Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The application of blood serum ATR-FTIR spectroscopy for the identification of Colorectal cancer

wanli yang
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
nan pang
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
jia shi
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
wei zhang
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
Jin Han
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
ao song
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
Chao yang
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
Jun Leng
Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences
Degao Zhang
decaux@163.com

Chongming Hospital Affiliated to Shanghai University of Medicine and Health Sciences

Research Article

Keywords: Colorectal cancer, Blood serum, ATR-FTIR spectroscopy, Principal component analysis, Machine learning

Posted Date: July 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10109264/v1

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

Additional Declarations: No competing interests reported.

Abstract

Colorectal cancer (CRC) poses a substantial public health challenge, highlighting the need for advancements in early diagnostic techniques. This study utilized attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy to perform a multivariate analysis of human serum. The research involved the analysis of blood samples from 97 individuals diagnosed with colorectal cancer and 98 healthy volunteers. Principal component analysis (PCA) was employed to interpret the infrared spectral data obtained from the serum samples. Specific spectral bands with intensity variations between the two groups were identified. The infrared spectral ranges of 3500 ~ 3000 cm^{-1} and 1600 ~ 1500 cm^{-1} demonstrated significant diagnostic potential for colorectal cancer, likely due to differences in protein conformation and nucleic acids. By applying machine learning algorithms to differentiate between colorectal cancer patients (n = 97) and healthy controls (n = 98), we achieved a sensitivity of up to 88.9%, a specificity of 92.9% and a accuracy of 90.9%. Receiver operating characteristic (ROC) analysis yielded an area under the curve (AUC) of 0.911. These findings underscore the potential of our serum-based ATR-FTIR spectroscopy method as a straightforward, minimally invasive, and reliable diagnostic tool for the detection of Colorectal cancer.

Introduction

Colorectal cancer (CRC) is the third most common malignant tumor in terms of incidence. The American Cancer Society's 2023 report estimates about 153,020 new colorectal cancer cases and 52,550 deaths [1]. Early-stage colorectal cancer patients have a 5-year survival rate nearing 90%, whereas those diagnosed at advanced stages have a rate of only 13.1%. Thus, enhancing early screening techniques is crucial for lowering CRC mortality [2].

Current diagnostic modalities encompass colonoscopy, sigmoidoscopy, serum markers, carcinoembryonic antigen (CEA) testing, fecal tests, and various imaging techniques, among others [2–4]. Despite the relatively low specificity of CEA, it remains valuable for the detection of tumor recurrence [5]. Fecal tests, including the guaiac-based fecal occult blood test (gFOBT) and the fecal immunochemical test (FIT), are employed for the screening of early asymptomatic colorectal cancer [2]. Nonetheless, a comprehensive examination necessitates colonoscopy, and both colonoscopy and pathological examination are frequently recommended for secondary screening [6]. Imaging modalities such as computed tomography (CT) are employed in the detection of colorectal cancer; however, this technique necessitates bowel preparation, the oral administration of contrast agents, and involves exposure to ionizing radiation [7]. In instances of abnormal findings, a follow-up colonoscopy remains essential. Beyond conventional diagnostic approaches, there is an imperative to develop novel methodologies that are straightforward, objective, and non-invasive. ATR-FTIR spectroscopy has emerged as a promising alternative due to its high sensitivity, non-destructive nature, and relatively low cost [8].

ATR-FTIR spectroscopy has emerged as a promising technique in cancer diagnostic testing, and it enables the measurement of the fingerprint spectra of compounds at the molecular level [9–11]. This technique takes advantage of the conformational changes or content variations in the bonding and vibrational patterns of functional groups within biomolecules (such as proteins, nucleic acids, lipids and carbohydrates) that occur during the development of cancer [12]. Therefore, ATR - FTIR can effectively distinguish between tumor samples and normal samples [12]. Numerous studies have concentrated on the application of FTIR spectroscopy for the detection of CRC in tissue samples. Cai et al. observed that the integrated absorbance of the nucleic acid band and the amide I band was minimal in healthy colonic tissues, whereas the lipid band exhibited the lowest absorbance in CRC tissues [13]. Geyik et al. observed an elevation in the concentrations of lipid-to-protein and nucleic acid-to-protein ratios in CRC tissues, accompanied by a reduction in the amide-I to amide-II ratio [14]. Similarly, Li et al. reported a decrease in lipid content alongside an increase in protein content within cancerous tissues, noting alterations in the intensities of peaks associated with proteins and lipids [15]. The observed spectral variations may be associated with the heightened energy demands and biosynthetic activities of cancerous tissues, suggesting that these spectral parameters could serve as potential diagnostic biomarkers. Recent studies have employed serum spectral analysis to differentiate between individuals with colorectal cancer and healthy group. Guleken et al. observed that, relative to the healthy group, the absorbance of the majority of peaks in the serum ATR-FTIR spectroscopy of CRC patients exhibited a decrease, with the exception of the peak at 1740 cm^{-1}, corresponding to lipid vibration, which showed an increase. Through the application of second derivative calculations in the protein region and deconvolution analysis of the amide I region, distinct variations in protein structure and the amide I band were identified between non-cancerous and colorectal cancer samples [16]. These findings indicate that ATR-FTIR spectroscopy holds substantial promise for the diagnosis of CRC.

Future research should focus on increasing sample sizes and incorporating a broader range of machine learning methodologies to enhance diagnostic accuracy. This study employed ATR-FTIR spectroscopy imaging technology in conjunction with machine learning techniques to analyze serum samples from patients with CRC and healthy individuals. By applying multivariate data analysis to detect spectral variations, we successfully examined 195 human serum samples, enabling the differentiation between CRC patients and healthy individuals. This approach offers novel methodologies and insights for the diagnosis of colorectal diseases.

