Characterization of Covid-19 infected pregnant women sera using laboratory indexes, vibrational spectroscopy, and machine learning classifications

Zozan Guleken [a,1,*], Paweł Jakubczyk [b], Paja Wiesław [b], Pancerz Krzysztof [c], Huri Bulut [d], Esra Öten [e], Joanna Depciuch [f,1,**], Nevzat Tarhan [g]

[a] Department of Physiology, Uskudar University Faculty of Medicine, Istanbul, Turkey
[b] College of Natural Sciences, University of Rzeszów, Poland
[c] State School of Higher Education in Zamosc, Poland
[d] Department of Medical Biochemistry, Faculty of Medicine Istinye University, Istanbul, Turkey
[e] Health Science University Istanbul Kanuni Sultan Suleyman Training and Research Hospital, Department of Obstetrics and Gynecology, Turkey
[f] Institute of Nuclear Physics Polish Academy of Science, 31-342, Krakow, Poland
[g] Uskudar University NP Hospital, Istanbul, Turkey

ARTICLE INFO

ABSTRACT

Herein, we show differences in blood serum of asymptomatic and symptomatic pregnant women infected with COVID-19 and correlate them with laboratory indexes, ATR FTIR and multivariate machine learning methods. We collected the sera of COVID-19 diagnosed pregnant women, in the second trimester (n = 12), third-trimester (n = 7), and second-trimester with severe symptoms (n = 7) compared to the healthy pregnant (n = 11) women, which makes a total of 37 participants. To assign the accuracy of FTIR spectra regions where peak shifts occurred, the Random Forest algorithm, traditional C5.0 single decision tree algorithm and deep neural network approach were used. We verified the correspondence between the FTIR results and the laboratory indexes such as: the count of peripheral blood cells, biochemical parameters, and coagulation indicators of pregnant women. CH_2 scissoring, amide II, amide I vibrations could be used to differentiate the groups. The accuracy calculated by machine learning methods was higher than 90%. We also developed a method based on the dynamics of the absorbance spectra allowing to determine the differences between the spectra of healthy and COVID-19 patients. Laboratory indexes of biochemical parameters associated with COVID-19 validate changes in the total amount of proteins, albumin and lipase.

1. Introduction

The novel Covid disease 2019 (COVID-19) pandemic is still spreading around the world. Due to changes in physiologic status identified with the pregnancy and immunology related to respiratory basis, the population is more defenseless against coronavirus contamination [1]. Ongoing investigations have shown that ACE2 expression related to immaturity in the placenta [2], might explain the susceptibility of women to the disease, particularly in the first trimester of pregnancy [3]. Ellington et all reported that the prevalence of COVID-19 in pregnant women was 9.0% in 2020 [1]. Summarizing, recent clinical information suggests that pregnant women are more suspectable to coronavirus and have a higher risk of infection. Rapid and unambiguous diagnostics of coronavirus has particular importance in effective screening of the COVID-19 infected patients. A variety of molecular COVID-19 testing techniques ranging from advanced tests applied in research centers, to simple and rapid tests used in points-of-care is being worked on to effectively diagnose COVID-19 patients [4]. These techniques, well known to specialists and clinicians, are based on deoxyribonucleic acid (DNA) amplification, antibody and antigen assays, might be ambiguous for general community. Up to date, there are several accessible diagnostic tests for COVID-19, with the perspective to

---

* Corresponding author.
** # equal senior authors
    E-mail addresses: zozan.guleken@uskudar.edu.tr (Z. Guleken), joanna.depciuch@ifj.edu.pl (J. Depciuch).
[1] Equal senior author.

prevent a COVID-19 crisis. These tests are mainly dependent on the four following techniques: (1) reverse transcription-polymerase chain reaction (RT-PCR), which is the current standard test for coronavirus, (2) loop-mediated isothermal amplification (LAMP), which is a simple, but less developed testing method, (3) lateral flow – hand-held single-use assay and (4) enzyme-linked immunosorbent assay (ELISA) [5].

