Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Development of novel spectroscopic and machine learning methods for the measurement of periodic changes in COVID-19 antibody level

Zozan Guleken a, Yeşim Tuyji Tok b, Paweł Jakubczyk c, Wiesław Paja d, Krzysztof Pancerz e, Yaroslav Shpotyuk c, Jozef Cebulski c, Joanna Depciuch f, * 

a Uskudar University, Faculty of Medicine, Department of Physiology, Turkey  
b Department of Medical Microbiology, Cerrahpaşa Medical Faculty, Istanbul University-Cerrahpaşa, Turkey  
c Institute of Physics, University of Rzeszów, Poland  
d Institute of Computer Science, University of Rzeszów, Poland  
e Institute of Philosophy, John Paul II Catholic University of Lublin, Poland  
f Institute of Nuclear Physics Polish Academy of Science, 31-342 Krakow, Poland

ARTICLE INFO

Keywords:  
Anti-SARS-CoV-2 antibody levels  
FTIR  
Raman spectroscopy  
Machine learning

ABSTRACT

In this research, blood samples of 47 patients infected by COVID were analyzed. The samples were taken on the 1st, 3rd and 6th month after the detection of COVID infection. Total antibody levels were measured against the SARS-CoV-2 N antigen and surrogate virus neutralization by serological methods. To differentiate COVID patients with different antibody levels, Fourier Transform InfraRed (FTIR) and Raman spectroscopy methods were used. The spectroscopy data were analyzed by multivariate analysis, machine learning and neural network methods. It was shown, that analysis of serum using the above-mentioned spectroscopy methods allows to differentiate antibody levels between 1 and 6 months via spectral biomarkers of amides II and I. Moreover, multivariate analysis showed, that using Raman spectroscopy in the range between 1317 cm^{-1} and 1432 cm^{-1}, 2840 cm^{-1} and 2956 cm^{-1} it is possible to distinguish patients after 1, 3, and 6 months from COVID with a sensitivity close to 100%.

1. Introduction

The coronavirus disease 2019 (COVID-19), which is caused by severe respiratory syndrome coronavirus 2 (SARS-CoV-2) was first reported in Wuhan, China in December 2019 and has spread among humans across the continents and eventually was recognized as a pandemic on March 11, 2020 [1,2]. SARS-CoV-2 is a zoonotic, positive polarity RNA virus, which includes structural proteins, namely, spike (S), envelope (E), membrane (M), nucleocapsid (N), as well as 16 non-structural proteins, that are involved in viral infection and replication [3,4]. SARS-CoV-2 infection initiates an immune response that targets these proteins and includes the production of neutralizing antibodies. The antibodies remain detectable in the plasma for months. These antibodies bind rapidly and strongly to viral antigens, blocking cellular infiltration and replication of the virus [5,6,7].

The plaque reduction neutralization test (PRNT) is the gold standard for demonstrating protective antibody titer in viral infections. However, the practical use of such a laborous method is limited due to the biosafety requirement (level-3 laboratory is mandatory). It is still unclear at what titer serum antibodies are protective from COVID-19, but studies have shown a high correlation between cell-based neutralization assays and serum antibody titers. In most patients, SARS-CoV-2 forms a self-limiting infection, and soon within 1–3 weeks on the average after antigens appear in the blood, the antibodies reach a level that can be serologically detected in the plasma [8,9,10,11]. Dynamic measurement of total anti-SARS-CoV-2 antibody levels and periodic monitoring of dynamic titer bear great importance in epidemiological studies such as determining exposure and infection rates in the population, monitoring the development of herd immunity. Antibody level monitoring can also be used to determine disease severity and to design preventive measures against the risk of re-infection [12]. The level of antibody reflects the dependence on the amides and lipids metabolisms [13]. It was shown, that people with symptomatic COVID-19 have an almost 15-fold reduction in the serum of the sphingosine consistent with the presence of acid ceramidase protein, as compared to asymptomatic COVID-19 patients [14]. Previous studies in cohorts selected from the society

* Corresponding author.  
E-mail address: joanna.depciuch@ifj.edu.pl (J. Depciuch).

https://doi.org/10.1016/j.measurement.2022.111258  
Received 8 February 2022; Received in revised form 20 April 2022; Accepted 22 April 2022  
Available online 26 April 2022  
0263-2241/© 2022 Elsevier Ltd. All rights reserved.
Table 1 (continued)

| ID | Sex   | Age | Symptoms                        | Comorbidities          |
|----|-------|-----|---------------------------------|------------------------|
| 43 | Female| 42  | Fever, diarrhea                 |                        |
| 44 | Female| 50  | Myalgia, loss of smell and taste |                        |
| 45 | Female| 32  | Asymptomatic                    |                        |
| 46 | Female| 56  | Fever, weakness, taste loss,    | Hypertension            |
|    |       |     | muscle pain                     |                        |
| 47 | Female| 62  | Fever, weakness                 |                        |

have shown that antibodies disappear in an average period of four months [15,16]. On the other hand, the community constantly encounters severe patients that require hospitalization, hence are faced with higher viral load.