Materials and methods

Ethics statement and consent to participate

This study was conducted with the authorization of Chongming Hospital, affiliated with Shanghai University of Medicine and Health Sciences, following approval by Chongming Hospital, affiliated with Shanghai University of Medicine and Health Sciences ethics committee (Approval number: xxx). In accordance with the Declaration of Helsinki, written informed consent was obtained from each participant. Clinical trial number: not applicable.

Sample and Sample Preparation

Serum samples were obtained from 97 patients diagnosed with colorectal cancer and 98 healthy blood donors, courtesy of the Laboratory Department at xxx. A volume of 1 ml of fasting peripheral venous blood was collected from both the colorectal cancer patients and the healthy donors, who served as the control group, using non-anticoagulant blood collection tubes. The blood samples were allowed to coagulate naturally at room temperature for 20 minutes via endogenous coagulation factors. Subsequently, the samples underwent centrifugation at 3500 rpm for 10 minutes to facilitate serum separation. The isolated serum was then transferred to cryotubes and stored at -80°C. Prior to analysis, the serum samples were thawed. The colorectal cancer patients had a mean age of 65.43 years, with an age range of 31 to 92 years, comprising 54 males and 43 females. The inclusion criteria for this study were as follows: (1) all patients had undergone initial treatment; (2) comprehensive clinical data were accessible; (3) colorectal cancer was confirmed in all patients via surgical intervention or pathological biopsy; and (4) there was an absence of other tumors within the digestive system. The exclusion criteria were: (1) a history of radiotherapy or chemotherapy; and (2) the presence of other primary malignant tumors. Additionally, a control group consisting of 98 healthy individuals was selected. These individuals had undergone physical examinations at our hospital during the same period and demonstrated normal values in all blood tests. The average age of the healthy blood donors was 51.34 years, with ages ranging from 29 to 80 years. The control group included 51 males and 47 females. All participants in this study were provided with comprehensive information and gave informed consent by signing the consent form.

Spectral Irradiation Analysis

The absorbance mode of the ATR-FTIR spectrometer (Bruker IFS 66v), situated at the National Synchrotron Radiation Laboratory in xxx, was employed to examine the spectral differences in serum samples from individuals diagnosed with colorectal cancer compared to those from unaffected individuals. A 10 μL aliquot of each serum sample was uniformly distributed onto a barium fluoride crystal (the ATR crystal) and subsequently allowed to air-dry for 10 minutes at ambient temperature. The spectral recording resolution for each sample was set at 4 cm^{-1}, with an imaging spot size of 100 μm × 100 μm. A total of 128 cumulative scans were performed for spectral acquisition. OPUS software was utilized to process the spectral measurements and imaging data.

Spectral preprocessing and data analysis

The OPUS software was utilized to perform essential calibration and normalization procedures on all spectra. Baseline correction was achieved using the rubber band method, while normalization was conducted using the vector method. Data analysis was executed with RStudio version 4.4.0. Average spectra were generated, and the standard deviation was calculated from the cancer samples and the raw spectra of the control samples. The spectral absorbance value was quantified by calculating the area

under the peak curve. This process involved transforming the infrared spectral curve into a discrete set of data points, segmenting these data points according to specified wavenumber ranges, and subsequently calculating the area for each segment. To enhance the spectral bands, reduce background noise, and reveal the intrinsic biochemical characteristics of individual serum samples from colorectal cancer patients and healthy donors, the second derivative of their respective spectra was computed. The second derivative, a common mathematical operation applied to infrared spectra, facilitates the identification of broad absorption bands. The second derivative, defined as the derivative of the first derivative, quantifies the rate of change of the first derivative and provides insights into the concavity and convexity of a function's graph. This mathematical operation is frequently applied to infrared spectra to identify broad absorption bands, with each band manifesting as a more pronounced minimum in the second derivative. This enhanced representation facilitates the detection of spectral changes. Following the preprocessing of the spectra and the calculation of their second derivatives, principal component analysis (PCA) was conducted. PCA, an unsupervised learning technique, was utilized to uncover patterns and relationships within the data autonomously and to identify directions of maximal variance [17]. The score plot of the principal components can be utilized to identify patterns or clusters present in the dataset.

Classification Models and ROC Analysis

The study employed k-Nearest Neighbors (kNN), Gaussian Process Regression (GPR), and Support Vector Machine (SVM) with a linear kernel as classification methodologies to differentiate the serum of patients with Colorectal cancer from that of healthy donors. Receiver Operating Characteristic (ROC) analysis was subsequently performed to assess the diagnostic performance across various probability and diagnostic thresholds. The initial dataset was randomly divided into a training set, comprising 70 cases of Colorectal cancer and 70 healthy donors, and a validation set, consisting of 27 cases of Colorectal cancer and 28 healthy donors. A five-fold cross-validation approach was applied to the training dataset. The optimized model was then retrained using the entire training dataset and utilized to predict outcomes on the test set, enabling the calculation of sensitivity, specificity, accuracy, and the area under the curve (AUC) metrics. The definitions of these metrics are as follows:

Sensitivity

The proportion of true positive samples correctly identified by the model among all actual positive samples.

Sensitivity = $\frac{\mathbf{TP}}{\mathbf{TP}+\mathbf{FN}}$

Specificity

The proportion of true negative samples correctly identified by the model among all actual negative samples.

Specificity = $\frac{TN}{TN+FP}$

Accuracy

The proportion of correctly classified samples among all samples.