In the last decade, transmission or attenuated total reflection (ATR) Fourier transform infrared spectroscopy (FT-IR) has emerged as a new and rapid detection technique for viral diseases or to foresee viral consequences in blood [6], serum, plasma [7,8] or contaminated cells [9], separate different viral infections [6], or check the type of infecting agents, such as bacterial or viral, based on the spectral measurements of blood data [10,11]. Remarkably, studies had shown that molecular and chemical changes in blood components in response to bacterial or viral infections can be recorded and reflected by IR spectra [12,13]. These changes can be visible as differences in the absorbance value or peak positions. The FTIR spectrum of blood collected from patients infected by HIV is characterized by quantitative and qualitative changes in the spectral regions corresponding to lipids (3010 cm^{-1}), carbohydrates (1299 cm^{-1} and 1498 cm^{-1}), glucose (1035 cm^{-1}) and amide I (1652 cm^{-1}) vibrations [6], while hepatitis B and C virus can be diagnosed by FTIR spectroscopy using band at 1093 cm^{-1} [7]. Cells infected by herpes viruses show higher amount of phosphate, which is visible in FTIR spectra as an increase of the absorbance in the range between 950 cm^{-1} and 1350 cm^{-1} [8].

The main advantage of the IR technique, potentially valuable in identifying the presence of COVID-19 in human blood serum, is the possibility to characterize all biological components and chemical structure in the measured biofluid, as well as a rapid differentiation of the sample [14]. To reliably categorize the samples, various modelling and numerical methods are applied. Deep learning calculations, through the use of complex data sets are often regarded as the basis for separating important data and image them with timely interpretable structures [15].

Herein, we report a multi-level study of women at different pregnancy stages infected with COVID-19 using FTIR spectroscopy, laboratory indexes and modeling based on machine learning methods, as well as Lissajous curves. We attempt using FTIR spectroscopy to find a spectroscopy marker, which is present in sera of pregnant women suffering from COVID-19. Thus, we show, which absorbance maxima in FTIR spectra determine the course of the disease (symptomatic or asymptomatic). As medical data confirming the patient's health is provided, we correlate and validate the FTIR results with the laboratory indexes comprising the count of peripheral blood cells, biochemical parameters and coagulation indicators. Finally, the novelty is combining the absorbance spectra dynamics with the Lissajous curves to show, which infrared regions are the most sensitive to the changes in blood serum collected from the pregnant women suffering from COVID-19. Consequently, this allows to define, which chemical compounds in our body cause a symptomatic or asymptomatic course of COVID-19 in pregnant women. The performed calculations could be used as a basis to develop useful spectrochemical research models for diagnostics in practical applications.

2. Methods

2.1. Population study

This study consisted of a group of 37 pregnant women admitted at the maternity hospital of Kanuni Sultan Suleyman Training and Research Hospital in the clinical Department of Obstetrics and Gynecology in Turkey, between November 2020 and May 2021, which were divided into 2 groups: (A) 11 pregnant women diagnosed COVID-19 negative – control group and (B) 26 pregnant women diagnosed COVID-19 positive. Additionally, (B) was divided into 3 groups: (1) 12 pregnant women in second trimester with asymptomatic COVID-19

(T2); (2) 7 women in second trimester with severe COVID-19 symptoms (T2s) and (3) 7 pregnant women in third trimester with asymptomatic COVID-19 (T3). Assessments of clinical evaluations of critical vital signs, chest X-ray assessments and laboratory indexes were confirmed in the supervision of physicians. Therefore, treatment management has been prescribed based on clinical findings and national guidelines. At the time of enrollment, personalized data collections were used to document physiologic clinical chronic health conditions of pregnant women or neonates. Subsequently, maternal and neonatal outcomes and clinical progress were recorded until the birth of pregnant women. The accuracy of our data was verified via independent researchers.