In recent years, spectroscopic methods combined with multivariate analyzes and machine learning methods have been used to differentiate tissues, cells and body fluids from sick and healthy people [17,18]. The possibility of using these methods in medical diagnostics as well as in the diagnostics of viruses is investigated. Khan et al. showed that Raman spectroscopy analysis of blood serum can distinguish (with 97% of precision, 100% of sensitivity and 95% of specificity) between a healthy patient and a patient infected with hepatitis B virus [19]. Furthermore, in the other work, the authors showed, that using FTIR spectroscopy it is possible to distinguish between females suffering from symptomatic and asymptomatic COVID-19 [20]. In the other work, machine learning methods showed, that using FTIR data it is possible to detect COVID-19 infection with 87% of sensitivity [21]. In addition, Liu et al. showed high sensitivity of surface-enhanced Raman scattering (SERS) in the detection of the anti-SARS-CoV-2 IgM/IgG [22]. These authors proposed a sensor, which gives 800 times higher detection of antibodies than the standard IgM and IgG sensor. Consequently, FTIR and Raman spectroscopy can be used to detect COVID-19. However, in our work, besides the possibility to distinguish between the blood serum of people with COVID-19 and healthy people, the main attention was paid to the use of spectroscopies methods to predict the SARS-CoV-2 antibody timeline, estimate reinfection risk and the need for the vaccination by periodically assessing total antibody levels in people, who have recovered from COVID-19. For this purpose, FTIR and Raman spectroscopy methods in combination with multivariate analyses were used. To investigate the possibility of FTIR and Raman methods to distinguish patients with different antibody levels, the obtained spectroscopic data were further analyzed by machine learning and neural network methods [2]. In addition, the correlation between antibody and amides as well as lipids levels were analyzed in the respective vibration regions of Raman and FTIR.

2. Materials and methods

2.1. Patient selection

For this research, 47 patients aged 26–77 (healthcare employees) infected by COVID (confirmed by SARS-CoV-2 RT–PCR testing) were involved. Demographic data such as age, gender, history of additional disease/drug use, symptoms and treatments were recorded for each patient on the prepared form. Collected data is shown in Table 1.

The results of biochemical (erythrocyte sedimentation rate (ESH), leukocyte, hemoglobin, platelet counts, coagulation, liver enzymes, D-dimer, C-reactive protein) and radiological (lung radiography / computerized tomography (CT)) examinations were recorded by inquiring from the hospital information system. Positive RT-PCR test of COVID-19, age over 18, and belonging to the group of health care employees were used as selection criteria. All the selected patients agreed to give a sample for antibody measurement more than once, including the first blood sample after 1 month from the first diagnosis, the second one after 3 months and the last one – after 6 months. The severity of
COVID-19 was determined according to the WHO (World Health Organization) criteria and the patients were evaluated separately as asymptomatic/symptomatic (mild-severe) [23].

2.2. Laboratory study

Peripheral blood samples were taken on the 1st, 3rd, and 6th month after the diagnosis of positive asymptomatic or symptomatic COVID-19. Total antibody level against SARS-CoV-2 N antigen was measured in ECLIA (Eлектrochemiluminescence immunoassay)-based, semi-quantitative Elecsys anti-N serological test (Roche Diagnostics Rotkreuz, Switzerland) in Cobas e601 device. The results were expressed as COI (cut-off index) and values above one (>1 COI) were evaluated as the reactivity following the manufacturer’s recommendations.