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

Area Under the Curve (AUC)

The Receiver Operating Characteristic (ROC) curve provides a comprehensive evaluation of the model's performance across various classification thresholds. The AUC value, which ranges from 0 to 1, quantifies the area beneath the ROC curve, with values approaching 1 indicative of superior classification capability. In this context, TP (True Positive) refers to true positives, TN (True Negative) to true negatives, FP (False Positive) to false positives, and FN (False Negative) to false negatives.

Statistical analysis

Data are presented as mean ± standard deviation. Statistical differences between the values of two groups were evaluated using the unpaired Student's t-test (two-tailed), with a significance threshold set at $P < 0.05$.

Results

Mean spectra and band distribution

The implementation of Fourier Transform Infrared Spectroscopy (FTIR) technology enabled a detailed analysis of spectral variations in serum samples from patients diagnosed with colorectal cancer compared to those from healthy individuals. The vibration modes of functional groups responsible for the infrared spectral absorption peaks have been identified and are detailed in Table 1. Figure 1A presents the average spectral profiles of 97 colorectal cancer patients and 98 healthy controls, alongside the mean differential spectra between these two groups. Figure 1B shows the average spectra and their corresponding standard deviations for each cohort. The application of the second derivative to the extended spectral region, derived from the initial absorption spectra of serum, facilitates the identification of differences in the second derivative between the two groups, as shown in Fig. 2. Notable differences between the colorectal cancer group and the healthy group are observed in the spectral regions of 1420 ~ 1380 cm^{-1}, 1600 ~ 1500 cm^{-1}, 1700 ~ 1600 cm^{-1}, 2990 ~ 2950 cm^{-1} and 3500 ~ 3000 cm^{-1}, as depicted in Fig. 1A and 1B. Compared to healthy individuals, patients with colorectal cancer exhibit a shift towards shorter wavelengths in the peak position and a decrease in absorbance within the 3500–3000 cm^{-1} range (Tables 2 and 3). This range corresponds to the vibrational motions associated with the stretching and bending of O-H and N-H bonds, primarily indicative of proteins or water [18]. Additionally, a notable shift towards higher frequencies, referred to as a blue shift, and a decrease in

absorbance is observed in the amide Ⅱ region and amide Ⅰ region (1600–1500 cm^{-1} and 1700–1600 cm^{-1}) among colorectal cancer patients (Tables 2 and 3). This shift is primarily related to the protein component and is sensitive to changes in protein secondary structure. These vibrations include the stretching motions and the bending motions of NH and bond stretching vibration of C-N [19, 20]. The findings of this study reveal notable alterations in protein structure and relative concentrations within the serum of individuals diagnosed with colorectal cancer. Additionally, a reduction in absorbance is detected within the range of 2990 ~ 2950 cm^{-1} (refer to Table 3), signifying modifications in the antisymmetric stretching vibrations of CH_2 and CH_3 groups, which are indicative of lipid presence [21]. This observation implies a relative reduction in lipid concentrations in the serum of colorectal cancer patients. Moreover, an decrease in absorbance is observed within the range of 1420 to 1380 cm^{-1} (refer to Tables 3) may be related to changes in serum pH or ionization states in cancer patients [22, 23]. In summary, patients with colorectal cancer exhibit marked differences in the concentrations of proteins, lipids, and nucleic acids in their serum, as well as structural modifications in the functional groups of proteins and nucleic acids, compared to healthy individuals.

Fig-1. The serum spectra of both the healthy donor group and the colorectal cancer patient group were examined. The average spectra of the healthy individual group (red) and the colorectal cancer group (blue) were analyzed.

Fig-2. The second derivative of the average serum FTIR spectra of the healthy group (red) and the Colorectal cancer group (blue) were studied.

Table 1
The bands frequencies assigned for the IR spectra

Range (cm^{-1})	**Peak position (cm^{-1})**	**Possible resolve**
990 ~ 1040	1029	C—O—, C—O—H
1040 ~ 1070	1054	C—O stretching
1070 ~ 1090	1083	P = O stretching
1150 ~ 1190	1165	C—O stretching
1230 ~ 1250	1240	P = O stretching
1430 ~ 1470	1455	$-CH_3$ deformation
1500 ~ 1600	1551	Amide
1600 ~ 1700	1655	Amide
1700 ~ 1800	1747	C = O stretching
2910 ~ 2950	2930	C—H stretching
2950 ~ 2990	2970	
3000 ~ 3500	3383	O—H, N—H stretching

Table 2
Spectral peak analysis of colorectal cancer

		Normal(n = 98)	Colorectal (n = 97)	*P-value*
Range (cm^{-1})	1070 ~ 1090	1080.463 ± 4.135	1078.237 ± 6.183	0.906
	1150 ~ 1190	1164.316 ± 8.712	1163.261 ± 8.233	0.149
	1230 ~ 1250	1247.802 ± 6.263	1246.243 ± 3.233	0.268
	1305 ~ 1325	1306.408 ± 2.761	1307.456 ± 7.281	0.393
	1380 ~ 1420	1400.500 ± 1.781	1399.838 ± 1.618	0.059
	1430 ~ 1470	1448.368 ± 3.034	1447.989 ± 3.518	0.131
	1500 ~ 1600	1542.612 ± 1.986	1540.025 ± 2.641	0.003
	1600 ~ 1700	1656.815 ± 3.091	1654.390 ± 1.998	0.044
	2950 ~ 2990	2952.124 ± 6.792	2951.047 ± 9.287	0.051
	3000 ~ 3500	3296.893 ± 5.581	3294.156 ± 4.957	< 0.001

