ELSEVIER

Nanomedicine: Nanotechnology, Biology, and Medicine
48 (2023) 102657

nanomedicine
Nanotechnology, Biology, and Medicine

nanomedjournal.com

Correlation between human colon cancer specific antigens and Raman spectra. Attempting to use Raman spectroscopy in the determination of tumor markers for colon cancer

Joanna Depciuch, PhD[a,1,*], Paweł Jakubczyk, Prof.[b], Wiesław Paja, PhD[c], Krzysztof Pancerz, PhD[d], Agnieszka Wosiak, Prof.[e], Monika Kula-Maximenko, PhD[f], İlhan Yaylım, Prof.[g], Güldal İnal Gültekin, Prof.[h], Nevzat Tarhan, Prof.[i], Mehmet Tolgahan Hakan, PhD[g], Dilara Sönmez, M.Sc[g], Devrim Sarıbal, Prof.[j], Soykan Arıkan, Prof.[k,l], Zozan Guleken, Prof.[m,1,*]

[a]*Institute of Nuclear Physics Polish Academy of Science, 31-342 Krakow, Poland*
[b]*Institute of Physics, University of Rzeszów, Poland*
[c]*Institute of Computer Science, University of Rzeszow, Poland*
[d]*Institute of Philosophy, John Paul II Catholic University of Lublin, Poland*
[e]*Institute of Information Technology, Lodz University of Technology, Poland*
[f]*The Franciszek Górski Institute of Plant Physiology, Polish Academy of Sciences, ul. Niezapominajek 21, 30-239 Kraków, Poland*
[g]*Istanbul University, Aziz Sancar Institute of Molecular Medicine, Istanbul, Turkey*
[h]*Department of Physiology, Okan University Faculty of Medicine, Istanbul, Turkey*
[i]*Uskudar University NP Hospital, Istanbul, Turkey*
[j]*Department of Biophysics, Cerrahpaşa Medical School, Istanbul, Turkey*
[k]*Istanbul Education and Research Hospital, Department of General Surgery, Istanbul, Turkey*
[l]*Cam and Sakura City Hospital, Istanbul, Turkey*
[m]*Uskudar University, Faculty of Medicine, Department of Physiology, Istanbul, Turkey*

Revised 6 December 2022

Abstract

Colorectal cancer is the second most common cause of cancer-related deaths worldwide. To follow up on the progression of the disease, tumor markers are commonly used. Here, we report serum analysis based on Raman spectroscopy to provide a rapid cancer diagnosis with tumor markers and two new cell adhesion molecules measured using the ELİSA method. Raman spectra showed higher Raman intensities at 1447 cm^{-1} 1560 cm^{-1}, 1665 $cm^{-1,}$ and 1769 cm^{-1}, which originated from CH_2 proteins and lipids, amide II and amide I, and C═O lipids vibrations. Furthermore, the correlation test showed, that only the CEA colon cancer marker correlated with the Raman spectra. Importantly, machine learning methods showed, that the accuracy of the Raman method in the detection of colon cancer was around 95 %. Obtained results suggest, that Raman shifts at 1302 cm^{-1} and 1306 cm^{-1} can be used as spectroscopy markers of colon cancer.

Keywords: Colon cancer; ICAM-1; CEACAM-1; Raman spectroscopy; Machine learning

Introduction

The disease of colon cancer (CRC) affects many people around the world. The disease was ranked second in terms of death rates and third in terms of new cases, according to Globocan 2018.[1] Also, more recent studies had shown that there are a significant number of rectal cancer deaths misclassified as

* Corresponding authors.
E-mail addresses: joanna.depiuch@ifj.edu.pl (J. Depciuch), zozanguleken@gmail.com, zozan.guleken@uskudar.edu.tr (Z. Guleken).
[1] Equal senior authors.

https://doi.org/10.1016/j.nano.2023.102657
1549-9634/

colon cancer.[2] Therefore, the number of people suffering from colon cancer is likely to be higher. Cancer biomarkers are often detectable in blood samples. This suggests that the incidence of the disease is higher than the known. Diagnostic tests such as blood tests are minimally invasive and simple.[3,4] ELISA generally exhibited high specificity and sensitivity. The detection effect is generally evaluated according to the area under the a percentile method and the receiver-operating curve ROC curve.[5]

Detection and identification of biomarkers are critical for the management of the disease and progress. Biomarkers could be found that alone or in combination with existing tests or tools could identify the predisposition or early stage of disease, thus enhancing risk stratification for screening. By selecting the appropriate chemotherapeutic drugs through biomarkers, clinicians will also be able to change diagnostic and treatment algorithms across a wide range of patients. Carcinoembryonic antigen (CEA), Cancer antigen (CA), and Alfa fetoprotein (AFP) are mainly detectable in the serum well-known soluble tumor markers which are used as a diagnostic and prognostic index of the disease. A colorectal cancer tumor marker, carcinoembryonic antigen (CEA),[6] has been described for the first time in 1965, and a preoperative test for CEA is recommended by the National Comprehensive Cancer Network the American Society of Clinical Oncology,[7,8] and the European Group on Tumor Markers for patients with nonmetastatic colorectal cancer. The American Joint Committee on Cancer has proposed including CEA in the staging system in order to correlate it with metastases and recurrence[7,9] Another tumor marker cancer antigen 125 (CA 125) is also specific to peritoneal dissemination in gastrointestinal cancers.[10] Additionally, current studies provided that the CA125 biomarkers can be used to not only diagnose CRC, but also to predict the spread of lymph nodes, vascular invasion, neural invasion, tumor differentiation, and staging, using them as tools for guiding treatments and prognoses[11] CA-15-3 is a heterogeneous glycoprotein antigen, with two monoclonal antibodies (115D8 and DF3). The concentration of markers is also elevated in metastatic cancer such as ovarian, colorectal, lung, stomach, and uterus cancers.[12] Nevertheless, CA15.3 correlates better with pretherapy breast cancer patients.[13] Carbohydrate antigen 19-9 (CA19-9) is a monoclonal antibody, that binds to E-Selectin. Both malignant and benign processes can cause an increase in serum CA19-9. Cancers of the pancreas, stomach, lungs, biliary tract and colon produce CA19-9 tumor markers.[14] The increase in CA19-9 level is also observed in patients with liver cirrhosis, acute cholangitis, diabetes mellitus, endometriosis, and bronchiectasis.[15] Alpha-fetoprotein (AFP) initially exists in human fetal sera.[16] Tumors that produce AFP are usually found in endodermal organs derived from the foregut (ref). Cancers that produce AFP are mostly found in the stomach, bile ducts, and pancreas, and they have a high rate of liver metastasis and a poor prognosis.[17] There is potential value in measuring serum AFP levels to predict the recurrence of colon cancer. Cancers that produce AFP display more aggressive biological behavior and have a worse prognosis than cancers that do not produce AFP.[18]

Often malignant tumors are characterized by both disruptions of the tissue structure and dysfunctional differentiation. The abnormal expression of various adhesion molecules induces loss of cell-cell adhesion, contributing to tumor invasion and metastasis such that this causes tumor differentiation and malignant invasion.[19]

In this study we used classical tumor markers which are used in clinics. Additionally, we researched two new markers that may be a tumor marker First is carcinoembryonic antigen-related cell adhesion molecules (CEACAM) which is surprisingly diverse functions in cell adhesion, in intracellular and intercellular signaling during complex biological processes such as cancer progression, inflammation, angiogenesis, and metastasis. The other is intercellular adhesion molecule (ICAM-1) that the elevated serum levels recently been described in patients with solid tumors, including colorectal cancers. These serum levels correlate in many studies with tumor stage and the development of metastases, although their impact on patient survival is at present unclear and more studies are required to use these two adhesion molecule as a colon cancer tumor marker.[20] The presence of ICAM-1 has been documented in a wide range of cell types within the tumor microenvironment (TME), including tumor and stromal cells, blood, and lymph vessels, as well as infiltrating and resident immune cells as a potential angiogenic and inflammatory promotor. Metastasis also resulted in increased ICAM-1 expression, in the same manner as the primary tumor.[21]