2.3. FTIR measurements

The Bruker Vertex 70v spectrometer equipped with an attenuated total reflection (ATR) diamond crystal plate was used to study the absorbance spectra of serum samples. The measurements were performed in 400–4000 cm⁻¹ region with 32 scans for each sample, and spectral resolution of 4 cm⁻¹. Collected blood samples were centrifuged (15 min, 3000 rpm) then serum was separated and stored (at −80 °C) till spectroscopic measurements. Before FTIR measurement, the samples were thawed, spotted on a CaF₂ slide and dried. The ATR crystal diamond plate was cleaned each time before the measurements [24]. All spectra were analyzed using OPUS 7.0 software. For each spectrum, the baseline correction, vector normalization and smoothing with Savitzky–Golay filter were performed [20].

2.4. FT-Raman measurements

Blood serum collected from patients after 1, 3 and 6 months after COVID-19 positive test (group I, II and III, respectively) was measured using FT-Raman spectrometer Nicolet XNR 9650. This spectrometer is equipped with a Nd: YAG laser operating at 1064 nm wavelength (power was set to 1 W). Before the measurement, the same sample preparation procedure (as for FTIR) was performed. Blood samples were centrifuged (15 min, 3000 rpm), then serum was separated and stored (at −80 °C). Before measurement, the samples were thawed. A volume of 4 µL of each serum sample dropped on a gold support was used for measurements. All spectra were recorded in the range from 150 to 3700 cm⁻¹ with 8 cm⁻¹ spectral resolution and 64 scans. Raman spectra were processed by the Omnic/Thermo Scientific software.

2.5. Multivariate analysis of obtained FTIR and FT-Raman spectra

To increase the distinguish efficiency between serum collected from patients after 1, 3 and 6 months after COVID-19 using FTIR and FT-Raman spectroscopies, as well as the similarity between samples in each analyzed group, Principal Component Analysis (PCA) and Hierarchical Component Analysis (HCA) were performed. Moreover, to identify the wavenumbers which could be used to differentiate serum from patients at different periods after COVID-19, the Partial Least Square (PLS) analysis was done. PCA and HCA analyses were performed using Past 3.0. software, while PLS was done using Origin 2019 software. PLS analyses were performed in the 800–4000 cm⁻¹ region of FTIR and 500–3700 cm⁻¹ region of Raman spectra, while PCA and HCA analyses were used for the range between 800 cm⁻¹ and 1800 cm⁻¹, as well as between regions, which were obtained from PLS analysis.

2.6. Analysis of FTIR absorbance dynamics

To mark the difference between the three analyzed groups and to confirm the conducted statistical analysis the absorbance dynamics method was applied. This method is based on the fact that in IR spectra, for a carefully selected range of wavenumbers, two types of absorbance dynamics can be noticed. In the case of the first type $\frac{dA}{dk} > 0$, and in the case of the second type $\frac{dA}{dk} < 0$, where A denotes the FTIR spectrum and k is the wavenumber. This means that for the first type of dynamics the absorption is increasing with k, while in the second is decreasing. The simple difference between such defined dynamics for wavenumber k can be used as an indicator of the differences between the IR spectra under consideration and the appropriately calibrated value of the first derivative is a good measure of the differences. This is a very accurate method with the resolution of the order of a single wavenumber. In this research, the analyses were performed in the 1500–1800 cm⁻¹ and 2700–3000 cm⁻¹ regions of FTIR spectra, as according to PLS analysis, only in these regions the significant differences in the absorbance for the analyzed group can be noticed.

2.7. Data analysis using machine learning methods

To acquire the knowledge about the accuracy of FTIR spectroscopy in separating evaluated samples, six machine learning methods were used:

- Random forest (RF), [25].
- C5.0 decision tree algorithm, [26].
- Deep Neural Networks (DNN) [27].
- k-nearest Neighbors (kNN) [28].
- XGBoost trees [29].
- Support Vector Machine (SVM) [30].

Appropriate datasets in a tabular form were created to classify the cases. The goal of building classification models was to check whether these models can distinguish between cases belonging to different categories (treated as decision classes): 1 month after COVID-19; 3 months after COVID-19; 6 months after COVID-19. The datasets consisted of rows (each row corresponds to one patient), columns representing features describing patients (wavenumbers of single peaks), and a decision column containing the categories mentioned earlier. Each dataset, before a feature selection process, contains 156 features (wavenumbers) and 1 decision attribute.

