layer with different connection strengths, which are represented by weights. The neurons in the hidden layer then process the inputs and pass their outputs to the next layer until the output layer is reached, which produces the final output of the network. The backpropagation calculates the error between the predicted output and the true labels and uses this error to adjust the weights of the connections between the neurons. By iterations, the network is able to learn and improve its predictions.

SEKNN is an ensemble method that utilizes random selection to generate subsets of the original features. By creating multiple models based on these subsets, the approach combines models with different feature sets that provide diverse perspectives on the data, resulting in

better training. This method is particularly useful for k nearest neighbor (KNN) since KNN is sensitive to changes in features. The SEKNN method is based on the using of multiple KNN models, which can help overcome the limitations of a single model trained on the entire feature set.

SVM is used due to its effectiveness in high-dimensional spaces, meaning it can handle data points that have many features or dimensions. SVM works by finding the hyperplane that best separates two classes in the feature space. This hyperplane is selected to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class. In SVM, a kernel function is used to transform the input features into a higher-dimensional space where a linear decision boundary can be identified. By mapping the original feature space to this higher-dimensional space, the algorithm can find a decision boundary that is capable of separating the data points into different classes.

The MSI value of each case was predicted based on the trained models. The resulting probability was used for binary classification using a threshold, where the patient was classified as either having MSI-H or not. To evaluate the generalization ability of the model, five-fold cross-validation was employed. The dataset was divided into five equally sized groups, with each group being used once as a test set while the remaining nine groups were used for training. This process was repeated five times, and the results were averaged to obtain the final performance. Performance indices, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC), were used to evaluate the models' performances. AUC was used to consider the trade-offs between sensitivity and specificity at different thresholds [31].

Image retrieval

A CBIR system can automatically extracts the characteristics of a query image and compare them to the existed target image database to obtain interested images in an objective and rapid way. The performance evaluation of a CBIR system is reliant on its ability to retrieve images in a rank order that corresponds to the similarity between the query image and the images in the database. The ranking is determined by measuring the similarity between the query image and the target images. To measure the effectiveness of the CBIR system, the ground truth relevance images of the targets must be labeled. The ratio of relevant images to retrieved images is used to establish the benchmark for the top k accuracy, which indicates the number of relevant images retrieved in the top k. In the experiment, multiple query images were used to test the CBIR system, and the mean average precision (mAP) was calculated for each top k cutoff [32]. TP means the total relevance of the top k. FP means the total irrelevance of the top k.

$$\mathrm{TP} = \sum_{n=1}^{k} R_n \tag{2}$$

$$\mathrm{FP} = \sum_{n=1}^{k} (1 - R_n) \tag{3}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

$$AP = \frac{1}{|R|} \sum_{k=1}^{|R|} Precision(R_k), mAP = \frac{1}{Q} \sum_{k=1}^{Q} AP_k, Q : query \tag{5}$$

Results

This study enrolled a total of 123 MSI-H tumors and 407 MSS tumors, which comprised 427 MSI-H and 1590 MSS colonoscopy images. Fig 2 demonstrate the flow diagram for the