CEACAM-1 The cancer-associated cell adhesion molecules (CEACAMs) are glycosylphosphatidylinositol (GPI)-linked immunoglobulins.[22] CEACAM-1 is a subtype of CEACAM also known as biliary glycoprotein I or CD66a.[23] Studies revealed that CEACAM-1 has a role in the inhibition of carcinogenesis.[24] ,In addition to promoting apoptosis of various cells such as pulmonary and mammary epithelial cells, oral keratinocytes, cancer cells, Jurkat T cells, and cardiomyocytes.[25,26] Recent studies have revealed that CEACAM-1 activity is critical in the progression of colorectal cancer, and therefore, CEACAM-1 may be a promising therapeutic target for the treatment of colon cancer.[27]

However, still new methods, that may help in diagnosis are sought. A promising technique is Raman spectroscopy. In this technique a lot of methods such as Raman scattering (RS) microscopy, stimulated Raman scattering, coherent Stokes and anti-Stokes Raman scattering were used to detect chemical changes in diseases blood fluids, tissues, cells.[28] Importantly, these methods are characterized by selectivity, sensitivity with small cost and high speed.[29] Furthermore, Raman spectroscopy is non-destructive methods.[30] A plenty of studies were used Raman spectroscopy for the diagnosis of cholangiocarcinoma,[31] gastric,[32] pancreatic,[33] lung,[34] breast,[35] endometrial,[36,37] and bladder[38] cancers. Several studies reported diagnostic sensitivity and specify of Raman spectroscopy in the diagnosis of colon cancer[39,40] or developed a highly accurate prediction model using machine learning models.[41]

Hence, we considered that serum analysis based on Raman spectroscopy whether could provide a rapid cancer diagnosis with tumor markers and two new cell adhesion molecules. Additionally, we constructed a model that predicts cancer with increased accuracy based on machine learning. The obtained results were compared with clinical levels of tumor markers in colon cancer patients in order to explore the clinical possibility of using this method for tumor marker detection. Additionally,

we determined whether Raman shifts correlate with tumor biomarkers.

Materials and matters

Clinical features of the patients

Istanbul University İstanbul-Cerrahpaşa Medical School's Ethics Committee approved the study protocols with permission number E-83045809-604.01.02-28,898 and date in date 25.01.2021. The information and biological samples of cancer patients were provided by an expert oncologist at Istanbul Research and Training Hospital who specializes in the histopathological diagnosis of gastric cancer and colon cancer. The study was conducted on 30 CRC-diagnosed patients, who were applied to the hospital in the last year and 45 healthy controls who were applied to the hospital but not diagnosed with any type of cancer. Properly, 75 blood serum samples were gathered for the study. The patients and healthy controls participating in our study are the people who applied to the State Hospital where colon cancer cases are concentrated in Turkey. At the beginning of the study, only age, gender, stage of the cancer depending on TNM system of these patients were taken into consideration, but their ethnic origins and blood groups were not taken into consideration. The patient participants were 36.6 % (n = 11) women, 63.3 %(n = 19) male the group. The average age for women were 61 ± 12.8; 56.8 ± 9.18 for men. The specific definition of cancer stage is generally described by The TNM system. The letters and numbers describe the tumor (T), lymph nodes (N), whether or not the cancer has spread or metastases (M). In this study, 3 of patients (10 %) were in T2 stage, 9 of them (30 %) were in T3 stage, 17 of them (56.6 %) were in T4 stage and finally 1 patient (3.3 %) were inoperate. 12 patients (34.2 %) were in N0, 7 patients (20 %) were N1, 3 patients were (8.57 %) N2 and 3 patients (20 %) and 1 patient(2.8 %) were inoperate. Additionally, 18 patients did not have metastasis (60 %), and 12 patients had (40 %) unfortunately had metastasis.

The Personal Data Act and the Act on Ethical Review for human experiments applied to data collected from patients who were involved in this experiment. The study participants have all signed consent forms after they received full information and were fully informed. Tumor markers and adhesion molecule levels were determined by the enzyme-linked immunosorbent assay (ELISA) method according to procedures recommended by the manufacturer (ref). The results were presented in pg/mL for endocan and ng/mL for sICAM-1 (according to.[42,43]

Samples preparation for the acquisition of Raman spectra

Collected blood serum was measured using FT-Raman spectrometer Nicolet NXR 9650, which is equipped with an Nd: YAG laser with 1064 nm wavelength. Before measurement, the follicular fluid samples were thawed. To obtain the spectra, a volume of 4 μL of each serum sample was dropped and waited about 8 min to dry. All samples were measured in the range from 150 to 3700 cm^{-1} with 8 cm^{-1} spectral resolution and 64 scans. The laser power was 1 W. Raman spectra were processed by the Omnic/Thermo Scientific software. For each spectrum smoothing using 9 points, vector normalization and baseline correction were done.[44,45]

Statistics - multivariate analysis and correlation test

In this study, Principal Component Analysis (PCA) was used to show distinguish between samples collected from control and people suffering from colon cancer. For this purpose, two Raman ranges were analyzed (800 cm^{-1}–1800 cm^{-1} and 2700 cm^{-1}–3000 cm^{-1}) was selected. Moreover, a correlation between tumor markers and selected Raman shifts, Pearson correlation test was performed. Both analyses were done using Past 3.0 software.

Analysis of intensity dynamics and spectroscopy markers of colon cancer

The analysis of Raman intensity dynamics[46,47] is a very useful approach in spectra comparison of healthy persons and patients with selected disease entities. It allows finding the Raman shifts regions which indicate different intensity dynamics between healthy and sick persons and helps to find the differences in the Raman spectrum under consideration. These differences come from the altered dynamics caused by the disease. If we assume that the average spectra of samples from the control group can be considered as reference spectra, then we can find differences in intensity dynamics for every colon cancer patient compared to these reference spectra. Such an approach allows finding regions of Raman shifts, that simultaneously marked different dynamics for all colon cancer patients. Therefore, these regions can be considered as potential markers to discriminate colon cancer patients only by their IR spectrum. Moreover, In order to quantitatively describe the effectiveness of this approach, we introduce a discrimination coefficient D_{Cancer} for colon cancer patients, which determines the detection efficiency of colon cancer patients and $D_{Control}$ for the healthy individuals which determines the detection efficiency of healthy individuals. The discrimination coefficients can be calculated from the Eq. (1)

$$D_{Cancer} = \frac{N_{Cancer\_altDyn}}{N_{Cancer}} \cdot 100\%, \quad D_{Control} = \frac{N_{Control\_unchDyn}}{N_{Control}} \cdot 100\% \quad (1)$$

where N_{Cancer_altDyn} denotes the number of colon cancer patients with altered dynamics detected by the intensity dynamics approach, N_{Cancer} is the total number of colon cancer patients, $N_{Control\_unchDyn}$ denotes the number of healthy individuals from the control group with unchanged dynamics detected by the intensity dynamics approach and $N_{Control}$ is the total number of the control group.

In other words the discrimination probability D_{Cancer}, gives the information on how many patients have altered intensity dynamics compared to the reference spectrum, whereas $D_{Control}$ how many individuals from the control group have unchanged dynamics. The calculations were performed in a Matlab Simulink environment (MathWorks, Natick, MA).