The experiments were performed using the R environment in which the following packages were used: mltools, class, randomForest, C5.0, kersas, xgboost and e1071. Additionally, the Boruta package [31] was used to perform the selection process for the most relevant features (wavenumbers) which have the greatest impact on the assignment of the category by evaluating the importance of each descriptive wavenumber. This approach reduced the original set of 156 features (wavenumbers) to approximately 61 to 106 features in case of the range of wavenumbers from 1500 to 1800 cm⁻¹ and from 166 of original features (wavenumbers) to approximately 47 to 138 features in case of the range of wavenumbers from 2700 to 3000 cm⁻¹ without degrading or improving

| Dataset | Wavenumbers | # of cases | # of features | Classes |
|---------|-------------|------------|--------------|---------|
| 1       | 1500 to 1800 cm⁻¹ | 84         | 156          | Group I and II |
| 2       | 1500 to 1800 cm⁻¹ | 84         | 61           | Group I and II |
| 3       | 1500 to 1800 cm⁻¹ | 85         | 156          | Group II and III |
| 4       | 1500 to 1800 cm⁻¹ | 85         | 106          | Group II and III |
| 5       | 1500 to 1800 cm⁻¹ | 81         | 151          | Group I and III |
| 6       | 1500 to 1800 cm⁻¹ | 81         | 78           | Group I and III |
| 7       | 2700 to 3000 cm⁻¹ | 84         | 166          | Group I and II |
| 8       | 2700 to 3000 cm⁻¹ | 84         | 47           | Group I and II |
| 9       | 2700 to 3000 cm⁻¹ | 85         | 166          | Group II and III |
| 10      | 2700 to 3000 cm⁻¹ | 85         | 79           | Group II and III |
| 11      | 2700 to 3000 cm⁻¹ | 81         | 166          | Group I and III |
| 12      | 2700 to 3000 cm⁻¹ | 81         | 138          | Group I and III |
3. Results

3.1. Biochemical results

In this study, biochemical tests, as well as physical methods such as FTIR and Raman spectroscopy in combination with multivariate and machine learning analyses were used to determine differences in the antibody level in patients infected by COVID-19. Investigations were performed after 1, 3, and 6 months (group I, II, and III, respectively) from the date of infection (as confirmed by RT-PCR test).

The SARS-CoV-2 N antibody levels in the different periods after COVID-19 are shown in Table 3. The highest level of antibodies, in general, was recorded after 1 month from infection by COVID-19. This level decreases periodically by the time after the disease. However, only a small loss (for 27 units) is observed in the mean value of antibody level after 1 and 6 months.

To show a correlation between the level of antibodies and the time since the COVID-19 disease, a correlation test was performed (Table 4).

The correlation test presented in Table 4 showed the correlation between antibody levels after 1 and 3 as well as 6 months from COVID-19 disease. Furthermore, a correlation between patients’ antibody levels after 3 and 6 months from COVID-19 was observed. As can be seen from Fig. 1, the total protein value decreases with the time since infection by COVID-19. However, it was noted, that only differences between patients after 1 and 6 as well as between patients after 3 and 6 months from COVID-19, were statistically significant. In the case of groups of patients after 1 and 3 months from COVID-19, the differences were not significant.

Albumin levels in COVID-19 patients were also changing with time (Fig. 2). The increase of albumin level was observed after 3rd month vs 1st month. Statistically, a significant increase was in the 6th month after disease, compared to 1st and 3rd months. The increase in albumin level increased from 4.41 after 1 month to 4.47 after 6 months. Consequently, the difference was very small.

3.2. FTIR and Raman results with multivariate analyses

The FTIR spectra of serum collected from patients in different periods after COVID-19 are presented in Fig. 3a. The same peaks corresponding to functional groups building phospholipids, nucleic acids, carbohydrates, proteins and lipids vibrations were observed, and respective differences in the absorbance values were noticed. Peaks at 1082 cm⁻¹ and 1165 cm⁻¹ correspond to the vibration of PO₂⁻ and bending vibration of C=O—H groups from phospholipids, nucleic acids, and carbohydrates, respectively [32]. Amide vibrations which are structural units building proteins (amide III, amide II and amide I), were identified at 1238 cm⁻¹, 1304 cm⁻¹, 1531 cm⁻¹ and 1635 cm⁻¹ wavenumbers [32]. The vibration of C=O (Amide I, C=O stretching) was located at 1392 cm⁻¹, while the peaks at 1454 cm⁻¹ are originated from C—H stretching vibrations [33]. The peaks around 3278 cm⁻¹ are originated from O—H stretching vibrations. The vibration modes of peaks at 2958 cm⁻¹, 2931 cm⁻¹ and 2872 cm⁻¹ originate from the asymmetric vibration of C—H in CH₃, asymmetric vibration of C—H in CH₂ and symmetric vibration of C—H in CH₃, respectively [34]. These functional groups in the structure of the lipids can be observed. Furthermore, the C=O vibrations of lipids were also noticed at 1740 cm⁻¹ [34].