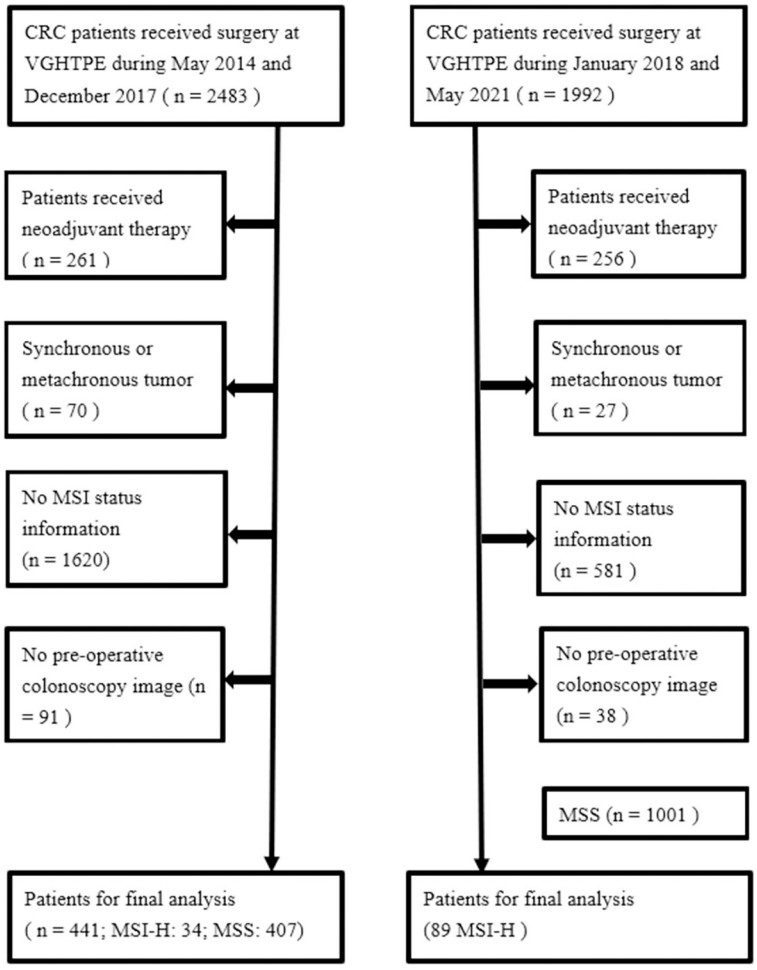

Fig 2. Flow diagram for the patients enrolled in the current study.

https://doi.org/10.1371/journal.pone.0292277.g002

patients enrolled in the current study. Table 1 provides a summary of the clinicopathologic characteristics of MSI-H and MSS tumors. When compared to MSS tumors, MSI-H tumors exhibit significant differences, including earlier staging, a predominant occurrence on the right side, higher rates of poor to undifferentiated grading, increased presence of tumor with ≥50% mucin component, elevated occurrences of lymphovascular invasion (LVI), perineural invasion, and signet ring cell components. Additionally, there is a tendency for these MSI-H tumors to be more prevalent in females (p = 0.066). Fig 3 demonstrate the gross images of MSI tumor and MSS tumor. To conduct the image retrieval process, the collected images were separated into the target image database (80%) and query images (20%). The similarity measurements between the query images and the target image database were based on image characteristics trained from the target image database, and the training involved five-fold cross-validation. Table 2 displays the classification performances of various learning networks, including DenseNet201, which was compared to ViT. Overall, the approaches of ViT were superior to DenseNet201. Among the approaches, combining pre-trained features in SVM outperformed the ways of fine-tuning and training from scratch. Based on the performance comparisons in Tables 3 and 4, SVM was selected as the best machine learning classifier. Fig 4 illustrates the receiver operating characteristic (ROC) curve and AUC value of the best performance achieved by DenseNet201 and ViT.

Table 1. Clinicopathologic profile of all colorectal cancer patients.

Characteristic (number)	MSI-H (123)	(%)	MSS (407)	(%)	p
Age					0.067
Mean ± SD	66.7±15.0		63.9±12.6		
Age ≧50 y/o					0.594
Yes	105	85.4	355	87.2	
No	18	14.6	52	12.8	
Gender					0.066
Male	61	49.6	240	59.0	
Female	62	50.4	167	41.0	
PreOP CEA level (ng/mL)					0.248
Mean ± SD	11.7±27.1		31.7±191.6		
Elevated PreOP CEA≧5 ng/mL					0.221
Yes	41	33.3	169	41.5	
No	82	66.7	237	58.2	
Stage					0.001
I	25	20.3	83	20.4	
II	69	56.1	124	30.5	
III	23	18.7	135	33.2	
IV	6	4.9	65	16.0	
Location					0.001
Right-sided colon	84	68.3	104	25.6	
Left-sided colon	28	22.8	218	53.6	
Rectum	11	8.9	85	20.9	
Grade of differentiation					0.001
Well to moderate	98	79.7	378	92.9	
Poor to undifferentiated	25	20.3	29	7.1	
Mucinous component					0.001
≧50%	25	20.7	19	4.7	
< 50%	96	79.3	384	95.3	
LVI					0.001
Yes	18	15.4	119	31.6	
No	99	84.6	257	68.4	
Perineural invasion					0.005
Yes	4	3.4	47	12.5	
No	113	96.6	329	87.5	
Signet ring cell component					0.001*
Yes	13	11.1	8	2.1	
No	104	88.9	368	97.9	

LVI: Lymphovascular invasion.

*Fisher's Exact Test.

https://doi.org/10.1371/journal.pone.0292277.t001

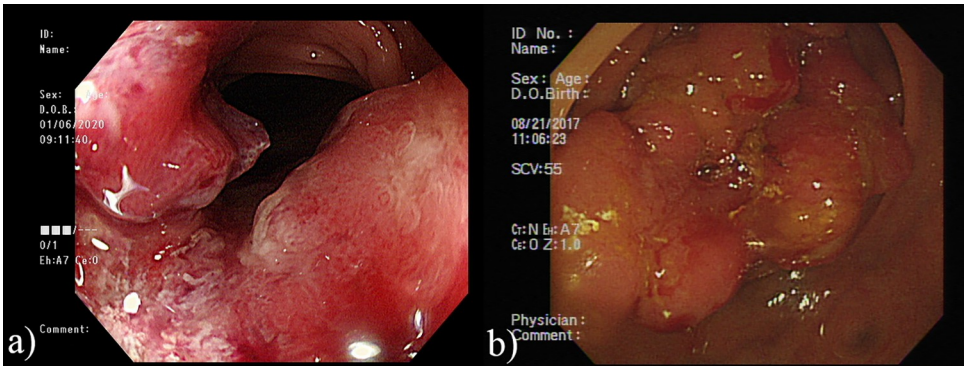

Fig 3. Demonstration of colonoscopy images of colon cancer with different MSI status. (a) MSI-H (b) MSS.

https://doi.org/10.1371/journal.pone.0292277.g003

Table 2. Classification performances of different learning networks.