Data analysis using machine learning methods

The spectroscopic data were also analyzed using selected machine learning methods. The goal was to determine the

efficiency of identifying sick and healthy patients. Three different ML methods were used: Deep Learning (DL),[48] Support Vector Machine (SVM),[49,50] and eXtreme Gradient Boosting trees (XGBoost).[51] Selected quality measures were used to evaluate the quality of the built models: accuracy, sensitivity, specificity, precision, F1 score, and Matthew correlation coefficient (MCC). In addition, the analysis of the quality of the models is presented using ROC curves and the values of the area under them. Such curves are a visualization of the relationship between the effectiveness of positive classifications (Sensitivity), True Positive Rate, and the failure to classify negative cases (1-Specificity), False Positive Rate, at each probability level. Thus, the ROC curve illustrates how large the percentage of misclassifications (positive and negative) will be for a given cut-off point. How does the assumed limiting the probability of assignment to a positive category (disease occurrence) simultaneously affect the correct and incorrect classification of these cases. In the case of classification models, i.e., models that predict the values of a qualitative variable, one of the primary ways to summarize is to assess the quality of the classification - of whether the model correctly assigns cases to categories of the dependent variable. The smaller the discrepancy between the actual values and those predicted by the statistical model, the better. Data from spectroscopic experiments have been converted to information systems, tables consisting of columns and rows. The columns represent the individual Raman shifts (260 for the 1800 cm^{-1} - 800 cm^{-1} range and 78 for the 3000 cm^{-1} - 2700 cm^{-1} range) and the rows the intensity values of these Raman shift for specific patients, the last column is the patient category: sick (positive, P), healthy (negative, N). The table contains results obtained for 74 patients, and learning cases, including 30 sick cases and 44 healthy cases. Based on the prepared datasets, ML algorithms-built learning models. These models were tested on test cases using leave-one-out cross-validation (LOOCV) approach. Calculations were performed in the R environment using selected computational packages: class, caret, keras, xgboost, and e1071. In addition, we used the Boruta feature selection method to identify all relevant Raman shifts from the original sets of features[52,53] and to create the subsets of relevant features.

That is, 6 sets were created for ML analysis:

1. set of cases described by 260 features, Raman shifts in the range 1800 cm^{-1} - 800 cm^{-1},
2. set of cases described by 68 features, Raman shifts in the range 1800 cm^{-1} - 800 cm^{-1}, obtained by all relevant feature selection algorithms,
3. set of cases described by 78 features, Raman shifts in the range 3000 cm^{-1} - 2700 cm^{-1},
4. a set of cases described by 21 features, Raman shifts in the 3000 cm^{-1} - 2700 cm^{-1}range, obtained by all relevant feature selection algorithms,
5. set of cases described by 338 features, Raman shifts from 1800 cm^{-1} - 800 cm^{-1} and 3000 cm^{-1} - 2700 cm^{-1} ranges,
6. set of cases described by 43 features, Raman shifts from 1800 cm^{-1} - 800 cm^{-1} and 3000 cm^{-1} - 2700 cm^{-1} ranges, obtained by all relevant feature selection algorithms.

Results

Biochemical results

The basic and demographic characteristics of the patients are presented in Table 1. We assessed histopathological distribution; tumor size (T stage), lymph node status (N stage), distant metastasis, lymphovascular invasion, and perineal invasion. As a reference range, CEA maintains a blood level of 0–2.5 nanograms per milliliter (ng/mL). In general, levels at 10 ng/mL, the extensive disease is suggested, and levels over 20 ng/mL suggest spreading cancer or maybe a recurrence indicator also. In our study, the CEA level was higher than the reference level (16.5 ± 24.1). Ca125 results are expressed in units per milliliter (U/mL). CA 125 has a reference range of 0–35 units/mL. The CA-125 level in colon cancer was 47.2 ± 81.7 which was higher than the reference level. Another tumor marker CA-15-3 normal test should be less than or equal to 30 U/mL which is in the range of reference level with the value of 14.4 ± 9.54. Additionally, CA-19-9 normal results are <37 U/mL. In this study, it was 50.3 ± 83.1 which was higher than the reference level. Alfa fetoprotein is another tumor marker that we investigated. AFP normal range 0.50–5.50 IU/mL, but AFP level between 10 ng/mL to 20 ng/mL is normal for adults. It was a 2.36 ± 1.88 level in colon cancer patients which was low but in the reference range. In pathologic conditions as well as in their resolution, ICAM-1(ng/mL) appears to regulate many essential cellular functions. In our study, the level of ICAM-1 was 335 ± 269 with a minimum of 122 ng/mL and maximum 1065 ng/mL. Malignant phenotypes are more likely to be caused by invasion of cancer cells than tumor growth. It has been reported that CEACAM1 is associated with poor survival rates in patients with colorectal cancer and hepatocellular carcinoma. Enhanced migration, invasion, metastasis, and poor prognosis are all related to CEACAM1 cytoplasmic domain balance and long cytoplasmic domain dominance.[54] Our CEACAM1 results were 14.7 ± 2.73 ng/mL which were down

Table 1
Descriptive of tumor markers.

	CEA (ng/mL)	CA-125 (U/mL)	CA-15-3 (U/mL)	CA-19-9 (U/mL)	AFP (ng/mL)	Soluble ICAM (ng/mL)	CEACAM-1 (ng/mL)
Mean	16.5	47.2	14.4	50.3	2.36	335	14.7
Median	6.59	17.9	12.3	22.6	2.22	217	13.9
Standard deviation	24.1	81.7	9.54	83.1	1.88	269	2.73
Minimum	0.0100	4.80	5.00	0.700	0.300	122	11.3
Maximum	98.5	291	45.0	378	8.50	1065	20.9

below the reference range (381.30–968.13 ng/mL) as shown in Table 1.

Also, we compared tumor markers with One-way Anova-Kruskal-Wallis comparison. ICAM-1 level was significant ($p < 0.05$) in the condition of the presence of distance metastasis and in the condition of mucinous components as presented in Table 3.

Raman measurement results

In this study, Raman spectroscopy was used to determine chemical differences between serum collected from healthy patients and people suffering from colon cancer, as well as to show the possibility of using this technique in differentiation between healthy and unhealthy groups of patients. Therefore, firstly, the chemical composition of serum was presented in Fig. 1. In the Raman spectra peaks corresponding to symmetric dioxy stretching of the phosphate backbone from nucleic acids and phospholipids were noticed around 810 cm^{-1}, 900 cm^{-1}, 1000 cm^{-1}, and 1080 cm^{-1}. Furthermore, asymmetric vibrations from polysaccharides was visible at 1182 cm^{-1}.[55] Raman shifts around 1280 cm^{-1}, 1370 cm^{-1} and 1460 cm^{-1} were originated from CH_2 proteins and lipids.[56] Amide II and amide I vibrations were observed at 1560 cm^{-1} and 1660 cm^{-1}, respectively.[55] The highest Raman shifts values were corresponded from C═O, CH_3 and (C═)C-H lipids vibrations (Raman shifts at 1769 cm^{-1}, around 2930 cm^{-1} and 3060 cm^{-1}).[55] Detailed marked and description of the peaks in the Control and Colon Cancer Raman spectra were presented in Fig. 1A.

In serum collected from patients suffering from colon cancer, higher Raman intensities of peaks corresponding to CH_2 lipids and proteins vibrations, as well as amides II, I and C═O vibrations from lipids, were visible in comparison with control group. The value of Raman intensity for other lipids vibrations and functional groups building nucleic acid and phospholipids vibrations was very similar for both analyzed group of patient, Fig. 1B. However, from the Raman spectra, it is not possible to show, if this spectroscopy method can be used to differentiate serum collected from healthy and un-healthy patients and which Raman shifts could be used as a spectroscopy marker of colon cancer. Therefore, PCA analysis were done, Fig. 2.