By comparing all three analyzed groups, it could be noticed that the lowest absorbance of all described functional groups in the FTIR spectrum of serum was noticed 6 months after COVID-19 (group III). The highest absorbance of C=O—H groups from carbohydrates and C=O, CH₂, CH₃ from lipids was noticed in patients 3 months after COVID-19 (group II) and slightly lower absorbance was observed in the spectrum for patients 1 month after COVID-19 (group I). However, the opposite
situation is observed for amides vibrations (the highest for group I and slightly lower for group II). In the group I, the absorbance level of peaks at 1392 cm$^{-1}$ and 1454 cm$^{-1}$ was almost the same, which was visible in the value of ratio 1392 cm$^{-1}$/1454 cm$^{-1}$ ($0.9963 \pm 0.0002$). While in the second group, the absorbance value of the peak at 1392 cm$^{-1}$ was lower than in the case of the peak at 1454 cm$^{-1}$ (ratio 1392 cm$^{-1}$/1454 cm$^{-1}$, $0.0909 \pm 0.0033$). In the group III, the absorbance of the peak at 1392 cm$^{-1}$ was higher than the peak at 1454 cm$^{-1}$ (ratio 1392 cm$^{-1}$/1454 cm$^{-1}$, $1.0157 \pm 0.0004$). Besides the absorbance intensity, differences in the shape of the spectrum were also detected.

PLS analysis presented in Fig. 3b showed, that significant differences in the absorbance between three analyzed groups were placed at

![Fig. 3. FTIR spectra of serum collected from patients 1, 3, and 6 months after COVID-19 (black, red, blue curves, respectively) (a); PLS analysis for IR range between 800 cm$^{-1}$ and 4000 cm$^{-1}$ (b); with diagnostic plots for groups I, II, III (c1, c2, c3, respectively).](image)

![Fig. 4. PCA analysis of IR range between 1500 cm$^{-1}$ and 1800 cm$^{-1}$ (a1) and between 2700 cm$^{-1}$ and 3000 cm$^{-1}$ (b1) of serum collected from patients after 1 (black dot), 3 (red dot) and 6 (blue dot) months from COVID-19 with PC1 (a2, b2) and PC2 (a3, b3) loading.](image)
wavenumbers between 1500 cm$^{-1}$ and 1800 cm$^{-1}$ as well as between 2700 cm$^{-1}$ and 3000 cm$^{-1}$. These regions correspond to the vibrations of amides (amide II and amide I) and lipids functional groups (C=O, CH$_2$ and CH$_3$). Furthermore, diagnostic plots obtained for groups showed, that experimental data fit well with the PLS model, Fig. 3c1, c2, c3. For groups II and III, fitting accuracy was around 75% in both cases, while in the case of group I, fitting accuracy was around 85%. Consequently, for these IR regions, PCA and HCA analyses were performed (Figs. 4 and 5).

PCA analysis of IR range between 1500 cm$^{-1}$ and 1800 cm$^{-1}$ (Fig. 4a1) showed a possibility to differentiate the patients being 6 months after COVID-19 from patients 1 and 3 months after this disease. The principal components PC1 and PC2 were calculated and illustrated in Fig. 4. In group III the values of PC2 were negative. As was shown in Fig. 4a3, PC2 with a negative value is originated from the IR range between 1620 cm$^{-1}$ and 1685 cm$^{-1}$. This range corresponded to amide I vibrations.

Also using IR range between 2700 cm$^{-1}$ and 3000 cm$^{-1}$ (Fig. 2b1) it was noted, that only samples from group III can be differentiated for the next two analyzed groups of patients. For patients 6 months after COVID-19, both components, PC1 and PC2, had a negative value. Loading plots presented in Fig. 4b2 and 4b3, showed, that component 2 plays the main role in the differentiation, in which negative values for wavenumbers between 2775 cm$^{-1}$ and 2830 cm$^{-1}$, as well as between 2860 cm$^{-1}$ and 2880 cm$^{-1}$, were noticed.