	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
DenseNet201 (train from scratch)	51%	85%	42%	30%	93%	0.75
DenseNet201 (fine-tune)	77%	64%	80%	48%	89%	0.80
DenseNet201 (pre-train+ML)	80%	47%	89%	55%	86%	0.80
ViT (train from scratch)	69%	57%	72%	42%	87%	0.74
ViT (fine-tune)	81%	49%	89%	56%	86%	0.77
ViT (pre-train+ML)	84%	47%	94%	68%	87%	0.86

https://doi.org/10.1371/journal.pone.0292277.t002

Table 3. Performance indices of combining DenseNet201 features in machine learning classifiers.

Classifier	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
LR	55%	51%	56%	24%	81%	0.53
ANN	79%	50%	87%	51%	87%	0.74
SEKNN	78%	43%	87%	48%	85%	0.67
SVM	80%	47%	89%	55%	86%	0.80

https://doi.org/10.1371/journal.pone.0292277.t003

Table 4. Performance indices of combining ViT features in machine learning classifiers.

Classifier	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
LR	72%	52%	78%	39%	86%	0.69
ANN	81%	52%	89%	56%	87%	0.80
SEKNN	82%	46%	92%	62%	86%	0.77
SVM	84%	47%	94%	68%	87%	0.86

https://doi.org/10.1371/journal.pone.0292277.t004

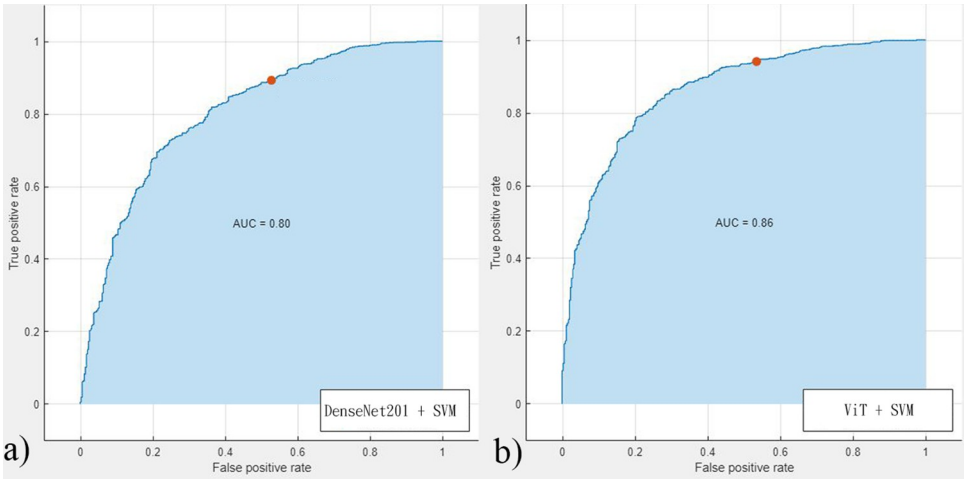

Fig 4. Receiver operating characteristic curve of different features combined in SVM to MSI prediction. (a) DenseNet201 features (b) ViT features.

https://doi.org/10.1371/journal.pone.0292277.g004

Table 5. Retrieval performances of different learning networks.

	Top-10
DenseNet201 (pre-train)	0.59
DenseNet201 (fine-tune)	0.79
ViT (pre-train)	0.72
ViT (fine-tune)	0.81

https://doi.org/10.1371/journal.pone.0292277.t005

Table 5 lists the top-10 image retrieval result. Based on the trained features in the classifications, ViT features had mAP = 0.81 which outperformed DenseNet201 features having mAP = 0.79. ViT still have better result of image retrieval than DenseNet201. Fig 5 shows the top-10 image retrieval results of the queries of MSI-H and MSS images based on ViT fine-tune features.