Multivariate, statistic and machine learning analysis of Raman spectra

From the PCA analysis performed for Raman region between 800 cm^{-1} and 1800 cm^{-1} it was visible, that components with the highest value of variance (PC1–74.58 %, PC2–10.07 %) were not show differentiation between healthy and un-healthy patients, Fig. 2 A1. However, using PC4 (2.71 %) and PC5 (1.61 %) components, it was visible that almost all samples collected from un-healthy patients had positive values of both components, while samples collected from control group showed negative value of PC4 and positive as well as negative value of PC5, Fig. 2 B1. Furthermore, it was visible, that samples collected from un-healthy patients were grouped in other quadrant of the coordinate system as a samples collected from healthy one, which mean, that these two groups of samples can be distinguish to each other. Loading plots of PC1, PC2, PC4 and PC5 for Raman region between 800 cm^{-1} and 1800 cm^{-1} were presented in Fig. 2 A2–B3. Furthermore, PCA analysis performed for lipid Raman region showed, that value of PC2 (4.74 %) was positive for samples collected from patients suffering from colon cancer, while almost all samples collected from healthy people showed negative value of this component, Fig. 2 B1. Consequently, this component can be used to differentiate these two groups of samples. Moreover, loading plots visible in Fig. 2 C3 showed, that Raman shift between 2750 cm^{-1} and 2800 cm^{-1} was characteristic for healthy samples, while ranges between 2700 cm^{-1} and 2750 cm^{-1}, as well

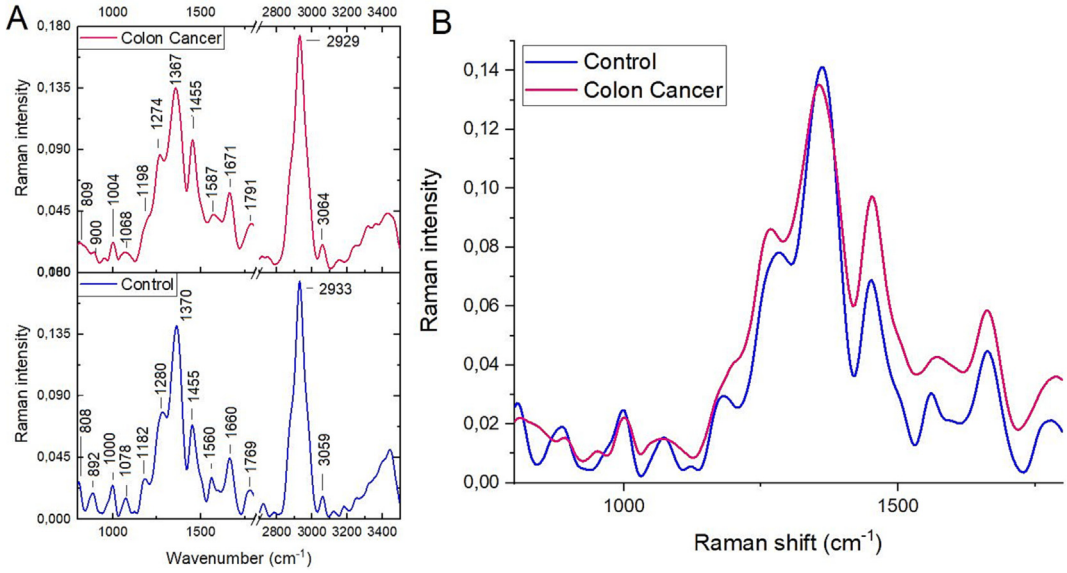

Fig. 1. Offset of average Raman spectra of serum collected from healthy patients (blue line) and patients suffering from colon cancer (pink line) (A), overlap of the obtained spectra in the Raman region between 800 cm^{-1} and 1800 cm^{-1} (B).

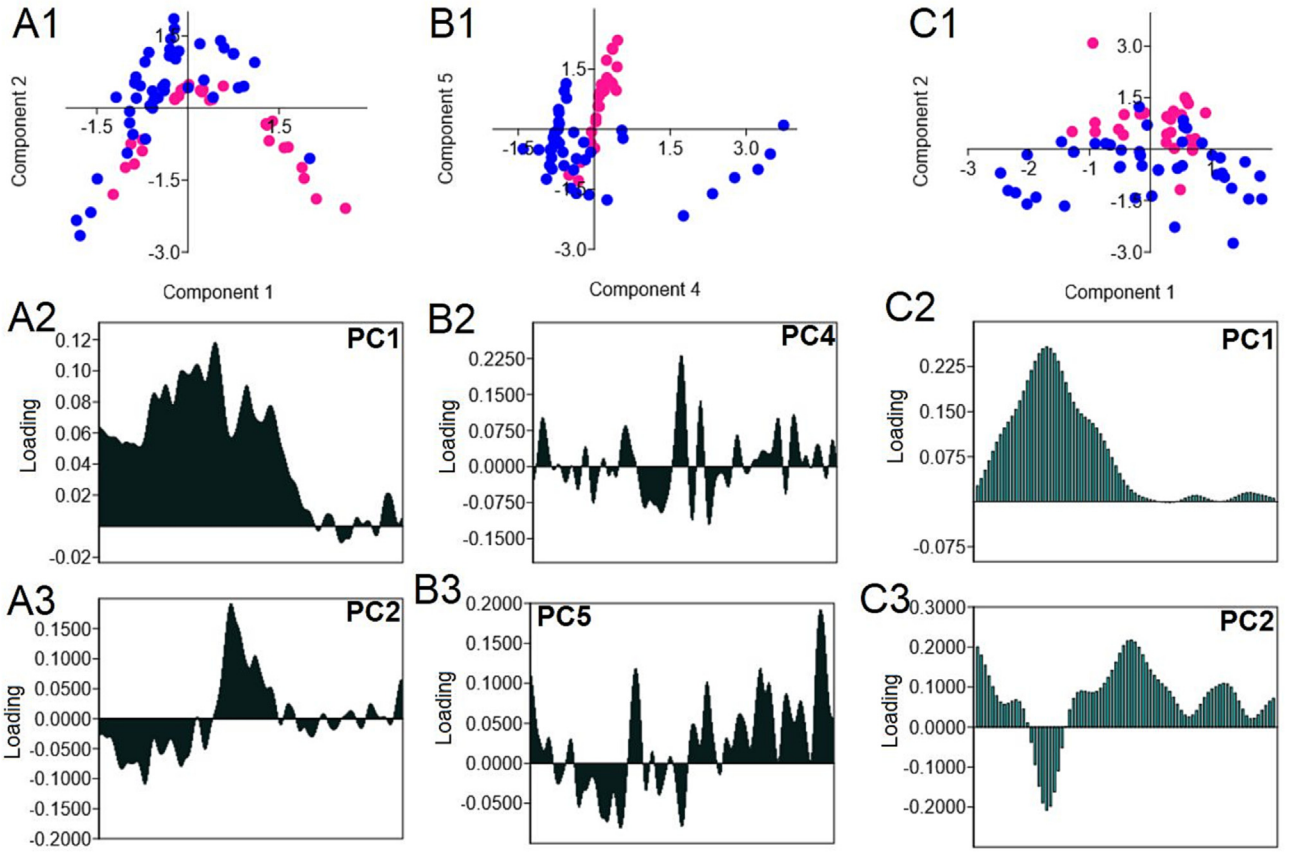

Fig. 2. PCA analysis of Raman range between: 800 cm^{-1} and 1800 cm^{-1} (A1, B1) with PC1, 2, 4, 5 loading plots (A2-B3) and between 2700 cm^{-1} and 3000 cm^{-1} (C1) with PC1 and PC2 loading plots (C2, C3), where blue dots mean healthy patients and pink dots – patients suffering from colon cancer.

as between 2800 cm^{-1} and 3000 cm^{-1} was characteristic for serum collected from healthy patients.

Linear models performed for healthy and un-healthy patients showed, that samples collected from healthy people showed value from 0.05 to 0.12, while value range for colon cancer patients was from 0.00 to 0.19, where almost all samples were placed apart from the values characteristic for patients from the control group, Fig. 3A. Different situation was placed for lipids Raman range, where samples from both group were overlapped to each other, Fig. 3B. However, for two models, the value of correlation was around 1 and 0.96, respectively for Raman range between 800 cm^{-1} and 1800 cm^{-1} and between 2700 cm^{-1} and 3000 cm^{-1}.

To show correlation between Raman shifts visible in Fig. 1 and consequently, to show which Raman shift could be correlated with two colon cancer markers using in diagnostic of this disease, Pearson correlation test was performed, Table 4.

Results obtained from correlation test showed, that only values of CEA colon cancer marker were correlated with Raman shifts. Moreover, it was visible, that all Raman shifts, which were taken for analysis showing correlation with CEA. However, in diagnostic it was very important to show the highest correlation and the most probability value of Raman shifts, therefore, more preciously analyses such as: analysis of intensity dynamics and spectroscopy markers of colon cancer and calculation of accuracy, selectivity and specificity of obtained results were done.