2.2. Specimen collection

We performed the experiments with blood serum following the universal safeguard as directed by the Institutional Review Board of the Istanbul University, Cerrahpaşa Medical Faculty Clinical research etic committee (26.04.2021 date, and with the number of E−54368345-199-83879) in Turkey. The contributors were fully informed and signed the consent form for this study. COVID-19 was tested by nasopharyngeal swabs in the hospital according to guidelines. Blood samples were obtained from all participants via venipuncture. Approximately after 5 min after collection, the samples were moved to the physiology laboratory from obstetrics and gynecology units. After collecting the participant's whole blood sample in the blood tubes without adding any reagents, we coagulated the blood by leaving it at room temperature for about 20–30 min, and then removing the clot by centrifugation for 15 min at 3000 rpm to separate the fibrinogen precipitate. Serum samples were kept in Eppendorf. In order to prevent proteolytic degradation, we stored the samples at -80 °C until the spectroscopic analysis.

2.3. FTIR spectra collection

Mid-IR spectrometer (FT/IR-4700, JASCO, Tokyo/Japan) equipped with ATR TM diamond crystal plate, with liquid-nitrogen cooled MCT detector was used for spectra acquisition. The spectra were collected with 128 scans for each sample, at a resolution of 4 cm^{-1} with zero-filling of the interferogram resulting in 4 cm^{-1} data spacing corresponding to 3684 data points. To avoid extensive water absorbance, the spectrum of the empty diamond/MCT of the ATR unit was recorded as background and subtracted repeatedly from each measured sample. 1 μL of each serum sample was dropped onto the IR-reflective glass [16] and subsequently dried [15]. The acquisitions were made three times per sample. The crystal of the device was washed prior each sample measurement. Flowing the preprocess, all spectral recordings were analyzed via JASCO Spectra Manager version 2 and OPUS software. The features were baseline and ATR corrected, vector normalized, 25 points smoothed with Savitzky–Golay from 4000 to 600 cm^{-1} bands and the data were averaged [17]. We used 128 scanned serum data spectral band areas between 4000 cm^{-1} to 400 cm^{-1} to see the structural differentiation related to COVID-19 infection. After baseline correction, the best fit for decomposing the amide I band in the spectral region of interest was obtained.

2.4. Multivariate analysis

To assess the information about the spectra variation among the samples, Partial Least Squares analysis (PLS) on the FTIR data was performed using the Origin 2019 software. With this analysis, we aimed to show, which IR regions and wavenumbers play the most significant role in distinguishing between the control and COVID-19 samples.

2.5. Machine learning methods applied to COVID data

To acquire the knowledge about the accuracy of FTIR spectroscopy in

separating evaluated samples, three machine learning methods were used: the Random Forest (RF) algorithm [18], as well as the standard C5.0 single decision tree algorithm [19] and Deep Neural Networks (DNN) algorithms approved by literature [20]. To conduct the analysis, appropriate datasets were created to classify the cases and to distinguish between them. The datasets consisted of rows (i.e., patients), columns describing patients, (wavenumber or single peaks) and a decision column containing the category of the condition: (1) – Asymptomatic second trimester; (2) –Asymptomatic third trimester; (3) – Third trimester with severe symptoms; (4) – Control group.

The experiments were performed using the R environment, as well as the Random Forest, C5.0 and Keras software packages. The Boruta package [20] was used to perform the selection process for the most important attributes, which have the greatest impact on the assignment of the condition category to evaluate and calculate the importance of each descriptive attribute of peaks. Our approach reduced the original set of 453 attributes to approximately 3–223, depending on the number of categories created without degrading or improving the quality of case classification. In this way, the analysis was performed using the eight datasets created:

- 4 categories with 453 attributes - all peaks, data dimension: 37 rows, 453 columns and the category column containing one of the four disease categories;
- 4 categories with 13 attributes selected (confirmed and tentative) – characteristic peaks selected and identified in the data, data dimension: 37 rows, 13 description columns, the category column containing one of the four disease categories;
- 4 categories with 7 attributes selected (confirmed) – characteristic peaks selected and identified in the data, data dimension: 37 rows, 7 description columns, the category column containing one of the four disease categories
- 3 categories with 453 attributes – all peaks, data dimension: 37 rows, 453 columns, and the category column containing one of the three disease categories (categories 1 and 2 have been combined into one category);
- 3 categories with 3 attributes selected (confirmed) – characteristic peaks selected and identified in the data, data dimension: 37 rows, 3 description columns, the category column containing one the three-disease category (categories 1 and 2 have been combined into one category);
- 2 categories with 453 attributes – all peaks, data dimension: 37 rows, 453 columns and the category column containing one of the two disease categories (categories 1, 2, and 3 have been combined into one category);
- 2 categories with 223 attributes selected (confirmed and tentative) – selected peaks characteristic and identified in the data, data dimension: 37 rows, 223 description columns, the category column containing one the two-disease category (categories 1, 2 and 3 have been combined into one category);
- 2 categories with 115 attributes selected (confirmed) – selected peaks characteristic and identified in the data, data dimension: 37 rows, 115 description columns, the category column containing one the two-disease category (categories 1, 2 and 3 have been combined into one category);