Table 3
Spectral peak analysis of colorectal cancer

		Normal(n = 98)	Colorectal (n = 97)	*P-value*
Range (cm^{-1})	1070 ~ 1090	0.177 ± 0.069	0.168 ± 0.062	0.816
	1150 ~ 1190	0.165 ± 0.088	0.167 ± 0.095	0.919
	1230 ~ 1250	0.221 ± 0.078	0.215 ± 0.061	0.718
	1305 ~ 1325	0.218 ± 0.054	0.221 ± 0.041	0.876
	1380 ~ 1420	0.312 ± 0.075	0.263 ± 0.054	0.048
	1430 ~ 1470	0.268 ± 0.060	0.265 ± 0.065	0.419
	1500 ~ 1600	0.661 ± 0.073	0.641 ± 0.057	< 0.001
	1600 ~ 1700	0.878 ± 0.027	0.883 ± 0.022	0.412
	2910 ~ 2950	0.328 ± 0.073	0.293 ± 0.052	0.022
	3000 ~ 3500	0.814 ± 0.091	0.758 ± 0.087	< 0.001

PCA results

To investigate the differential effects of proteins and nucleic acids in distinguishing individuals with colorectal cancer from healthy individuals, a Principal Component Analysis (PCA) model was developed to analyze the infrared spectra of serum across specific wavelength ranges. These ranges, including 1600–1500 cm^{-1} and 3500–3000 cm^{-1}, were selected due to notable changes in both absorbance and peak position. PCA is a widely used multivariate statistical technique for reducing data dimensionality and examining the underlying structure of a dataset. The analysis facilitates the identification of the principal sources of variation within the dataset and the extraction of the most representative features. The results indicated that the PCA score plot, when visualized in the PC1-PC2-PC3 space, effectively differentiated between the two serum groups. The implementation of PCA modeling along the PC2 and PC3 axes demonstrated a clear level of segregation, categorizing the samples into two distinct groups: the cluster of healthy donors corresponded to the region characterized by negative PC2 and PC3 values, while the cluster of patients corresponded to the region characterized by positive PC2 and PC3 values. Notably, PC1 did not significantly contribute to the K-means clustering analysis (Fig. 3). These findings imply that the primary spectral differences observed between the colorectal cancer group and the healthy group are linked to variations in protein conformation and nucleic acids within serum samples.

Fig-3. The results of the PCA analysis were focused on the spectral regions of 3500 ~ 3000 cm^{-1} and 1600 ~ 1500 cm^{-1}. A scatter plot of the PCA analysis scores was presented in Fig. 3.

Classification Models and ROC RESULTS

Figure 4 shows the Receiver Operating Characteristic (ROC) curves of the Support Vector Machine (SVM), Gaussian Process Regression (GPR), and k-Nearest Neighbors (kNN) models, based on the spectral data from the 1600 ~ 1500 cm^{-1} and 3500 ~ 3000 cm^{-1} wavenumber regions. Notably, the kNN model demonstrated superior diagnostic performance, successfully differentiating 97 cases of colorectal cancer from 98 healthy control samples. The SVM model achieved balanced sensitivity and specificity rates of 88.9% and 92.9%, respectively, with an overall accuracy of 90.9% and an Area Under the Curve (AUC) of 0.911, indicative of excellent diagnostic capability. In contrast, the kNN model reported a sensitivity of 85.2%, specificity of 85.7%, and accuracy of 85.5%, while the GPR model exhibited a sensitivity of 85.2%, specificity of 89.3%, and accuracy of 87.3%, as detailed in Table 4. The SVM classification method displayed a slight inclination towards higher sensitivity relative to specificity, whereas the kNN and GPR models showed a marginal preference for specificity over sensitivity.

Fig-4. Receiver operating characteristic (ROC) analysis of SVM, kNN and GPR model (AUC, area under the ROC curve;) were focused on the spectral regions of 3500 ~ 3000 cm^{-1} and 1600 ~ 1500 cm^{-1}.

Table 4
ROC analysis of kNN, SVM and GPR model

Sensitivity %	**kNN**	**SVM**	**GPR**
	85.2	**88.9**	**85.2**
Specificity %	85.7	92.9	89.3
Accuracy %	85.5	90.9	87.3
ROC (AUC)	0.818	0.911	0.851

Discussion

Colorectal cancer is the third most commonly diagnosed cancer in the world, with approximately 2 million new cases added each year in recent years[24]. Consequently, enhancing the prompt identification of Colorectal cancer stands as a pivotal objective within the realm of clinical medicine. In recent times, the utilization of serum for ATR-FTIR spectroscopy analysis has become a novel, rapid, and convenient cancer detection technology, showing great potential [25–28].

Serum is rich in lipids, proteins, carbohydrates, and various other biomolecules, making it a valuable indicator of physiological and pathological changes in the human body [29]. ATR-FTIR spectroscopy is capable of characterizing all classes of biomolecular species linked to alterations in metabolic reaction products and enzyme activities within serum. This study reveals significant differences in the biochemical composition of serum between patients with colorectal cancer and healthy individuals. The spectral regions of 3500 ~ 3000 cm^{-1} and 1600 ~ 1500 cm^{-1} are particularly critical for the diagnosis of colorectal cancer. A decrease in absorbance in the 3500 ~ 3000 cm^{-1} range indicates changes related to

the stretching and bending of O-H and N-H bonds, primarily suggesting a reduction in the concentration of proteins or carbohydrates. Furthermore, the observation of a blue shift in the amide region (1600–1500 cm^{-1}) in patients with colorectal cancer indicates alterations in the structural configuration of protein functional groups. Specifically, within the 1600 – 1500 cm^{-1} range, the findings of Guo et al. [12] also demonstrated that this spectral region can effectively distinguish colorectal cancer from liver cancer and healthy individual serum.