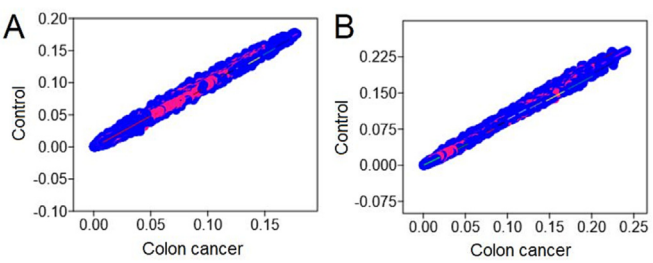

Fig. 3. Linear model of Raman range between: 800 cm^{-1} and 1800 cm^{-1} (A) and between 2700 cm^{-1} and 3000 cm^{-1} (B), where blue dots mean healthy patients and pink dots – patients suffering from colon cancer.

Table 2
One-way Anova Kruskal-Wallis comparison results of the dependent variable as ICAM-1 and grouping variable as Presence of distance metastasis and mucinous component.

Dependent variable/Grouping variable	χ^2	df	p
ICAM-1 level/Presence of distance metastasis	4807	1	0.028*
ICAM-1 level/Musinous component	4.154	1	0.042*

* Noted as $p < 0.05$; statistically significant.

Detailed analysis of Raman data allows to distinguish the regions of Raman shifts which marked the differences in intensity dynamics for Colon cancer patients and Control group. These regions together with the calculated discrimination probabilities are presented in Table 4.

As was visible in Table 4, the discriminations probabilities in the calculated regions of Raman shifts, are very high and allow to detect cancerous patients in terms of Raman spectrum with very high probabilities. Moreover, the Raman shifts at 1302 and 1306 cm^{-1} in the Raman spectrum of serum from colon cancer patients, were visible with 100 % probabilities, which suggest, that this two Raman shifts could be used as a spectroscopic marker of this disease. However, it is also worth to mention that, in order to applied this approach to the clinical trials, it is necessary to create general, objective reference spectra (assumed here as average of control group) for selected groups of patients.

The results of the analysis of the six datasets using machine learning algorithms are shown in Table 5. These results showed the percentage of selected parameters such as accuracy, sensitivity and specificity, as well as area under ROC curve (ROC AUC).

Obtained results showed, that depending on the using algorithm and dataset the accuracy was from 59 % (SVM, 3 dataset) to 96 % (DL and SVM, 6 dataset), while the specificity was from 89 % (DL, 3 dataset) to 100 % (SVM, 3 dataset). Furthermore, compared results obtained for non-reduced and reduced Raman shifts (from 260 to 68, from 78 to 21 and from 338 to 41) it was visible, that the percentages of analyzed parameters were similar, which means, that the using range of Raman shifts can be effectively reduced. In addition, the ROC AUC parameter obtained using the XGBoost model for all 6 sets reaches a value in the range from 95 % to 99 %, which is almost perfect. The ROC curves obtained for each set were shown in Fig. 4.

Discussion

In this paper, we reported an accurate and minimally invasive method for the diagnosis of colorectal cancer and colon cancer specific antigens. Tumor markers provide information about cancer, such as how aggressive it is, which type of treatment it might respond to, or whether it is responding to treatment, such as markers present in cancerous cells or other cells of the body in response to certain benign (noncancerous) conditions. Furthermore, our results suggest that ICAM-1, a cellular adhesion molecule, is involved in the progression and spread of colorectal cancer. In this study we discovered ICAM-1 level is a biomarker for the of presence of distance metastasis and correlates with the musinous components of the tumor. Moreover, the possibility of use Raman spectroscopy in detection of colon cancer from serum and to show which Raman shifts can be used a spectroscopy colon cancer marker, were investigated.

Table 4
Discrimination probabilities for colon cancer patients and control group. The rows are indexed by regions of Raman shifts, which can be used as potential markers to discriminate colon cancer patients by Raman spectrum.

Raman shift (cm^{-1})	D_{Cancer}	$D_{Control}$
854	87 %	100 %
858	90 %	98 %
1279	90 %	80 %
1302	100 %	68 %
1306	100 %	66 %
1726	83 %	65 %
2775	97 %	73 %
2779	93 %	68 %

The value of CEA was higher in people suffering from colon cancer, where higher value around 62 % than reference one give the 5-year disease-free survival was 84.6 % vs. 69.8 %,[57] and the 5-year overall survival was 74.5 % vs. 63.4 %.[58] In our studied group, the values of CEA were higher around 800 %, Table 2. Therefore, the selected group is representative. In the colon cancer patients, differences in the amount of amides and lipids vibrations were visible in Fig. 1. Molecular biology methods made it possible to determine that around 50 % patients suffering from this cancer had mutation of p53 and KRAS gene, which give information about protein playing role in intracellular signal transduction.[59,60] Moreover, changes in the synthesis, desaturation, elongation and mitochondrial oxidation of the lipids were observed in colon cancer cells, which can lead to structural changes in membranes and disruption of energy homeostasis.[61–63] Furthermore, Brozek Pluska et al., also showed, that peaks corresponding to amides and lipids vibrations can be used to distinguish colon cancer and non-cancer tissues.[64] Consequently, it is showed, that analyzed of serum was similar chemical changes such as cancer tissue. Continually, Brozek Pluska et al., also wrote, that Raman shift corresponding to lipids, proteins and carotenoids functional groups and the ratio between these Raman shifts were the most useful biomarkers in spectroscopic diagnostics of colon.[59] While our study showed, that spectroscopy marker of colon cancer was placed at 1302 cm^{-1}

Table 3
Pearson correlation test between Raman shift and value of colon cancer markers (CEA, CA19–9).

Peaks (cm^{-1})	3059	2928	1796	1665	1560	1447	1364	1280	1182	1066	1000	892	804
Cancer marker													
CEA	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.94
CA19-9	0.19	0.20	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19