HCA analysis showed, that almost all patients after 1 month from COVID-19 created one similarity group, when vibrations of amide II and amide I are analysed (Fig. 5a). Furthermore, for these IR range, patients after 6 months from COVID-19 created two similarity groups, which were separated from each other. Moreover, these groups were also separated from the other two analyzed groups of patients (after 1 and 3 months). In the case of patients after 3 months from COVID-19, three similarity groups were noticed. However, what is important, these groups were more similar with patients from group I, than III. In the IR range between 2700 cm$^{-1}$ and 3000 cm$^{-1}$, HCA analysis didn’t show significant and well observable separation between three analyzed groups and similarity between samples collected from patients from I, II or III groups (Fig. 5b).
Fig. 7. PCA analysis of Raman range between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) (a1) and between 2840 cm\(^{-1}\) and 2956 cm\(^{-1}\) (b1) of serum collected from patients after 1 (black dot), 3 (red dot) and 6 (blue dot) months from COVID-19 with PC1 (a2, b2) and PC2 (a3, b3) loading.

Fig. 8. HCA analysis of Raman range between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) (a) and between 2840 cm\(^{-1}\) and 2956 cm\(^{-1}\) (b) of serum collected from patients after 1 (black dot), 3 (red dot) and 6 (blue dot) months from COVID-19.
In the Raman spectra of serum collected from patients after COVID-19 (Fig. 6), peaks attributed to proteins and amino acids vibrations were visible at 808 cm$^{-1}$, 874 cm$^{-1}$ and 1005 cm$^{-1}$ [35]. Raman shifts at 1360 cm$^{-1}$ and 1460 cm$^{-1}$ are shown to correspond to CH$_3$ and CH$_2$ vibrations from proteins and lipids, respectively [35]. Amide I vibrations are located around 1650 cm$^{-1}$ [35]. The highest values of Raman shifts were originated from the functional groups building lipid structures such as cholesterol and cholesterol ester: 1788 cm$^{-1}$, 2929 cm$^{-1}$ and 3053 cm$^{-1}$ [36].

According to Raman spectra (Fig. 6a), the highest intensities of amino acids vibrations were noticed for patients 1 month after COVID, in comparison with next two analyzed groups, in which the intensities in these Raman range were almost the same. The highest intensity of peaks originated from CH$_3$ and CH$_2$ vibrations from proteins and lipids were visible in spectrum of patients 3 months after COVID-19, while the smallest in the spectrum of serum collected from people 6 months after COVID-19. In the case of vibrations of functional groups building cholesterol and esters, similar intensity was observed in spectrum of I and II groups, while in the III group, the intensity was lower.

PLS analysis presented in Fig. 6b showed, that very wide range of
Raman shifts can be helpful in the differentiation of serum collected from patients after different time from COVID-19. Square plot showed, that ranges between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) and between 2840 cm\(^{-1}\) and 2956 cm\(^{-1}\), which correspond to proteins and lipids, could be helpful in distinguishing analyzed samples. Moreover, diagnostic plots showed, that fit with PLS model for groups I and II was around 80%, while for group III - 75% (Fig. 6 c1-c3).

PCA analysis of Raman spectra in the range between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) showed, that it is possible to distinguish three analyzed groups to each other (Fig. 7 a1). For samples from group I the values of PC1 and PC2 components were positive, while for group II the value of PC1 was negative and PC2 - positive. In the case of group III the value of component 2 was negative while PC1 had positive as well as negative values. Loading plots showed, that the negative value of PC2 was in the range between 1365 cm\(^{-1}\) and 1424 cm\(^{-1}\) and this range can be helpful in differentiation between group III and groups II and I (Fig. 5a2, a3).

Furthermore, in the Raman spectra range between 2840 cm\(^{-1}\) and 2956 cm\(^{-1}\), it is possible to distinguish three analyzed groups, because all samples for each group appeared in a different part of the coordinate system (Fig. 7b1). Almost all samples from group I had positive values of PC1 and PC2. In group II, values of PC1 and PC2 were negative, while in group III, PC1 was positive, while PC2 – negative.