All experiments were performed using leave-one-out cross-validation method. This procedure is used to estimate the performance of machine learning algorithms when they are used to make predictions on data not used to train the model. It is a computationally expensive procedure to perform, although it results in a reliable and unbiased estimate of the model performance. The model is evaluated for every held-out observation. The final result is then calculated by taking the mean of all the individual evaluations. It allows splitting the set of examples into 37 pairs of training (36/37 cases) and test (1/37 cases) sets containing mutually exclusive examples.

Table 1

Characteristics of the approved participants. Total pregnant women with COVID-19 infection (n = 27).

Maternal baseline characteristics	
Maternal age, mean (SD)	30.8 + 6.4
RT-PCR assay of a maternal nasopharyngeal swab	
Positive n (%)	50 (100)
Negative n (%)	Null
Pregnancy BMI, kg/m2	29.7 + 4.8
Positive chest X-ray (CT) n (%)	11 (42)
Severe case (CT) n (%)	3 (12)
Pharmacological treatment n (%)	
Azithromycin	6 (23)
Plaquenil	16 (61)
Clexane	26 (100)
Lopinavir + ritonavir (Kaletra)	3 (12)

Table 2

Maternal and pregnancy outcomes of the approved participants. Total population n = 26.

Delivery mode	
Vaginal, n (%)	54 (14)
Cesarean section, n (%)	46 (12)
Covid related cesarean n (%)	19 (5)
GA at delivery, weeks median (range)	38 (36–41)
Induction of delivery related to COVID-19	6 (23.07)
Infected neonates, positive, n (%)	1 (3.8)
Apgar_1 Score mean (SD)	
T2	7.16 (0.79)
T3	6.85 (1.64)
T2s	7.28 (0.45)

2.6. Statistics

The clinical outcomes are expressed as number of cases (n) and percent (%) and mean with standard deviation (SD). The statistics were made by one ways of ANOVA, followed by Bonferroni * compared to the T2 + compared to T3. FTIR analysis was performed using Past 3.0 software. Mann–Whitney *U* test was used as the nonparametric test with a p-value threshold of 0.05. The analyses were performed using SPSS Statistics, together with GraphPad Prism 6. Moreover, to show the correlation between FTIR data and laboratory index values, Pearson correlation test was performed using the Past 3.0 software. Taking into account, that statistically significant changes between the analyzed groups were present in the white blood cells (WBC), lymphocytes, triglycerides, D-dimer, CRP, ferritin, albumin and total protein levels, so in the entire lipids and proteins fraction, we correlate the values of these laboratory parameters with average values of IR region corresponding to proteins (1500 cm^{-1} - 1700 cm^{-1}) and lipids (2880 cm^{-1} - 2980 cm^{-1}).

3. Results

3.1. Population

Seven patients in the second trimester (T2s) were classified as severe cases. One of the pregnant women was taken to the emergency. We reported maternal characteristics and pharmacological treatment of the studied groups in Table 1. None of the newborns was tested positive for COVID-19 genome detection via a swab of the nasopharynx. A radiological chest X-ray confirmation of interstitial pneumonia was obtained on admission for all COVID-19 diagnosed pregnant women. Pharmacological treatment during the hospitalization is reported in Table 1 in the title of the baseline characteristics of the participants.

We followed all of the pregnant women until birth. We performed an Apgar score to evaluate whether the child is healthy or not. Neonates were healthy. Maternal and pregnancy outcomes are listed in Table 2.