Discussion

Apart from certain clinical and pathologic characteristics, MSI status can also serve as a biomarker for anticipating the response to particular therapies. MSI status is particularly useful in

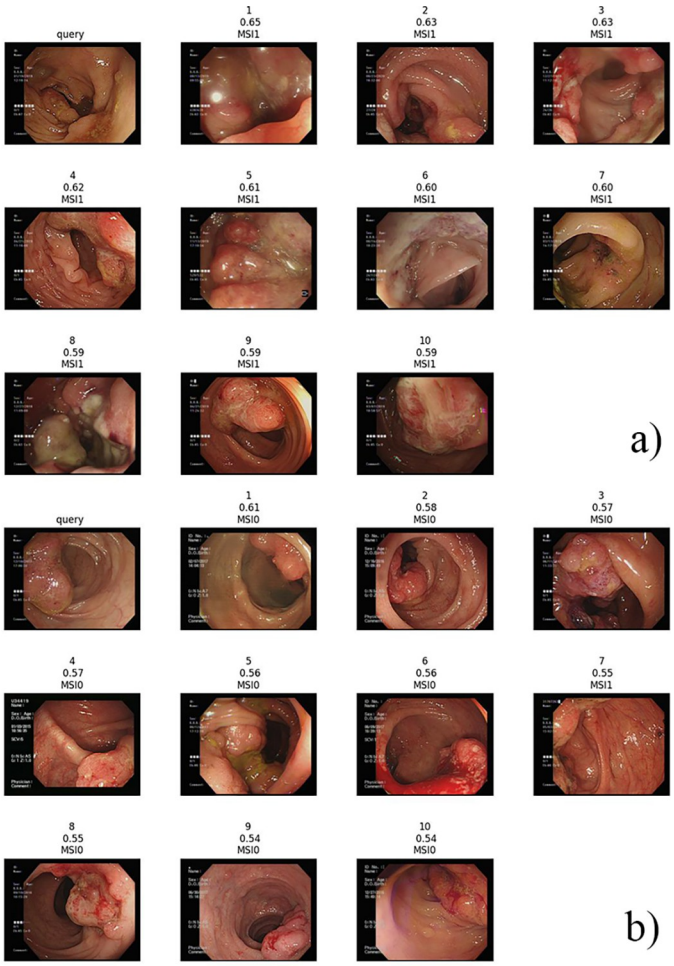

Fig 5. Retrieved top 10 colonoscopy images using ViT fine-tune features (a) the query of MSI-H (b) the query of MSS.

https://doi.org/10.1371/journal.pone.0292277.g005

guiding treatment decisions for stage II colorectal cancers, and MSI-high tumors have demonstrated a high degree of responsiveness to immune checkpoint inhibitors (ICI) [4]. As a result, MSI testing is a recommended component of the standard of care for all patients diagnosed with CRC [33]. Numerous studies have employed deep learning for the prediction of MSI status, with the majority of them using histopathologic images [34–44]. Additionally, radiomic signatures have been demonstrated to predict genetic alterations [45–50]. The goal of this study is to utilize deep learning models for the prediction of MSI status through the image retrieval of colonoscopy images of primary tumors. Colonoscopy is a widely used and convenient method for evaluating tumor status in clinical examinations. It has the advantage of presenting real-time image information, which can reduce time and costs compared to using histopathologic images and radiomic signatures. To our knowledge, this is the first study utilizing colonoscopy images for the prediction of MSI status in CRC.

By utilizing various training methods, the ViT architectures demonstrated superior performance compared to DenseNet201, which is another type of CNN architecture. One possible reason for this is that the attention mechanism used in ViT enables it to better analyze tumors and surrounding tissues through similarity measurements between image patches. SVM was utilized in this study to combine deep learning features, and the receiver operating characteristic curve for DenseNet201 and ViT features were 0.80 and 0.86, respectively. The NPVs were approximately 86–87%, suggesting a potential reduction in the expenses associated with MMR testing in routine clinical practice. Furthermore, leveraging pre-trained features enhances the efficiency and practicality of using deep learning.

The current study has several limitations. Firstly, the number of patients enrolled is limited, and only a small percentage (7.7%) of patients in the initial cohort had MSI-H tumors. To address this, we enrolled another MSI cohort, but this may impact the clinical applicability of the current model in real-world settings. The next step should be to perform further external validation using multicenter patients to generalize the clinical utility of the model. As more data becomes available, both training from scratch and fine-tuning methods are likely to yield better results. However, the increased amount of data also means that computational resources and training time will be more demanding. Secondly, we omitted the inclusion of family history, a crucial parameter in clinical Lynch syndrome assessment. Additionally, we did not incorporate clinicopathological features like age, staging, tumor location, and tumor differentiation into our predictive model, even though they may have predictive value for MSI status. A future study should aim to integrate additional diagnostic information to improve the accuracy of the model. The use of image retrieval to distinguish between various genetic backgrounds, such as sporadic or hereditary MSI-H, has not yet been fully explored. This aspect would be a subject of further investigation in our study. Third, further MLH1 methylation status and/or genetic testing, such as, next-generation sequencing is needed in patients with the loss of one or more MMR markers to differentiate sporadic MSI-H patients from Lynch syndrome. However, we didn't have this information. How the background of MSI-H tumor affects the decision of ViT features remained elusive. To tackle this concern, it is necessary for us to initiate a prospective study to acquire this information and establish more robust evidence for practical application. In addition, we utilized IHC for MMR proteins to detect the MSI status, rather than PCR-based. IHC for MMR proteins is the initial step to screen for Lynch syndrome [51]. Previous studies had proven to reveal a high coincidence rate of the two methods for detecting MSI status up to more than 90% [52–54]. IHC and the PCR method had high consistency in MSI status. Compared with PCR, the IHC method is the preferred single screening test and is more economical and more convenient for clinical operations. While there have been studies aimed at distinguishing MSI-H CRCs from MSS CRCs based on differences in their histopathological and pathomorphological characteristics, such as the prevalence