HCA analysis of Raman spectra in the range between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) showed, that all samples collected from patients after 3
showed, that three separate similarity groups were created. It
3.3. Analysis of FTIR absorbance dynamics

Furthermore, HCA analysis in the spectral range corresponding to
Group I and II
61 selected
features
kNN (k-
3)
97.62 1 0.95 1
DNN 67.86 0.63 0.8 0.57
XGBoost 95.24 0.95 0.95 0.96
SVM 71.43 0.63 0.95 0.5

Group II and II
156 original
features
kNN (k-
3)
98.82 0.98 1 0.98 1
DNN 97.65 1 0.96 1
XGBoost 97.65 1 0.96 1
SVM 97.65 1 0.96 1

Group I and III
156 original
features
kNN (k-
3)
98.82 0.98 1 0.98 1
DNN 98.82 1 0.98 1
XGBoost 97.65 1 0.96 1
SVM 96.47 1 0.93 1

Group I and III
78 selected
features
kNN (k-
3)
100 1 1 1 1
DNN 100 1 1 1
XGBoost 100 1 1 1
SVM 100 1 1 1

3.4. Machine learning

The accuracy of classification depends on the range of wavenumbers and
the machine learning model used. It was in the range from 67.86% to 100.00% in case of the 1st range of wavenumbers and in the range
from 58.33% to 100.00% in case of the 2nd range of wavenumbers. Most
models achieved an accuracy of over 90%. Only for the DNN model the
accuracy was sometimes lower, reaching 58 or 67%. It is worth noting,
however, that in the case of analysis of the set of patients from groups I
and III, the achieved classification accuracy for almost all models was at
the level of 100%, i.e. the models unambiguously identified patients
from these groups. Similarly, additional parameters describing the
quality of classification reached the maximum value of 1.
In Tables 5 and 6 additional values of parameters characterizing the quality of classification are presented. The parameters used are: precision, sensitivity and specificity and the results obtained for the 2-class sets are very good. The value of the sensitivity parameter shows that almost all patients from one group in comparison to another one are correctly identified as having proper features. The value of the specificity parameter shows that almost all patients from another group are identified as not having proper features.

In addition, a decision tree diagram was constructed using the C5.0 algorithm on the set of patients from groups I and III for both ranges of FTIR wavenumbers which are shown in Fig. 11. From this simple diagram it can be seen that wavenumbers of 1700 and 2701 cm\(^{-1}\) perfectly distinguish between patients 1 and 6 months after COVID-19.

The results obtained from the supplied data, provide a basis for further extensive research on a large number of learning examples.

4. Discussion

After an illness, our body produces antibodies. The higher the level of these antibodies, the less likely you are to get sick once in a given disease [37]. Therefore, it’s extremely important to know the timelines with high antibody levels. Especially, nowadays, when COVID-19 pandemic spread in the world. Thanks to this knowledge, it may be possible to know, after what time since infected by COVID-19, the patients should e. g. have to be vaccinated to feel safe. Therefore, in this study, we did investigate the level of SARS-CoV-2 N antibody (Anti-N) in patients after 1, 3 and 6 months from COVID-19. Moreover, spectroscopy methods such as FTIR and Raman in combination with multivariate and machine learning methods were employed to distinguish the patients with different antibody levels.

Löfstrom et al. showed, that the level of n-antibodies significantly increased in patients from 1 to 3 months and significantly decreased from patients 3 to 6 months [38]. In this study, the highest decrease of antibody level after 6 months from COVID-19 is also noticed (see Table 3). Fortunately, in case of COVID-19 infection, a decline in antibody concentration after 3 months could be response for the risk of reinfection already after 9–12 months [41]. Furthermore, as showed in this work, the total protein and albumin levels also decrease with the extension of time after COVID-19 as presented in Figs. 1, 2, respectively. Several papers showed, that low serum levels of total protein, albumin, prealbumin is visible in patients after COVID disease [42,43]. Furthermore, FTIR and Raman spectra demonstrated the most visible differences in the intensity of peaks corresponding to amides vibrations, as seen in Fig. 3a and 6a, respectively. Amides are the most prevalent structures of living things, which build an important organic molecules and various biomolecules such as peptides, proteins and nucleic acids [44]. Consequently, PLS analyses of obtained IR and Raman spectra (Fig. 3b and Fig. 6b, respectively) showed, that vibrational regions of amides can be used in differentiation serum collected from patients after 1, 3 and 6 months from COVID-19. PCA and HCA analyses showed, that it is possible to distinguish three groups of patients using Raman spectroscopy in the range between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) (Fig. 7a1, Fig. 8a1), while in the case of FTIR spectra, using the range between 1500 cm\(^{-1}\) and 1800 cm\(^{-1}\) (Fig. 4a1), only patients after 6 months from COVID-19 can be differentiated from the patients after 1 and 3 months. Interestingly, the Raman range, which was noticed in this work, coincides with the position of the peak that was used to test the anti-SARS-CoV-2 IgM/IgG level using the SERS technique (1328 cm\(^{-1}\)) [22]. Moreover, Payne et al., showed, that the strongest peaks corresponding to spike proteins from COVID-19 were located in the Raman range, which was used to differentiate samples in this work [45]. Also, Carlomagno et al., showed, that peak in the Raman spectrum located at 1317 cm\(^{-1}\), which is originated from structural proteins of COVID-19 can be used to differentiate samples collected from people with positive and negative results of COVID-19 test [46]. This suggests, that in the Raman spectra, vibrations of protein functional groups play a very important role in the detection of COVID-19 or antibodies.