Table 3
The laboratory indexes of peripheral blood cells.

|	T2 Mean (SD)	T3 Mean (SD)	T2s Mean (SD)	Reference (Unit)
Lymphocytes	1.54 ± 0.62	2.11 ± 0.955	1.14 ± 0.26	1.3–3.5 (10ˆ3 μL)
Neutrophils	5.08 ± 1.64	5.53 ± 2.42	5.53 ± 1.63	2.1–6.1 (10ˆ3 μL)
White blood cells (WBC)	6.42 ± 1890	8.27 ± 2457	7.06 ± 1611	4.3–10.3 (10ˆ3 μL)
Platelets counts	221.40 ± 5.80	245.80 ± 5.70	214.50 ± 5.40	156 - 373 (10ˆ3 μL)
Biochemical parameters	**T2 Mean (SD)**	**T3 Mean (SD)**	**T2s Mean (SD)**	
Amylase	65.80 ± 27.50	72.70 ± 23.80	112.00 ± 105.00	28 - 100 (U/L)
Lipase	33.90 ± 14.80	40.80 ± 22.40	47.10 ± 37.00	10 - 40 (U/L)
Triglyceride	185.00 ± 124.00	272.00 ± 140.00	315.00 ± 71.50	150 - 200 (mg/dL)
AST	20.20 ± 7.60	28.90 ± 8.59	25.00 ± 11.70	15 - 42 (IU/L)
ALT	14.20 ± 6.45	17.00 ± 4.97	18.90 ± 9.53	10 - 40 (U/L)
Coagulation indicators	**T2 Mean (SD)**	**T3 Mean (SD)**	**T2s Mean (SD)**	
CRP	13.80 ± 10.10	19.8 ± 30.60	59.80 ± 36.80	<3.00 (mg/L)
Ferritin	29.20 ± 26.00	58.4 ± 42.70	37.70 ± 9.47	20 - 200 (mg/ng)
D-dimer	1.54 ± 0.59	2.93 ± 1.94	37.70 ± 9.47	0,00–0.50 μg/mL
Pro-calcitonin	0.04 ± 0.02	0.06 ± 0.06	0.06 ± 0.04	0–0.05 μg/L
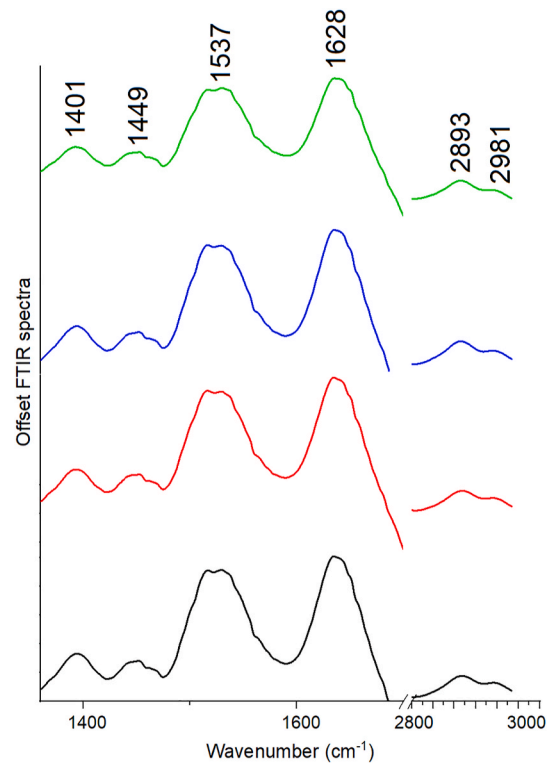
Fig. 2. Representative FTIR spectra of serum collected from COVID-19 infected pregnant women: T2 (black spectrum); T3 (red spectrum); T2s (blue spectrum) and without COVID-19 disease (green spectrum).