of mucinous adenocarcinoma and aggressive histological features in MSI-H tumors [55], there is currently no evidence suggesting that colonoscopy can grossly differentiate these features. However, it remains uncertain whether colonoscopy can successfully identify these distinct characteristics.

This study proposed the utilization of pre-trained ViT features to achieve a rapid and substantial MSI classification outcome, which is crucial for clinical applications. In addition to a numerical value indicating the MSI classification, the study also presented additional evidence through an image retrieval method, similar to the approach suggested by Komura et al. [19] in their study. Through a top-10 image display, physicians can observe whether images with similar compositions possess the same MSI classification, thereby increasing confidence in the decision-making. A future research direction could be to investigate the image differences among various MSI statuses. Similar to previous studies utilizing radiomic signatures [45–50], discovering more evidence from a larger pool of colonoscopy images would enhance the results' credibility and enable widespread use in clinical settings.

Conclusions

CRCs characterized by MSI-H have a more favorable prognosis. In this study, to achieve a rapid and cost-effective outcome with limited cases, we proposed the use of ViT features extracted from colonoscopy images. By combining pre-trained features in SVM, the classification results exhibited 84% accuracy and an AUC of 0.86. Compared to conventional CNNs, ViT based on the patch embedding and self-attention mechanism addresses the issue of limited receptive fields and gradient disappearance. The experiment also presented a CBIR result, with a mAP of 0.81, illustrating that images with similar content have the same MSI status. This proposed image classification and retrieval procedure, in conjunction with colonoscopy examination, makes MSI prediction more accessible and convenient for clinical use.

Supporting information

S1 Checklist. STROBE statement—Checklist of items that should be included in reports of observational studies.
(DOCX)

S1 File. Supporting information zip file contains image features.
(ZIP)

Acknowledgments

Jen-Kou Lin, Tzu-Chen Lin, Wei-Shone Chen, Shung-Haur Yang, Shih-Ching Chang, Huann-Sheng Wang, Yuan-Tzu Lan, Hung-Hsin Lin, Sheng-Chieh Huang, and Hou-Hsuan Cheng provided and cared for the study patients.

Author Contributions

Conceptualization: Chung-Ming Lo, Jeng-Kai Jiang.

Data curation: Chun-Chi Lin.

Formal analysis: Chun-Chi Lin.

Funding acquisition: Chung-Ming Lo.

Methodology: Chung-Ming Lo.

Project administration: Jeng-Kai Jiang.

Resources: Chun-Chi Lin.

Validation: Chun-Chi Lin.

Writing – original draft: Chung-Ming Lo, Chun-Chi Lin.

Writing – review & editing: Chung-Ming Lo, Jeng-Kai Jiang.