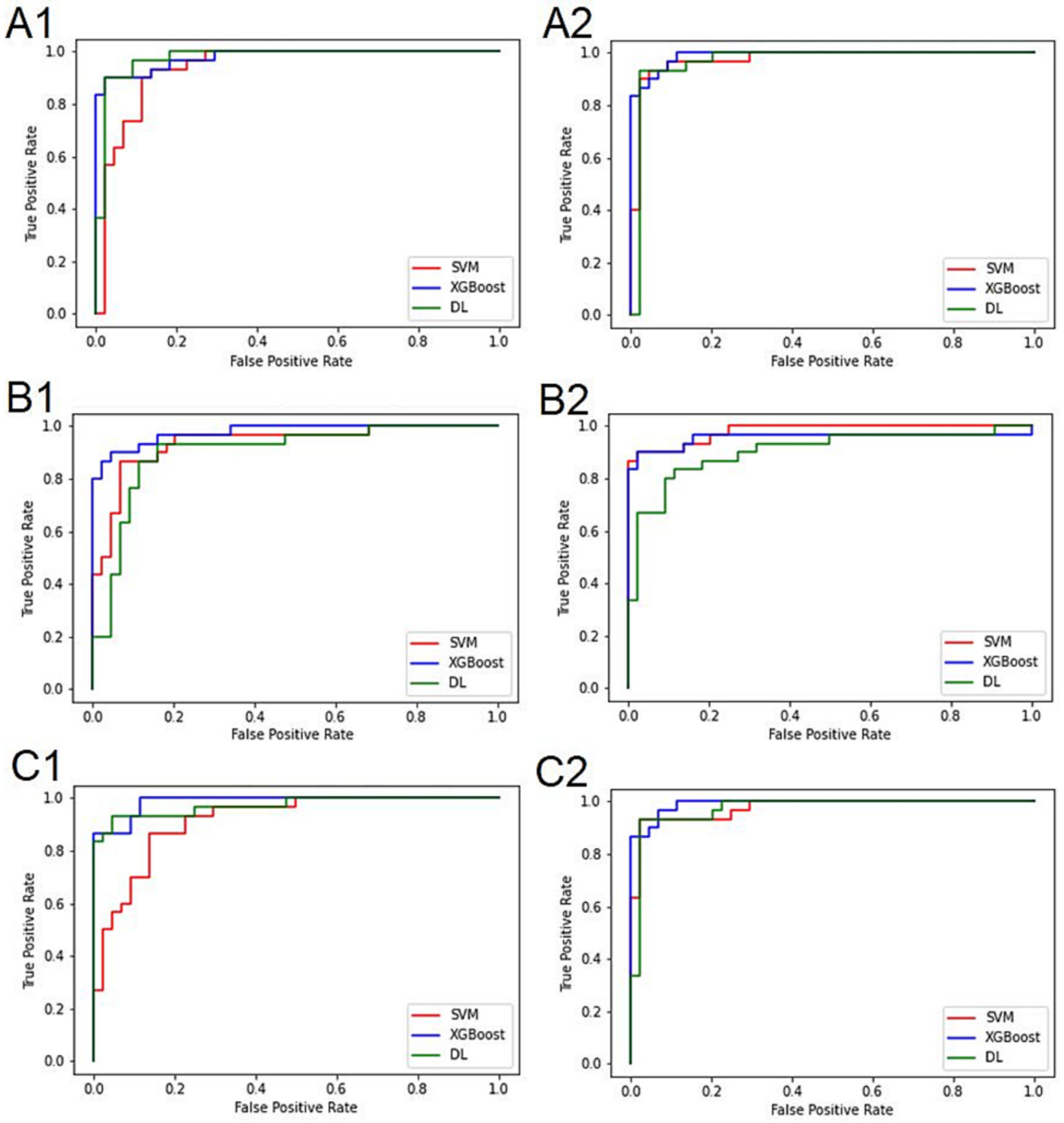

Fig. 4. ROC curves of Raman ranges between 1800 cm^{-1} and 800 cm^{-1} (A); 3000 cm^{-1} and 2700 cm^{-1} (B) and both regions (C) obtained by three different algorithms (SVM – red curves; XGBoost – blue curves and DL – green curves) for 1 dataset (A1); 2 dataset (A2); 3 dataset (B1); 4 dataset (B2); 5 dataset (C1) and 6 dataset (C2).

and 1306 cm^{-1}, Table 5. Petersen et al., also showed, that lipids Raman region differentiate cancer and non-cancer tissues.[65] This differentiation was also visible in PCA analysis presented in Fig. 2 B1. Ito et al., showed, that amides vibrations can be used in diagnostic of colon cancer from serum. They calculated the accuracy of these factor, which had value around 100 %. In other study, the diagnostic accuracy of Raman spectroscopy in colon cancer using multivariate analysis such as linear discriminant analysis (LDA) was around 83 %.[66] Our data showed, that using machine learning methods, the value of this factor was around 95 %, Table 5.

Summarizing, our data showed, that Raman spectroscopy marker of colon cancer were placed at 1302 cm^{-1} and 1306 cm^{-1} Moreover, using this method distinguish between serum collected from patients suffering from colon cancer and healthy one with accuracy around 100 %, were obtained.

Table 5
Mean values of the quality assessment parameters (accuracy – acc, sensitivity – sens. and specificity – spec.) of the obtained learning models, using 6 datasets and 3 learning models, algorithms.

Dataset	No. of features	Algorithm	acc (%)	sens. (%)	spec. (%)	ROC AUC (%)
1	260	DL	95	90	98	98
		SVM	80	57	95	94
		XGBoost	91	90	91	98
2	68	DL	95	93	95	97
		SVM	95	90	98	97
		XGBoost	93	90	95	99
3	78	DL	85	80	89	90
		SVM	59	0	100	94
		XGBoost	91	92	91	98
4	21	DL	84	67	95	90
		SVM	93	83	100	98
		XGBoost	91	90	91	96
5	338	DL	93	90	95	97
		SVM	80	57	95	92
		XGBoost	91	90	91	99
6	43	DL	96	93	98	97
		SVM	96	93	98	98
		XGBoost	93	90	95	99

However, what is very important to develop Raman spectroscopy technique, e.g. to show chemical changes in in individual macromolecules, where first mutations could be responsible for expansion of the diseases. The first attempts were shown at work Shi et al., where using stimulated Raman scattering (DO-SRS) microscopy it was possible to image protein biosynthesis without tissue bias, and simultaneously visualize lipid and protein metabolism and reveal their different dynamics.[67] In situ investigation of Raman spectroscopy was also found in Zhang et al. work, where glucose metabolism was traced.[68] Finally, de novo lipogenesis, 3D lipid droplet morphology, and lipid peroxidation under various methionine and insulin concentrations in the triple negative breast cancer were presented by Raman spectroscopy.[69]

CRediT authorship contribution statement

J.D, Z.G.: Conception and design, data analysis and interpretation, original draft, review and editing, approval of final article.

P.J., W. P., K. P., A. W.; data analysis, approval of final article.

M. K-M.: measurement, approval of final article.

I.Y., G.I.G., N.T., M.T.H., D.S., D.S., S.A.: biochemical tests, approval of final article.

Declaration of competing interest

The authors have no disclosures.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**:209-49, https://doi.org/10.3322/caac.21660.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CACancer J Clin* 2019;**69**:7-34, https://doi.org/10.3322/caac.21551.
3. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. *Ann Intern Med* 2009;**150**:162-9, https://doi.org/10.7326/0003-4819-150-3-200902030-00005.
4. Graser A, Stieber P, Nagel D, Schäfer C, Horst D, Becker CR, et al. Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma in an average risk population. *Gut* 2009;**58**:241-8, https://doi.org/10.1136/gut.2008.156448.
5. Dou Y, Lv Y, Zhou X, He L, Liu L, Li P, et al. Antibody-sandwich ELISA analysis of a novel blood biomarker of CST4 in gastrointestinal cancers. *OncoTargets Ther* 2018;**11**:1743, https://doi.org/10.2147/OTT.S149204.
6. Gold P, Freedman SO. Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques. *J Exp Med* 1965;**121**:439-62, https://doi.org/10.1084/jem.121.3.439.
7. Konishi T, Shimada Y, Hsu M, Tufts L, Jimenez-Rodriguez R, Cercek A, et al. Association of preoperative and postoperative serum carcinoembryonic antigen and colon cancer outcome. *JAMAOncol* 2018;**4**:309-15, https://doi.org/10.1001/jamaoncol.2017.4420.
8. Duffy MJ, van Dalen A, Haglund C, Hansson L, Holinski-Feder E, Klapdor R, et al. Tumour markers in colorectal cancer: european group on tumour markers (EGTM) guidelines for clinical use. *Eur J Cancer* 2007;**43**:1348-60, https://doi.org/10.1016/j.ejca.2007.03.021.
9. Huh JW, Oh BR, Kim HR, Kim YJ. Preoperative carcinoembryonic antigen level as an independent prognostic factor in potentially curative colon cancer. *J Surg Oncol* 2010;**101**:396-400, https://doi.org/10.1002/jso.21495.
10. Omar YT, Al-Naqeeb N, El Nas SA, Awwad AH, Foudeh MO, Safadi NB, et al. Serum levels of CA 125 in patients with gastrointestinal cancers. *Tumor Biol* 1989;**10**:316-23, https://doi.org/10.1159/000217631.
11. Gao Y, Wang J, Zhou Y, Sheng S, Qian SY, Huo X. Evaluation of serum CEA, CA19-9, CA72-4, CA125 and ferritin as diagnostic markers and factors of clinical parameters for colorectal cancer. *Sci Rep* 2018;**8**:2732, https://doi.org/10.1038/s41598-018-21048-y.
12. Malati T. Tumour markers: an overview. *indianJ Clin Biochem* 2007;**22**:17-31, https://doi.org/10.1007/BF02913308.
13. Frenette PS, Thirlwell MP, Trudeau M, Thomson DMP, Joseph L, Shuster JS. The diagnostic value of CA 27–29, CA 15–3, mucin-like carcinoma antigen, carcinoembryonic antigen and CA 19–9 in breast and gastrointestinal malignancies. *Tumor Biol* 1994;**15**:247-54, https://doi.org/10.1159/000217898.
14. Lakemeyer L, Sander S, Wittau M, Henne-Bruns D, Kornmann M, Lemke J. Diagnostic and prognostic value of CEA and CA19-9 in colorectal cancer. *Diseases* 2021;**9**:21, https://doi.org/10.3390/diseases9010021.
15. Kim S, Park BK, Seo JH, Choi J, Choi JW, Lee CK, et al. Carbohydrate antigen 19–9 elevation without evidence of malignant or pancreatobiliary diseases. *Sci Rep* 2020;**10**, https://doi.org/10.1038/s41598-020-65720-8.
16. Bergstrand CG, Czar B. Demonstration of a new protein fraction in serum from the human fetus. *Scand J Clin Lab Invest* 1956;**8**:174, https://doi.org/10.3109/00365515609049266.
17. Benson AB, Abrams TA, Ben-Josef E, Bloomston PM, Botha JF, Clary BM, et al. Hepatobiliary cancers. *J Natl Compr Cancer Netw* 2009;**7**:350, https://doi.org/10.6004/jnccn.2009.0027.
18. Ren F, Weng W, Zhang Q, Tan C, Xu M, Zhang M, et al. Clinicopathological features and prognosis of AFP-producing colorectal cancer: a single-center analysis of 20 cases. *Cancer Manag Res* 2019;**11**:4557-67, https://doi.org/10.2147/CMAR.S196919.
19. Ohene-Abuakwa Y, Pignatelli M. Adhesion molecules as diagnostic tools in tumor pathology. *Int J Surg Pathol* 2000;**8**:191-200, https://doi.org/10.1177/106689690000800306.
20. Abe Y, El-Masri B, Kimball KT, Pownall H, Reilly CF, Osmundsen K, et al. Soluble cell adhesion molecules in hypertriglyceridemia and