3.2. Laboratory indexes

It is known that we do not have the exact treatment protocol for the COVID-19 disease. The main monitoring parameters are respiratory rate, oxygen saturation and body temperature. Clinical follow-up of the count of lymphocytes, CRP, D-dimer and ferritin define the grade of the disease. The gradual increase in these markers is considered as cytokine storm or macrophage activation syndrome. Although we still do not have an exact scoring system, an increase in the level of CRP, D-dimer, ferritin, and triglyceride levels with a decrease in the count of thrombocytes and fibrinogens is defined as cytokine storm. On the other hand, it is known that the levels of neutrophil and procalcitonin are important markers for the consideration of the COVID-19 disease. The laboratory indexes including: peripheral blood cells, biochemical parameters and coagulation indicators of the COVID-19 infected pregnant women in the second trimester (T2), third trimester (T3), and severe second trimester (T2s) are presented in Table 3.

Peripheral blood counts were different among groups, as seen in Table 3. White blood cells were increased significantly in T2 vs T3 ($p < 0.01$). The counts of neutrophils were the highest for the T3 group and were slightly over the reference range. The number of lymphocytes was significantly decreased in the second-trimester pregnant women with severe COVID-19 infection, compared to T2 and T3 women ($p < 0.05$). Conversely, the platelets were in the reference range and there were no significant differences between the groups. The lipase and triglyceride levels of the T2s group were much higher than in the T2 and T3 pregnant women ($p < 0.05$). D-dimer levels were in all groups higher than the reference level. Statistically, there was an increase of the D-dimer in the T3 group compared with the T2 group ($p < 0.05$) as illustrated in Table 3. A significant increase in the level of CRP in the T2s group vs T2 and T3 ($p < 0.05$) was observed. Although, the ferritin levels were in the reference range, ferritin was higher in the T2s group compared to the T3 and T2s. The pro-calcitonin level was almost in the reference range, but
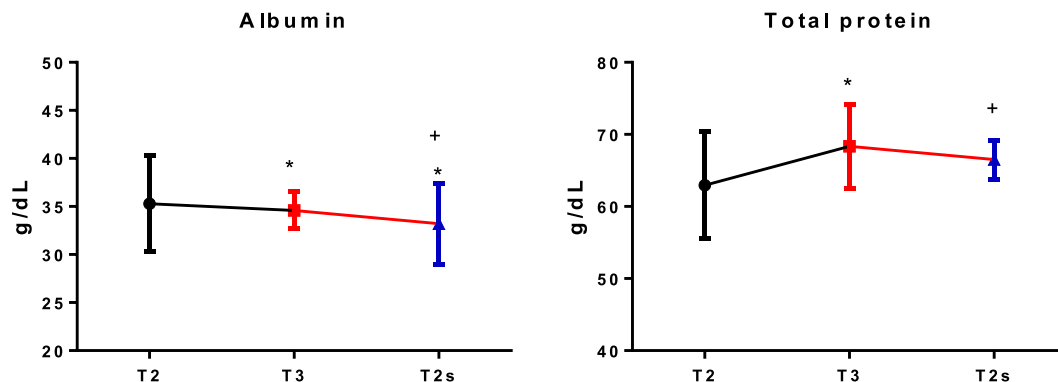
Fig. 1. Levels of albumin and total protein (g/dL). * Compared to the T2, + compared to T3. *$p < 0.05$, **$p < 0.01$, +$p < 0.05$, ++$p<0.01$ were considered significant.

Table 4
Description of vibrations presented in FTIR spectra of sera collected from women with and without COVID-19 with the marked shift of peaks.

Wavenumber (cm^{-1})							
Con.	T2	Δ = Con-T2	T3	Δ = Con-T3	T2s	Δ = Con-T2s	Vibrations
1401	1397	4	1397	4	1396	5	C=O stretching of COO$^-$
1449	1448	1	1448	1	1448	1	CH$_2$ scissoring
1537	1530	7	1531	6	1530	7	Amide II
1628	1631	−3	1632	−4	1633	−5	Amide I
2893	2887	6	2887	6	2888	5	Symmetric CH$_3$ stretching
2981	2975	6	2973	8	2977	4	Asymmetric CH$_3$ stretching

due to the measurement error its value was not statistically significant among the groups. All these changes may be aggravating factors for the course of COVID-19 disease.