References

1. Sung H. et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," (in eng), *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021. https://doi.org/10.3322/caac.21660 PMID: 33538338
2. Boland C. R. and Goel A., "Microsatellite instability in colorectal cancer," (in eng), *Gastroenterology*, vol. 138, no. 6, pp. 2073–2087.e3, Jun 2010. https://doi.org/10.1053/j.gastro.2009.12.064 PMID: 20420947
3. Lin C. C. et al., "The prognostic role of microsatellite instability, codon-specific KRAS, and BRAF mutations in colon cancer," (in eng), *J Surg Oncol*, vol. 110, no. 4, pp. 451–7, Sep 2014. https://doi.org/10.1002/jso.23675 PMID: 24964758
4. Vilar E. and Gruber S. B., "Microsatellite instability in colorectal cancer-the stable evidence," (in eng), *Nat Rev Clin Oncol*, vol. 7, no. 3, pp. 153–62, Mar 2010. https://doi.org/10.1038/nrclinonc.2009.237 PMID: 20142816
5. Ludford K. et al., "Neoadjuvant Pembrolizumab in Localized Microsatellite Instability High/Deficient Mismatch Repair Solid Tumors," (in eng), *J Clin Oncol*, p. Jco2201351, Jan 9 2023. https://doi.org/10.1200/JCO.22.01351 PMID: 36623241
6. Trojan J. et al., "Complete Pathological Response After Neoadjuvant Short-Course Immunotherapy with Ipilimumab and Nivolumab in Locally Advanced MSI-H/dMMR Rectal Cancer," (in eng), *Oncologist*, vol. 26, no. 12, pp. e2110–e2114, Dec 2021. https://doi.org/10.1002/onco.13955 PMID: 34431576
7. Cercek A. et al., "PD-1 Blockade in Mismatch Repair-Deficient, Locally Advanced Rectal Cancer," (in eng), *N Engl J Med*, vol. 386, no. 25, pp. 2363–2376, Jun 23 2022. https://doi.org/10.1056/NEJMoa2201445 PMID: 35660797
8. Rodriguez-Bigas M. A. et al., "A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines," (in eng), *J Natl Cancer Inst*, vol. 89, no. 23, pp. 1758–62, Dec 3 1997. https://doi.org/10.1093/jnci/89.23.1758 PMID: 9392616
9. Benson A. B. et al., "Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology," (in eng), *J Natl Compr Canc Netw*, vol. 19, no. 3, pp. 329–359, Mar 2 2021. https://doi.org/10.6004/jnccn.2021.0012 PMID: 33724754
10. Berg K. D., Glaser C. L., Thompson R. E., Hamilton S. R., Griffin C. A., and Eshleman J. R., "Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction," (in eng), *J Mol Diagn*, vol. 2, no. 1, pp. 20–8, Feb 2000. https://doi.org/10.1016/S1525-1578(10)60611-3 PMID: 11272898
11. Torshizi Esfahani A., Seyedna S. Y., Nazemalhosseini Mojarad E., Majd A., and Asadzadeh Aghdaei H., "MSI-L/EMAST is a predictive biomarker for metastasis in colorectal cancer patients," (in eng), *J Cell Physiol*, vol. 234, no. 8, pp. 13128–13136, Aug 2019. https://doi.org/10.1002/jcp.27983 PMID: 30549036
12. Boland C. R. et al., "A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer," (in eng), *Cancer Res*, vol. 58, no. 22, pp. 5248–57, Nov 15 1998. PMID: 9823339
13. Cheng D. T. et al., "Comprehensive detection of germline variants by MSK-IMPACT, a clinical diagnostic platform for solid tumor molecular oncology and concurrent cancer predisposition testing," (in eng), *BMC Med Genomics*, vol. 10, no. 1, p. 33, May 19 2017. https://doi.org/10.1186/s12920-017-0271-4 PMID: 28526081
14. Alexander J., Watanabe T., Wu T. T., Rashid A., Li S., and Hamilton S. R., "Histopathological identification of colon cancer with microsatellite instability," (in eng), *Am J Pathol*, vol. 158, no. 2, pp. 527–35, Feb 2001. https://doi.org/10.1016/S0002-9440(10)63994-6 PMID: 11159189
15. Echle A. et al., "Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning," *Gastroenterology*, vol. 159, no. 4, pp. 1406–1416. e11, 2020. https://doi.org/10.1053/j.gastro.2020.06.021 PMID: 32562722

16. Yamashita R. et al., "Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study," *The Lancet Oncology*, vol. 22, no. 1, pp. 132–141, 2021. https://doi.org/10.1016/S1470-2045(20)30535-0 PMID: 33387492

17. Chang X. et al., "Predicting colorectal cancer microsatellite instability with a self-attention-enabled convolutional neural network," *Cell Reports Medicine*, 2023. https://doi.org/10.1016/j.xcrm.2022.100914 PMID: 36720223

18. T. Peng, M. Boxberg, W. Weichert, N. Navab, and C. Marr, "Multi-task learning of a deep k-nearest neighbour network for histopathological image classification and retrieval," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*: *22nd International Conference*, *Shenzhen*, *China*, *October 13–17*, *2019*, *Proceedings*, *Part I 22*, 2019: Springer, pp. 676–684.

19. Komura D. et al., "Universal encoding of pan-cancer histology by deep texture representations," *Cell Reports*, vol. 38, no. 9, p. 110424, 2022. https://doi.org/10.1016/j.celrep.2022.110424 PMID: 35235802

20. Dosovitskiy A. et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

21. Lo C.-M. and Hung P.-H., "Assessing Ischemic Stroke with Convolutional Image Features in Carotid Color Doppler," *Ultrasound in Medicine & Biology*, vol. 47, no. 8, pp. 2266–2276, 2021. https://doi.org/10.1016/j.ultrasmedbio.2021.03.038 PMID: 34001404

22. Liang C.-W., Fang P.-W., Huang H.-Y., and Lo C.-M., "Deep Convolutional Neural Networks Detect Tumor Genotype from Pathological Tissue Images in Gastrointestinal Stromal Tumors," *Cancers*, vol. 13, no. 22, p. 5787, 2021. https://doi.org/10.3390/cancers13225787 PMID: 34830948

23. Lo C.-M., Yeh Y.-H., Tang J.-H., Chang C.-C., and Yeh H.-J., "Rapid Polyp Classification in Colonoscopy Using Textural and Convolutional Features," in *Healthcare*, 2022, vol. 10, no. 8: MDPI, p. 1494. https://doi.org/10.3390/healthcare10081494 PMID: 36011151