potential significance on monocyte adhesion. *Arterioscler Thromb Vasc Biol* 1998;**18**:723-31, https://doi.org/10.1161/01.ATV.18.5.723.
21. Benedicto A, Herrero A, Romayor I, Marquez J, Smedsrød B, Olaso E, et al. Liver sinusoidal endothelial cell ICAM-1 mediated tumor/endothelial crosstalk drives the development of liver metastasis by initiating inflammatory and angiogenic responses. *Sci Rep* 2019;**9**:1-12, https://doi.org/10.1038/s41598-019-49473-7.
22. Oettle H, Post S, Neuhaus P, Gellert K, Langrehr J, Ridwelski K, et al. Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: a randomized controlled trial. *JAMA* 2007;**297**:267-77, https://doi.org/10.1001/jama.297.3.267.
23. Beauchemin N, Draber P, Dveksler G, Gold P, Gray-Owen S, Grunert F, et al. Redefined nomenclature for members of the carcinoembryonic antigen family. *Exp Cell Res* 1999;**252**:243-9, https://doi.org/10.1006/excr.1999.4610.
24. Gebauer F, Wicklein D, Horst J, Sundermann P, Maar H, Streichert T, et al. Carcinoembryonic antigen-related cell adhesion molecules (CEACAM) 1, 5 and 6 as biomarkers in pancreatic cancer. *PLoS One* 2014;**9**, https://doi.org/10.1371/journal.pone.0113023.
25. Yaylim I, Butt G, Khalid S, Farooqi AA. Role of CEACAM in different cancers, recent trends cancer biol. Spotlight signal. Cascades microRNAs cell signal. *Pathways MicroRNAs Cancer Biol* 2018:293-300, https://doi.org/10.1007/978-3-319-71553-7_16.
26. Li N, Yang JY, Wang XY, Wang HT, Guan BX, Zhou CJ. Carcinoembryonic antigen-related cell adhesion molecule 1 is expressed and as a function histotype in ovarian tumors. *Ann Diagn Pathol* 2016;**20**:7-12, https://doi.org/10.1016/j.anndiagpath.2015.10.012.
27. Han ZM, Huang HM, Sun YW. Effect of ceacam-1 knockdown in human colorectal cancer cells. *Oncol Lett* 2018;**16**:1622-6, https://doi.org/10.3892/ol.2018.8835.
28. Fung AA, Shi L. Mammalian cell and tissue imaging using Raman and coherent Raman microscopy. *WIREs Systems Biology and Medicine* 2020e1501, https://doi.org/10.1002/wsbm.1501.
29. Shi L, Fung AA, Andy Z. Advances of stimulated Raman scattering imaging in tissues and animals. *Quant Imaging Med Surg* 2021;**11**(3), https://doi.org/10.21037/qims-20-712.
30. Butler HJ, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, et al. Using Raman spectroscopy to characterize biological materials. *Nat Protoc* 2016;**11**:664-87, https://doi.org/10.1038/nprot.2016.036.
31. Depciuch J, Parlinska-Wojtan M, Serin KRahmi, Bulut H, Ulukaya E, Tarhan N, et al. Differential of cholangiocarcinoma disease using Raman spectroscopy combined with multivariate analysis. *Spectrochim Acta - Part A Mol Biomol Spectrosc* 2022;**272**, https://doi.org/10.1016/j.saa.2022.121006.
32. Guleken Z, Bulut H, Gültekin Gİ, Arıkan S, Yaylım İ, Hakan MT, et al. Assessment of structural protein expression by FTIR and biochemical assays as biomarkers of metabolites response in gastric and colon cancer. *Talanta* 2021;**231**122353, https://doi.org/10.1016/J.TALANTA.2021.122353.
33. Wang G, Lipert RJ, Jain M, Kaur S, Chakraboty S, Torres MP, et al. Detection of the potential pancreatic cancer marker MUC4 in serum using surface-enhanced Raman scattering. *Anal Chem* 2011;**83**:2554-61, https://doi.org/10.1021/ac102829b.
34. Kaznowska E, Łach K, Depciuch J, Chaber R, Koziorowska A, Slobodian S, et al. Application of infrared spectroscopy for the identification of squamous cell carcinoma (lung cancer). Preliminary study. *Infrared Phys Technol* 2018;**89**:282-90, https://doi.org/10.1016/j.infrared.2018.01.021.
35. Depciuch J, Kaznowska E, Zawlik I, Wojnarowska R, Cholewa M, Heraud P, et al. Application of raman spectroscopy and infrared spectroscopy in the identification of breast cancer. *Appl Spectrosc* 2016;**70**:251-63, https://doi.org/10.1177/0003702815620127.
36. Guleken Z, Bulut H, Bulut B, Paja W, Parlinska-Wojtan M, Depciuch J. Correlation between endometriomas volume and Raman spectra. Attempting to use raman spectroscopy in the diagnosis of endometrioma. *Spectrochim Acta Part A Mol Biomol Spectrosc* 2022121119, https://doi.org/10.1016/J.SAA.2022.121119.
37. Guleken Z, Bulut H, Depciuch J, Tarhan N. Diagnosis of endometriosis using endometrioma volume and vibrational spectroscopy with multivariate methods as a noninvasive method. *Spectrochim Acta - Part A Mol Biomol Spectrosc* 2022;**264**120246, https://doi.org/10.1016/j.saa.2021.120246.
38. Li S, Li L, Zeng Q, Zhang Y, Guo Z, Liu Z, et al. Characterization and noninvasive diagnosis of bladder cancer with serum surface enhanced Raman spectroscopy and genetic algorithms. *Sci Rep* 2015;**5**:1-7, https://doi.org/10.1038/srep09582.
39. Sato S, Sekine R, Kagoshima H, Kazama K, Kato A, Shiozawa M, et al. All-in-one raman spectroscopy approach to diagnosis of colorectal cancer: analysis of spectra in the fingerprint regions. *J Anus Rectum Colon* 2019;**3**:84-90, https://doi.org/10.23922/jarc.2018-039.
40. Lin D, Feng S, Pan J, Chen Y, Lin J, Chen G, et al. Colorectal cancer detection by gold nanoparticle based surface-enhanced raman spectroscopy of blood serum and statistical analysis. *Opt Express* 2011;**19**:13565, https://doi.org/10.1364/oe.19.013565.
41. Ito H, Uragami N, Matsuo K, Yokoyama N, Inoue H, Miyazaki T, et al. Highly accurate colorectal cancer prediction model based on Raman spectroscopy using patient serum, world. *J Gastrointest Oncol* 2020;**12**:1311-24, https://doi.org/10.4251/wjgo.v12.i11.1311.
42. Kamezaki S, Kurozawa Y, Iwai N, Hosoda T, Okamoto M, Nose T. Serum levels of soluble ICAM-1 and VCAM-1 predict pre-clinical cancer. *Eur J Cancer* 2005;**41**:2355-9, https://doi.org/10.1016/j.ejca.2005.07.005.
43. Wright PF, Nilsson E, Van Rooij EM, Lelenta M, Jeggo MH. Standardisation and validation of enzyme-linked immunosorbent assay techniques for the detection of antibody in infectious disease diagnosis. *Rev Sci Tech* 1993;**12**:435-50, https://doi.org/10.20506/rst.12.2.691.
44. Guleken Z, Tok YTuyji, Jakubczyk P, Paja W, Pancerz K, Shpotyuk Y, et al. Development of novel spectroscopic and machine learning methods for the measurement of periodic changes in COVID-19 antibody level. *Measurement* 2022:111258, https://doi.org/10.1016/J.MEASUREMENT.2022.111258.
45. Guleken Z, Bulut H, Gültekin Gİ, Arıkan S, Yaylım İ, Hakan MT, et al. Assessment of structural protein expression by FTIR and biochemical assays as biomarkers of metabolites response in gastric and colon cancer. *Talanta* 2021;**231**, https://doi.org/10.1016/j.talanta.2021.122353.
46. Chaber R, Kowal A, Jakubczyk P, Arthur C, Łach K, Wojnarowska-Nowak R, et al. A preliminary study of FTIR spectroscopy as a potential non-invasive screening tool for pediatric precursor B lymphoblastic leukemia. *Molecules* 2021;**26**:1174, https://doi.org/10.3390/molecules26041174.
47. Guleken Z, Jakubczyk P, Wiesław P, Krzysztof P, Bulut H, Öten E, et al. Characterization of Covid-19 infected pregnant women sera using laboratory indexes, vibrational spectroscopy, and machine learning classifications. *Talanta* 2022;**237**122916, https://doi.org/10.1016/j.talanta.2021.122916.
48. Goodfellow Y, Bengio I, Courville Y, Bengio A. *Deep Learning*. The MIT Press; 2016 (accessed May 25, 2021) https://mitpress.mit.edu/books/deep-learning.
49. Zaknich A. Statistical learning theory. *Neural Netw Intell Signal Process* 2003:359-96, https://doi.org/10.1142/9789812796851_0014.
50. Guleken Z, Ünübol B, Bilici R, Sarıbal D, Toraman S, Gündüz O, et al. Investigation of the discrimination and characterization of blood serum structure in patients with opioid use disorder using IR spectroscopy and PCA-LDA analysis. *J Pharm Biomed Anal* 2020;**190**113553, https://doi.org/10.1016/j.jpba.2020.113553.
51. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min*; 2016. p. 785-94, https://doi.org/10.1145/2939672.2939785. 13–17-Augu.
52. Rudnicki WR, Wrzesień M, Paja W. All relevant feature selection methods and applications. *Stud Comput Intell* 2015;**584**:11-28, https://doi.org/10.1007/978-3-662-45620-0_2.