Furthermore, Principal Component Analysis (PCA) effectively distinguished serum samples from healthy individuals and those with colorectal cancer. A range of classification techniques, alongside three distinct machine learning algorithms, were utilized to differentiate between 100 cancer cases and 100 healthy control samples. These approaches yielded promising statistical metrics, highlighting the significant potential of this technology. Notably, the SVM model proved to be the most effective, achieving a sensitivity of 88.9%, a specificity of 92.9%, and an accuracy of 90.9% (refer to Table 4). Additionally, Receiver Operating Characteristic (ROC) analysis produced an Area Under the Curve (AUC) value of 0.911 (Fig. 4), further corroborating the model's excellent diagnostic capability. Despite the promising outcomes of this proof-of-concept study, several research questions remain unresolved. Primarily, the limited sample size necessitates further investigations with larger cohorts to substantiate our findings. Future research should also aim to investigate the impact of spectral changes on the pathological staging of colorectal cancer and the influence of pathogenic factors, such as Helicobacter pylori. Furthermore, it is imperative to focus efforts on the development of more precise and dependable diagnostic models utilizing ATR-FTIR spectroscopy. This study presents an innovative approach for the detection of colorectal cancer by identifying biomarkers in serum samples through the application of ATR-FTIR spectroscopy. Meanwhile, conducting extensive prospective clinical trials is necessary. It should also be noted that the findings are limited by the sample size, and future studies with larger cohorts are needed to improve specificity. Deeper bioinformatics analyses could further refine the diagnostic signatures and warrant further exploration.

Conclusion

The principal aim of clinical fundamental research related to ATR-FTIR spectroscopy is to devise novel and improved techniques for diagnosing diseases that are currently difficult or expensive to detect using conventional methods. The results of this study illustrate that the integration of ATR-FTIR spectroscopy with principal component analysis facilitates the differentiation of biomarkers, including proteins, lipids, and nucleic acids, between healthy and colorectal cancer samples. The data reveal that the spectra and spectral parameters of serum samples from individuals with colorectal cancer exhibit significant differences compared to those from healthy individuals. Notably, the spectral ranges of 3500 ~ 3000 cm^{-1} and 1600 ~ 1500 cm^{-1}, were identified as the most critical criteria for distinguishing colorectal cancer serum samples from healthy ones. These differences can be attributed to alterations in nucleic acids and significant changes in protein conformation within the serum samples during carcinogenesis. The diagnostic approach utilizing infrared spectroscopy, as implemented by the SVM model, achieves a

sensitivity of 88.9% and a specificity of 92.9%, thereby providing a theoretical foundation for future comprehensive research and practical applications.

Declarations

Ethicals declarations

This study was conducted with the authorization of Chongming Hospital, affiliated with Shanghai University of Medicine and Health Sciences, following approval by Chongming Hospital, affiliated with Shanghai University of Medicine and Health Sciences ethics committee (Approval number: xxx). In accordance with the Declaration of Helsinki, written informed consent was obtained from each participant.

Clinical trial number: not applicable.

Funding

This work was supported by Chongming District Science and Technology Committee (CKY202509) Shanghai University of Medicine and Health Sciences, Chongming Branch Scientific Program (CMYY2024-02), (CMYY2024-04). Scientific Research Foundation of SUMHS (SSF-23-24-008). Education and Industry Collaboration in Joint Training Programs (No.231107070285529). The 2025 Rising Star Project of Chongming Dis-trict Science and Technology Commission in Shanghai (CKY2025-65).

Author Contribution

Degao Zhang, Jun Leng and Chao Yang conceived the concept, designed and supervised the experiments. Wanli Yang, Nan Pang and Jia Shi performed most of the experiments, data analysis and drafted the article. Wei Zhang, Jin Han and Ao Song participated in part of the cytological experiments and data analysis.

Acknowledgement

This work was supported by the National Synchrotron Radiation Laboratory in Hefei.

References

1. Siegel RL et al (2023) Colorectal cancer statistics, 2023. CA Cancer J Clin 73(3):233–254
2. Housini M et al (2024) Colorectal cancer: Genetic alterations, novel biomarkers, current therapeutic strategies and clinical trials. Gene 892:147857