24. Lo C.-M. et al., "Modeling the Survival of Colorectal Cancer Patients Based on Colonoscopic Features in a Feature Ensemble Vision Transformer," *Computerized Medical Imaging and Graphics*, p. 102242, 2023. https://doi.org/10.1016/j.compmedimag.2023.102242 PMID: 37172354

25. Lo C.-M. and Lai K.-L., "Deep learning-based Assessment of Knee Septic Arthritis using Transformer Features in Sonographic Modalities," *Computer Methods and Programs in Biomedicine*, p. 107575, 2023. https://doi.org/10.1016/j.cmpb.2023.107575 PMID: 37148635

26. Deng J., Dong W., Socher R., Li L.-J., Li K., and Fei-Fei L., "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009: Ieee, pp. 248–255.

27. Lam C. S., Koon H. K., Chung V. C.-H., and Cheung Y. T., "A public survey of traditional, complementary and integrative medicine use during the COVID-19 outbreak in Hong Kong," *PloS one*, vol. 16, no. 7, p. e0253890, 2021. https://doi.org/10.1371/journal.pone.0253890 PMID: 34197523

28. Alkadri S. et al., "Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure," *Computers in Biology and Medicine*, vol. 136, p. 104770, 2021. https://doi.org/10.1016/j.compbiomed.2021.104770 PMID: 34426170

29. Nosrati V. and Rahmani M., "An ensemble framework for microarray data classification based on feature subspace partitioning," *Computers in Biology and Medicine*, vol. 148, p. 105820, 2022. https://doi.org/10.1016/j.compbiomed.2022.105820 PMID: 35872409

30. Xia J. et al., "Performance optimization of support vector machine with oppositional grasshopper optimization for acute appendicitis diagnosis," *Computers in Biology and Medicine*, vol. 143, p. 105206, 2022. https://doi.org/10.1016/j.compbiomed.2021.105206 PMID: 35101730

31. Chen J. et al., "Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet," *Computer methods and programs in biomedicine*, vol. 200, p. 105878, 2021. https://doi.org/10.1016/j.cmpb.2020.105878 PMID: 33308904

32. Smith J. R., "Quantitative assessment of image retrieval effectiveness," *Journal of the American Society for Information Science and Technology*, vol. 52, no. 11, pp. 969–979, 2001.

33. N. C. C. Network. (Mar, 14th). *Colon Cancer. Version 3.2022* [Online]. Available: https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf.

34. Yamashita R. et al., "Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study," (in eng), *Lancet Oncol*, vol. 22, no. 1, pp. 132–141, Jan 2021. https://doi.org/10.1016/S1470-2045(20)30535-0 PMID: 33387492

35. Park J. H. et al., "Artificial Intelligence for Predicting Microsatellite Instability Based on Tumor Histomorphology: A Systematic Review," (in eng), *Int J Mol Sci*, vol. 23, no. 5, Feb 23 2022. https://doi.org/10.3390/ijms23052462 PMID: 35269607

36. Fujii S. et al., "Rapid Screening Using Pathomorphologic Interpretation to Detect BRAFV600E Mutation and Microsatellite Instability in Colorectal Cancer," (in eng), *Clin Cancer Res*, vol. 28, no. 12, pp. 2623–2632, Jun 13 2022. https://doi.org/10.1158/1078-0432.CCR-21-4391 PMID: 35363302

37. Qiu W. et al., "Evaluating the Microsatellite Instability of Colorectal Cancer Based on Multimodal Deep Learning Integrating Histopathological and Molecular Data," (in eng), *Front Oncol*, vol. 12, p. 925079, 2022. https://doi.org/10.3389/fonc.2022.925079 PMID: 35865460

38. Lou J. et al., "PPsNet: An improved deep learning model for microsatellite instability high prediction in colorectal cancer from whole slide images," (in eng), *Comput Methods Programs Biomed*, vol. 225, p. 107095, Oct 2022. https://doi.org/10.1016/j.cmpb.2022.107095 PMID: 36057226

39. Leiby J. S., Hao J., Kang G. H., Park J. W., and Kim D., "Attention-based multiple instance learning with self-supervision to predict microsatellite instability in colorectal cancer from histology whole-slide images," (in eng), *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2022, pp. 3068–3071, Jul 2022. https://doi.org/10.1109/EMBC48229.2022.9871553 PMID: 36085965

40. Chang X. et al., "Predicting colorectal cancer microsatellite instability with a self-attention-enabled convolutional neural network," (in eng), *Cell Rep Med*, vol. 4, no. 2, p. 100914, Feb 21 2023. https://doi.org/10.1016/j.xcrm.2022.100914 PMID: 36720223

41. Guo B., Li X., Yang M., Jonnagaddala J., Zhang H., and Xu X. S., "Predicting microsatellite instability and key biomarkers in colorectal cancer from H&E-stained images: achieving state-of-the-art predictive performance with fewer data using Swin Transformer," (in eng), *J Pathol Clin Res*, Feb 1 2023.