However, the analysis of dynamic of FTIR spectra (Fig. 9) showed that the exact distinction between patients after 1 and 3 months from COVID-19 can be obtained by investigating the regions between 1637 cm\(^{-1}\) and 1643 cm\(^{-1}\), 1714 cm\(^{-1}\) and 1724 cm\(^{-1}\), 1768 cm\(^{-1}\) and 1770 cm\(^{-1}\) with accuracy close to 100% (Table 5). Furthermore, FTIR and Raman spectra, as well as analysis of FTIR spectra dynamics, showed, that differences in antibody levels in patients at different periods after

---

**Fig. 11.** Diagram of a decision tree built from a set of patients from groups I and III for range of wavenumbers 1500–1800 cm\(^{-1}\) (a) and 2700–3000 cm\(^{-1}\) (b).
COVID-19 could be also differentiated in the lipids region of spectra as a function of time after recovery from COVID-19. It was shown that FTIR and Raman spectroscopy as new diagnostic methods in distinguishing patients after 1, 3, and 6 months from COVID-19, were investigated. Obtained FTIR and Raman spectra showed that in patients after 6 months from COVID-19, the level of amides II and amide I were lower than in patients after 1 month from COVID-19. Moreover, multivariate analysis showed, that using Raman spectroscopy in the range between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) and between 2840 cm\(^{-1}\) and 2956 cm\(^{-1}\) allows distinguishing patients after 1, 3 and 6 months from COVID-19. The absorbance dynamics method shows the existence of continuous change in the dynamics of FTIR spectra as a function of time after recovery from COVID-19. It was possible to indicate the ranges of wavenumbers in which such changes occur suggesting their possible use as markers that can examine the rate of recovery from the disease. Furthermore, the results obtained using 6 different machine learning techniques showed the possibility to almost perfectly differentiate the different stages of recovery. It was also possible to use selected wavenumbers as certain thresholds to distinguish patients between different levels of SARS-CoV-2 antibody.

Obtained results and performed analysis illustrate a high potential of FTIR and Raman spectroscopy as new diagnostic methods in distinguishing SARS-CoV-2 antibody levels at patients in different periods after COVID-19 infection.

5. Conclusions

In this study, the possibility of using FTIR and Raman spectroscopy to identify differences in the antibody levels in patients after 1, 3 and 6 months from COVID-19, were investigated. Obtained FTIR and Raman spectra showed that in patients after 6 months from COVID-19, the level of amides II and amide I were lower than in patients after 1 month from COVID-19. Moreover, multivariate analysis showed, that using Raman spectroscopy in the range between 1317 cm\(^{-1}\) and 1432 cm\(^{-1}\) and between 2840 cm\(^{-1}\) and 2956 cm\(^{-1}\) allows distinguishing patients after 1, 3 and 6 months from COVID-19. The absorbance dynamics method shows the existence of continuous change in the dynamics of FTIR spectra as a function of time after recovery from COVID-19. It was possible to indicate the ranges of wavenumbers in which such changes occur suggesting their possible use as markers that can examine the rate of recovery from the disease. Furthermore, the results obtained using 6 different machine learning techniques showed the possibility to almost perfectly differentiate the different stages of recovery. It was also possible to use selected wavenumbers as certain thresholds to distinguish patients between different levels of SARS-CoV-2 antibody.

Obtained results and performed analysis illustrate a high potential of FTIR and Raman spectroscopy as new diagnostic methods in distinguishing SARS-CoV-2 antibody levels at patients in different periods after COVID-19 infection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Ezhiyan, I. Suresh, N. Nesakumar, SARS-CoV-2, MERS-CoV and SARS-CoV-2: a diagnostic challenge, Measurement 168 (2021) 108335, https://doi.org/10.1016/j.measurement.2020.108335.

[2] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Measurement 167 (2021) 108288, https://doi.org/10.1016/j.measurement.2020.108288.