To determine the differences in the amount of the protein level, we measured the albumin level and the total protein level in all groups (Fig. 1). In T2 pregnant women group the albumin levels were 36.67 (3.46), in the T3 group 34.23 (1.69) and T2s group 29.15 (1.8) with a mean (SD). Statistically, there was a significant decrease of the albumin level (Fig. 1a) in the T3 and T2s groups compared to the T2 group (p < 0.05) and a significant decrease in the T2s group compared to the T3 group. Additionally, we measured the total protein values in all the groups (Fig. 1b). The total protein value for the T2 group was 67.60 (4.52), for the T3 group, it was 66.75 (4.56) and for the T2s group 59.85 (7.2) with a mean (SD). Although albumin levels were increased in the T3 group, the total protein value was not decreased significantly (p < 0.01). But there was an important reduction in the protein level in the T2s group compared to T2 and T3 group (p < 0.01).

3.3. FTIR measurements

In this study, we used FTIR spectroscopy to obtain the information about chemical changes, which occurred in blood serum of pregnant women, who become infected with COVID-19.

In Fig. 2 visible peaks were marked in the collected FTIR spectra. These peaks correspond to functional groups building proteins and lipids structures. The peak at 1401 cm^{-1} corresponds to CH$_2$, as well as scissoring vibrations of CH$_2$ groups from carbohydrates and proteins were observed. Amide II and amide I vibrations were located at 1537 cm^{-1} and 1628 cm^{-1}, while vibrations of lipids functional groups were noticed at 1737 cm^{-1}, 2893 cm^{-1}, and 2981 cm^{-1}, respectively [21–26]. The description and the positions of the peaks visible in Fig. 1 from all analyzed groups, were assembled in Table 4.

When we compare the control group with groups of women suffering from COVID-19, structural changes in the biomolecules must have occurred, as peaks shift were observed. Indeed, the shift of peaks corresponding to stretching vibrations of C=O from COO−, amide I and amide II, as well as symmetric and asymmetric vibrations of CH$_3$ groups, were visible in all three groups of COVID-19 women.

In Table 4 differences in positions of peaks between COVID-19 groups, which passed COVID asymptomatically and very seriously are visible. This could suggest, that the course of COVID-19 depends on the protein fraction. Therefore, we decided to make a deconvolution of the amide I region, which provides information about the secondary structure of the protein fraction [27] (Fig. 3).

When we compare the amide I region obtained for COVID-19 (Fig. 3a) (Fig. 3b) (Fig. 3c) and control (Fig. 3d) groups, a different structure of the 1600 cm^{-1} – 1700 cm^{-1} range, was observed. Consequently, a different number of deconvolution curves was obtained. In COVID-19 groups, five curves were visible, while in the control group, a higher number of fitted curves was noticed. These curves originate from the secondary structure of proteins: α-helix and β-sheet [28]. Therefore,
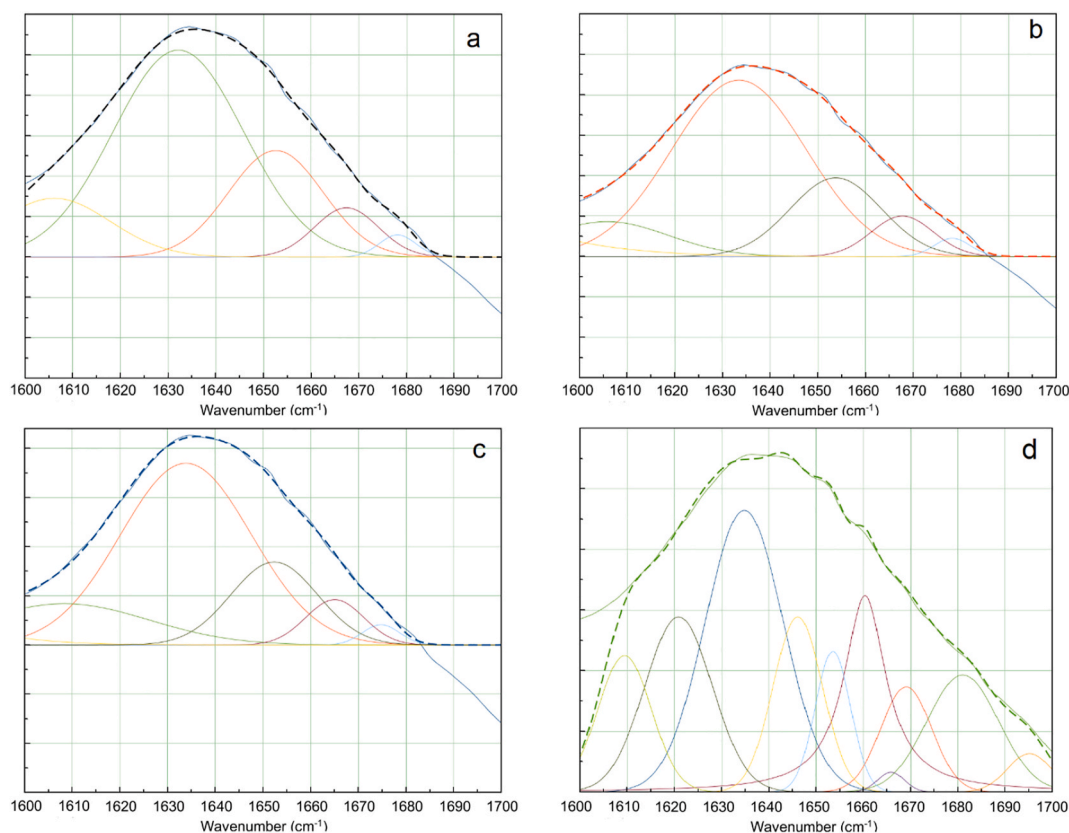