42. Lee S. H., Song I. H., and Jang H. J., "Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer," (in eng), *Int J Cancer*, vol. 149, no. 3, pp. 728–740, Aug 1 2021. https://doi.org/10.1002/ijc.33599 PMID: 33851412

43. Hildebrand L. A., Pierce C. J., Dennis M., Paracha M., and Maoz A., "Artificial Intelligence for Histology-Based Detection of Microsatellite Instability and Prediction of Response to Immunotherapy in Colorectal Cancer," (in eng), *Cancers (Basel)*, vol. 13, no. 3, Jan 21 2021. https://doi.org/10.3390/cancers13030391 PMID: 33494280

44. Echle A. et al., "Artificial intelligence for detection of microsatellite instability in colorectal cancer-a multicentric analysis of a pre-screening tool for clinical application," (in eng), *ESMO Open*, vol. 7, no. 2, p. 100400, Apr 2022. https://doi.org/10.1016/j.esmoop.2022.100400 PMID: 35247870

45. Wu X. et al., "Deep Learning Features Improve the Performance of a Radiomics Signature for Predicting KRAS Status in Patients with Colorectal Cancer," (in eng), *Acad Radiol*, vol. 27, no. 11, pp. e254–e262, Nov 2020. https://doi.org/10.1016/j.acra.2019.12.007 PMID: 31982342

46. Yang L. et al., "Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer?," (in eng), *Eur Radiol*, vol. 28, no. 5, pp. 2058–2067, May 2018. https://doi.org/10.1007/s00330-017-5146-8 PMID: 29335867

47. He K., Liu X., Li M., Li X., Yang H., and Zhang H., "Noninvasive KRAS mutation estimation in colorectal cancer using a deep learning method based on CT imaging," (in eng), *BMC Med Imaging*, vol. 20, no. 1, p. 59, Jun 1 2020. https://doi.org/10.1186/s12880-020-00457-4 PMID: 32487083

48. Hu J. et al., "Predicting Kirsten Rat Sarcoma Virus Gene Mutation Status in Patients With Colorectal Cancer by Radiomics Models Based on Multiphasic CT," (in eng), *Front Oncol*, vol. 12, p. 848798, 2022. https://doi.org/10.3389/fonc.2022.848798 PMID: 35814386

49. Ying M. et al., "Development and validation of a radiomics-based nomogram for the preoperative prediction of microsatellite instability in colorectal cancer," (in eng), *BMC Cancer*, vol. 22, no. 1, p. 524, May 9 2022. https://doi.org/10.1186/s12885-022-09584-3 PMID: 35534797

50. Pei Q. et al., "Pre-treatment CT-based radiomics nomogram for predicting microsatellite instability status in colorectal cancer," (in eng), *Eur Radiol*, vol. 32, no. 1, pp. 714–724, Jan 2022. https://doi.org/10.1007/s00330-021-08167-3 PMID: 34258636

51. Chen W., Swanson B. J., and Frankel W. L., "Molecular genetics of microsatellite-unstable colorectal cancer for pathologists," (in eng), *Diagn Pathol*, vol. 12, no. 1, p. 24, Mar 4 2017. https://doi.org/10.1186/s13000-017-0613-8 PMID: 28259170

52. Hissong E., Crowe E. P., Yantiss R. K., and Chen Y. T., "Assessing colorectal cancer mismatch repair status in the modern era: a survey of current practices and re-evaluation of the role of microsatellite instability testing," (in eng), *Mod Pathol*, vol. 31, no. 11, pp. 1756–1766, Nov 2018. https://doi.org/10.1038/s41379-018-0094-7 PMID: 29955148

53. Chen M. L. et al., "Comparison of microsatellite status detection methods in colorectal carcinoma," (in eng), *Int J Clin Exp Pathol*, vol. 11, no. 3, pp. 1431–1438, 2018. PMID: 31938240

54. Ye M., Ru G., Yuan H., Qian L., He X., and Li S., "Concordance between microsatellite instability and mismatch repair protein expression in colorectal cancer and their clinicopathological characteristics: a retrospective analysis of 502 cases," (in eng), *Front Oncol*, vol. 13, p. 1178772, 2023. https://doi.org/10.3389/fonc.2023.1178772 PMID: 37427134

55. Mei W. J., Mi M., Qian J., Xiao N., Yuan Y., and Ding P. R., "Clinicopathological characteristics of high microsatellite instability/mismatch repair-deficient colorectal cancer: A narrative review," (in eng), *Front Immunol*, vol. 13, p. 1019582, 2022. https://doi.org/10.3389/fimmu.2022.1019582 PMID: 36618386