3. Biller LH, Schrag D (2021) Diagnosis and Treatment of Metastatic Colorectal Cancer: A Review. JAMA 325(7):669–685
4. Mitsala A et al (2021) Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era. Curr Oncol 28(3):1581–1607
5. Khanmohammadi M et al (2009) Diagnosis of colon cancer by attenuated total reflectance-Fourier transform infrared microspectroscopy and soft independent modeling of class analogy. Med Oncol 26(3):292–297
6. Gupta S (2022) Screening for Colorectal Cancer. Hematol Oncol Clin North Am 36(3):393–414
7. Sha J et al (2020) Computed tomography colonography versus colonoscopy for detection of colorectal cancer: a diagnostic performance study. BMC Med Imaging 20(1):51
8. Yang X et al (2021) Diagnosis of Lung Cancer by ATR-FTIR Spectroscopy and Chemometrics. Front Oncol 11:753791
9. Tiernan H, Byrne B, Kazarian SG (2020) ATR-FTIR spectroscopy and spectroscopic imaging for the analysis of biopharmaceuticals. Spectrochim Acta Mol Biomol Spectrosc 241:118636
10. Tomas RC et al (2022) Detection of breast cancer by ATR-FTIR spectroscopy using artificial neural networks. PLoS ONE 17(1):e0262489
11. Sitnikova VE et al (2020) Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis. Talanta 214:120857
12. Guo S et al (2022) Fast and Deep Diagnosis Using Blood-Based ATR-FTIR Spectroscopy for Digestive Tract Cancers. Biomolecules, 12(12)
13. Song CL et al (2019) Fourier transform infrared spectroscopic imaging of colon tissues: evaluating the significance of amide I and C-H stretching bands in diagnostic applications with machine learning. Anal Bioanal Chem 411(26):6969–6981
14. Akin Geyik G et al (2022) A rapid diagnostic approach for gastric and colon cancers via Fourier transform mid-infrared spectroscopy coupled with chemometrics from paraffin-embedded tissues. Spectrochim Acta Mol Biomol Spectrosc 267(Pt 2):120619
15. Li Q et al (2017) Colorectal Cancer and Colitis Diagnosis Using Fourier Transform Infrared Spectroscopy and an Improved K-Nearest-Neighbour Classifier. Sens (Basel), 17(12)
16. Guleken Z et al (2021) Assessment of structural protein expression by FTIR and biochemical assays as biomarkers of metabolites response in gastric and colon cancer. Talanta 231:122353
17. Jolliffe IT, Cadima J (2016) *Principal component analysis: a review and recent developments.* Philos Trans A Math Phys Eng Sci, 374(2065): p. 20150202
18. Elmi F et al (2017) Application of FT-IR spectroscopy on breast cancer serum analysis. Spectrochim Acta Mol Biomol Spectrosc 187:87–91
19. Yang H et al (2015) Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. Nat Protoc 10(3):382–396

20. Ganim Z et al (2008) Amide I two-dimensional infrared spectroscopy of proteins. Acc Chem Res 41(3):432–441
21. Sutton CC, Franks GV, da Silva G (2015) Modeling the antisymmetric and symmetric stretching vibrational modes of aqueous carboxylate anions. Spectrochim Acta Mol Biomol Spectrosc 134:535–542
22. Wang R, Wang Y (2021) Fourier Transform Infrared Spectroscopy in Oral Cancer Diagnosis. Int J Mol Sci, 22(3)
23. Wang X et al (2014) FTIR spectroscopic comparison of serum from lung cancer patients and healthy persons. Spectrochim Acta Mol Biomol Spectrosc 122:193–197
24. Eng C et al (2024) Colorectal cancer. Lancet 404(10449):294–310
25. Chen H et al (2022) Rapid and sensitive detection of esophageal cancer by FTIR spectroscopy of serum and plasma. Photodiagnosis Photodyn Ther 40:103177
26. Yang B et al (2021) Detection of breast cancer of various clinical stages based on serum FT-IR spectroscopy combined with multiple algorithms. Photodiagnosis Photodyn Ther 33:102199
27. Su KY, Lee WL (2020) Fourier Transform Infrared Spectroscopy as a Cancer Screening and Diagnostic Tool: A Review and Prospects. Cancers (Basel), 12(1)
28. Kepesidis KV et al (2021) Breast-cancer detection using blood-based infrared molecular fingerprints. BMC Cancer 21(1):1287
29. Psychogios N et al (2011) The human serum metabolome. PLoS ONE 6(2):e16957