53. Paja W. Feature selection methods based on decision rule and tree models. *Smart Innov Syst Technol* 2016;**57**:63-70, https://doi.org/10.1007/978-3-319-39627-9_6.
54. Takeuchi A, Yokoyama S, Nakamori M, Nakamura M, Ojima T, Yamaguchi S, et al. Loss of CEACAM1 is associated with poor prognosis and peritoneal dissemination of patients with gastric cancer. *Sci Rep* 2019;**9**:1-9, https://doi.org/10.1038/s41598-019-49230-w.
55. Yu Y, Ramachandran PV, Wang MC. Shedding new light on lipid functions with CARS and SRS microscopy. *Biochim Biophys Acta - Mol Cell Biol Lipids* 2014;**1841**:1120-9, https://doi.org/10.1016/j.bbalip.2014.02.003.
56. Depciuch J, Tołpa B, Witek P, Szmuc K, Kaznowska E, Osuchowski M, et al. Raman and FTIR spectroscopy in determining the chemical changes in healthy brain tissues and glioblastoma tumor tissues. *Spectrochim Acta - Part A Mol Biomol Spectrosc* 2020;**225**117526, https://doi.org/10.1016/j.saa.2019.117526.
57. Ozawa H, Kotake K, Hosaka M, Hirata A, Nakagawa Y, Fujita S, et al. Incorporation of serum carcinoembryonic antigen levels into the prognostic grouping system of colon cancer. *Int J Colorectal Dis* 2017;**32**:821-9, https://doi.org/10.1007/s00384-017-2772-1.
58. Spindler BA, Bergquist JR, Thiels CA, Habermann EB, Kelley SR, Larson DW, et al. Incorporation of CEA improves risk stratification in stage II colon cancer. *J Gastrointest Surg* 2017;**21**:770-7, https://doi.org/10.1007/s11605-017-3391-4.
59. Cappell MS. Pathophysiology, clinical presentation, and management of colon cancer. *Gastroenterol Clin North Am* 2008;**37**:1-24, https://doi.org/10.1016/j.gtc.2007.12.002.
60. Tan C, Du X. KRAS mutation testing in metastatic colorectal cancer. *World J Gastroenterol* 2012;**18**:5171-80, https://doi.org/10.3748/wjg.v18.i37.5171.
61. Huang C, Freter C. Lipid metabolism, apoptosis and cancer therapy. *Int J Mol Sci* 2015;**16**:924-49, https://doi.org/10.3390/ijms16010924.
62. Sánchez-Martínez R, Cruz-Gil S, García-Álvarez MS, Reglero G, De Molina AR. Complementary ACSL isoforms contribute to a non-Warburg advantageous energetic status characterizing invasive colon cancer cells. *Sci Rep* 2017;**7**:1-15, https://doi.org/10.1038/s41598-017-11612-3.
63. Mika A, Kobiela J, Czumaj A, Chmielewski M, Stepnowski P, Sledzinski T. Hyper-elongation in colorectal cancer tissue - cerotic acid is a potential novel serum metabolic marker of colorectal malignancies. *Cell Physiol Biochem* 2017;**41**:722-30, https://doi.org/10.1159/000458431.
64. Brozek-Pluska B, Musial J, Kordek R, Abramczyk H. Analysis of human colon by raman spectroscopy and imaging-elucidation of biochemical changes in carcinogenesis. *Int J Mol Sci* 2019;**20**:3398, https://doi.org/10.3390/ijms20143398.
65. Petersen D, Mavarani L, Niedieker D, Freier E, Tannapfel A, Kötting C, et al. Virtual staining of colon cancer tissue by label-free raman micro-spectroscopy. *Analyst* 2017;**142**:1207-15, https://doi.org/10.1039/c6an02072k.
66. Li X, Yang T, Yu T, Li S. Discrimination of serum raman spectroscopy between normal and colorectal cancer. *Opt InfoBase Conf Pap* 2011;**8087**:391-5, https://doi.org/10.1117/12.889304.
67. Shi L, Zheng C, Shen Y, Chen Z, Silveira E, Zhang I, et al. Optical imaging of metabolic dynamics in animals. *Nat Commun* 2018;**9**:2995, https://doi.org/10.1038/s41467-018-05401-3.
68. Zhang L, Shi L, Shen Y, Miao Y, We M, Qian N, et al. Spectral tracing of deuterium for imaging glucose metabolism. *Nat Biomed Eng* 2019;**3**(5):402-13, https://doi.org/10.1038/s41551-019-0393-4.
69. Fung AA, Hoang K, Zha H, Chen D, Zhang W, Shi L. *Imaging Sub-Cellular Methionine and Insulin Interplay in Triple Negative Breast Cancer Lipid Droplet Metabolism. Frontier Oncology, Cancer Metabolism*; 2022, https://www.frontiersin.org/articles/10.3389/fonc.2022.858017/full.