Figures

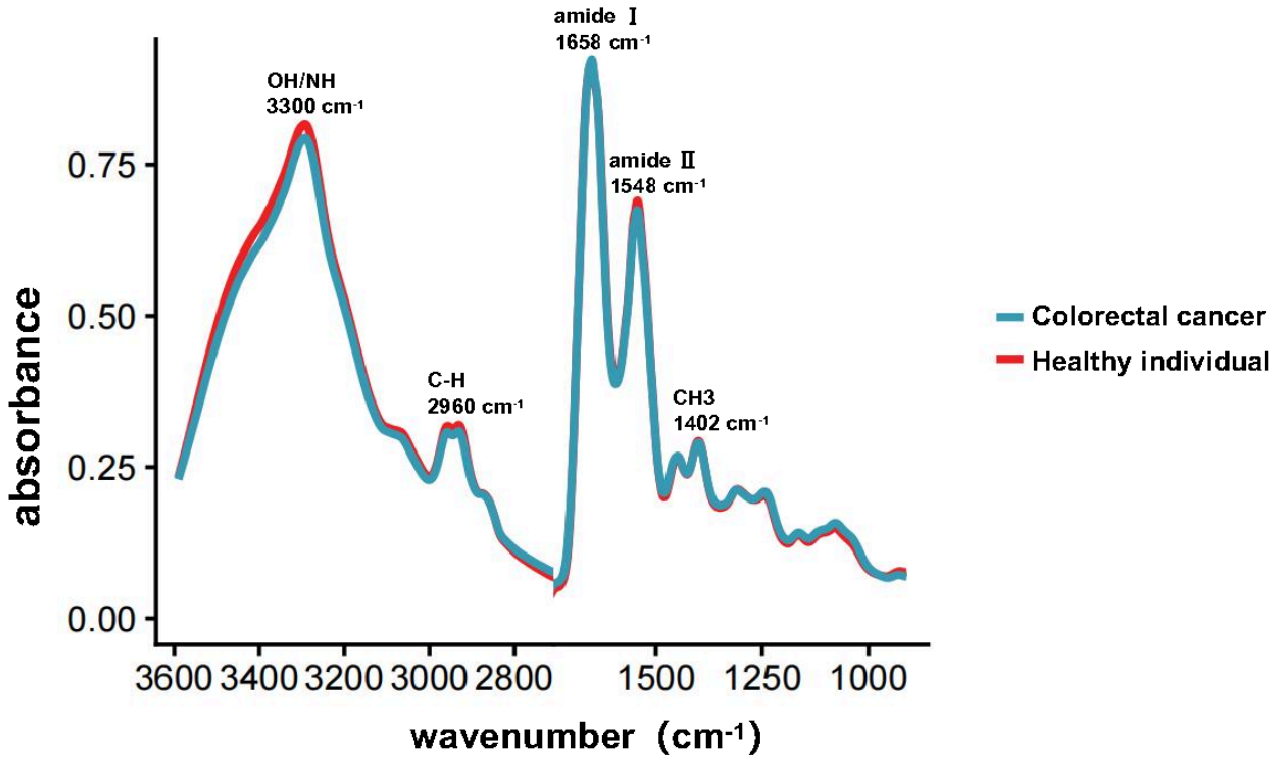


Figure 1

The serum spectra of both the healthy donor group and the colorectal cancer patient group were examined. The average spectra of the healthy individual group (red) and the colorectal cancer group (blue) were analyzed.

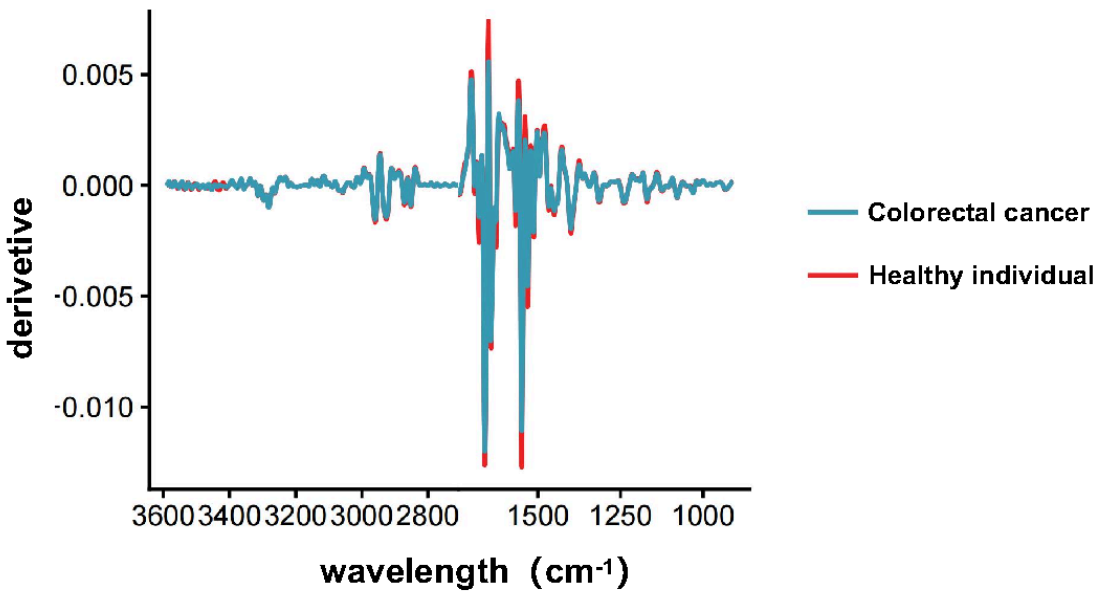


Figure 2

The second derivative of the average serum FTIR spectra of the healthy group (red) and the Colorectal cancer group (blue) were studied.

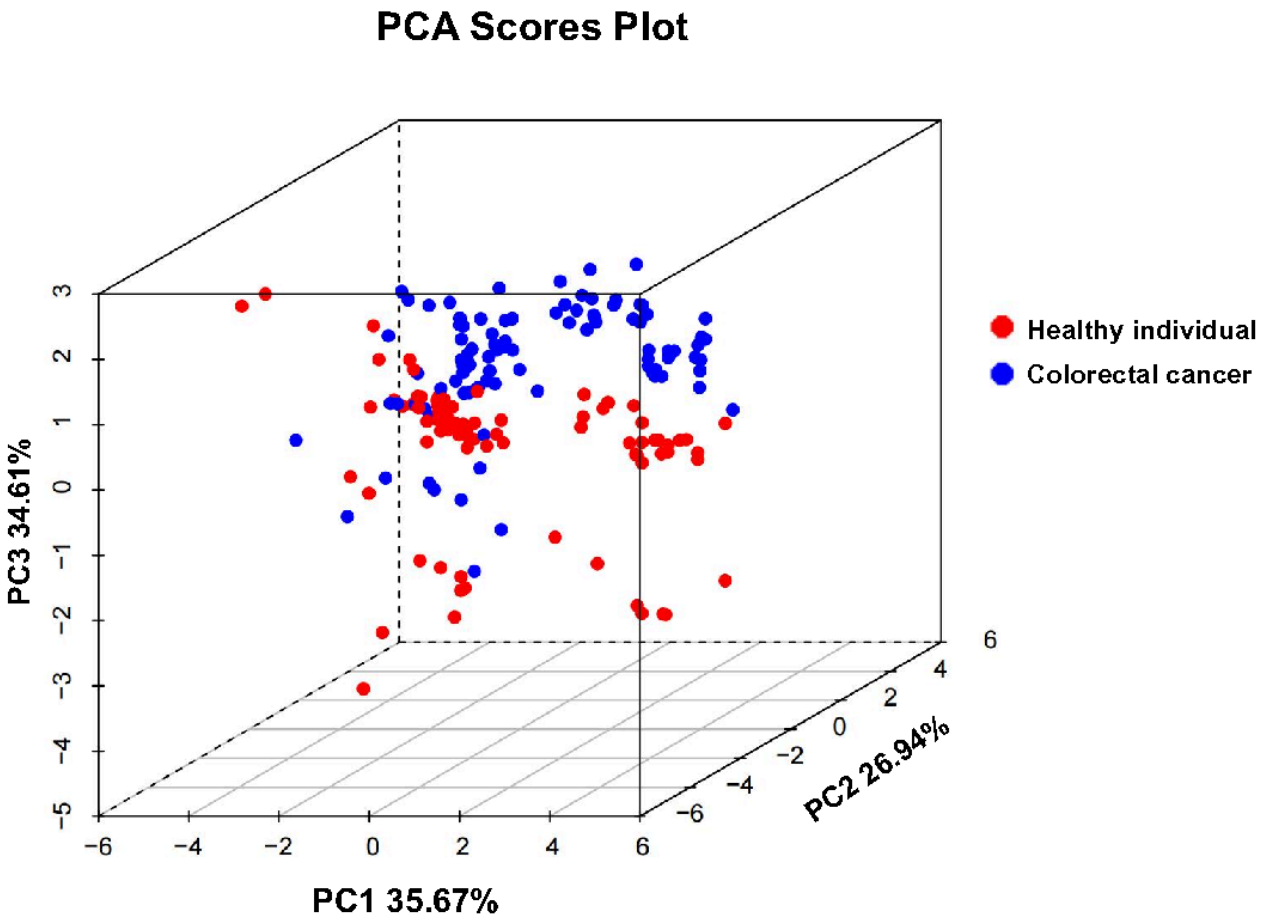


Figure 3

The results of the PCA analysis were focused on the spectral regions of 3500 ~ 3000 cm^{-1} and 1600 ~ 1500 cm^{-1}. A scatter plot of the PCA analysis scores was presented in Figure 3.

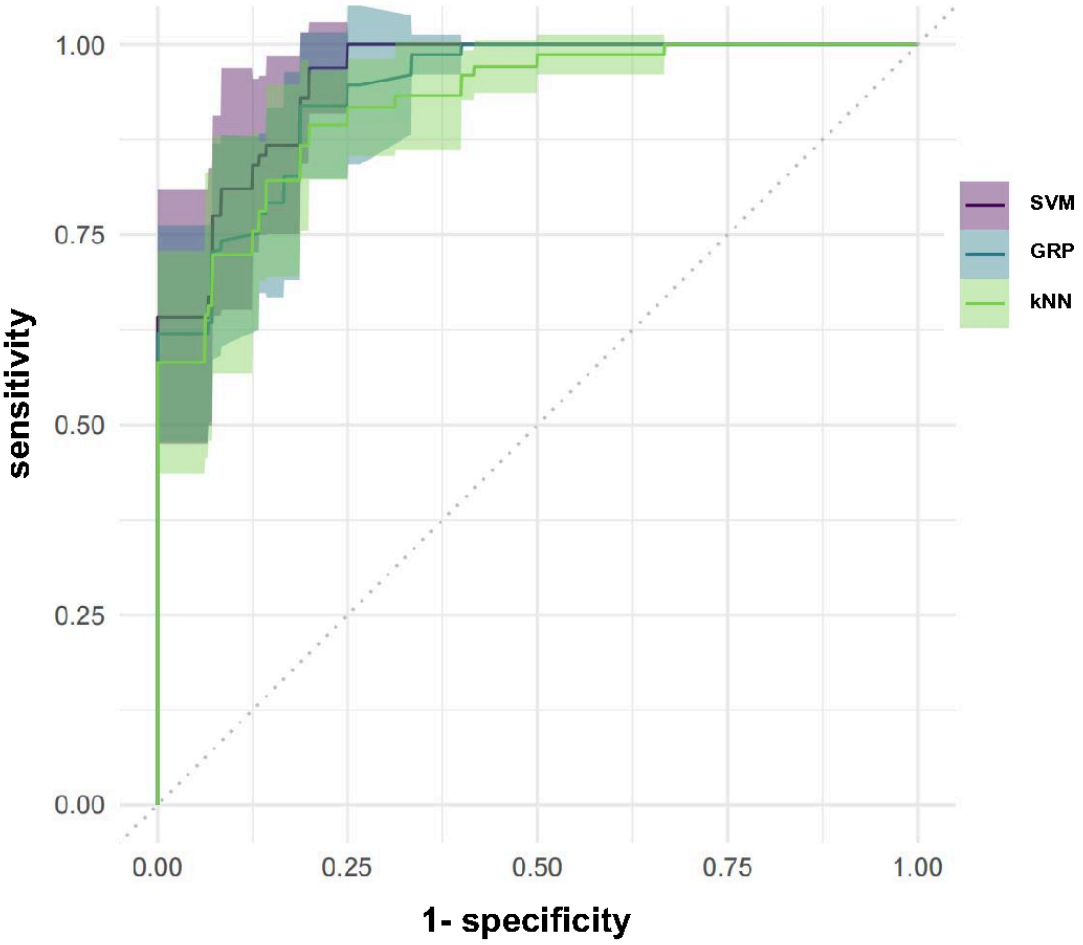


Figure 4

Receiver operating characteristic (ROC) analysis of SVM, kNN and GPR model (AUC, area under the ROC curve;) were focused on the spectral regions of 3500 ~ 3000 cm^{-1} and 1600 ~ 1500 cm^{-1}.