Fig. 3. Deconvolution of the amide I band of serum collected from COVID-19 infected pregnant women: T2 (a); T3 (b); T2s (c) and without COVID-19 disease (d).

Table 5

Peak positions of α-helix and β-sheet, as well as their area values obtained after deconvolution of amide I region of women with and without COVID-19.

Con.	Peak area	T2	Peak area	T3	Peak area	T2s	Peak area	Vibrations	α/β ratio
1609	1.598	1606	2.139	1606	1.393	1608	1.872	cross-β	Con. = 0.340
1621	2.584	1632	8.782	1633	7.717	1633	6.371	cross-β	T2 = 0.269
1646	1.886							α-helix	T3 = 0.230
1653	1.087	1652	3.284	1653	2.336	1652	1.926	α-helix	T2s = 0.211
1660	2.738	1667	1.023	1667	0.817	1665	0.704	anti-parallel β-sheet	
1681	1.837	1678	0.262	1678	0.210	1674	0.202	anti-parallel β-sheet	

The obtained results showed, that in COVID-19 groups, the value of α-helix and the β-sheet ratio is between 0.21 and 0.27, while in the control group – 0.34. This means, that COVID-19 could have caused changes in the protein fraction.
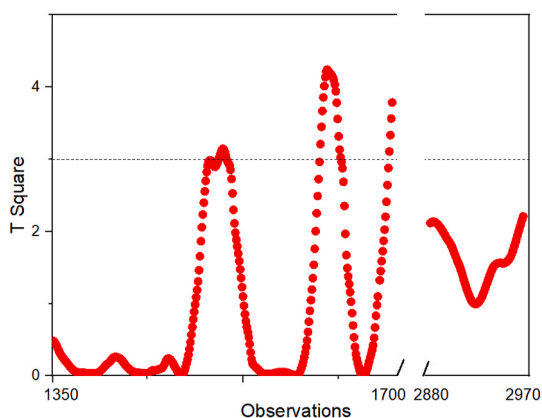
Fig. 4. PLS plot with the marked line separating the wavenumber values statistically significant in distinguishing between control and COVID samples.

the ratio between these two kinds of secondary structures was calculated and presented in Table 5.

3.4. Multivariate analysis

FTIR spectra and deconvolution of amide I region showed that deformation vibrations of CH groups may be used as a spectroscopic marker for the COVID-19 disease in pregnant women. Moreover, in the protein fraction, a marker responsible for the course of the disease was identified. However, to confirm these results and show, which IR region plays the most important role in distinguishing between the control and COVID-19 samples, Partial Least Squares analysis (PLS) was performed (Fig. 4).

The PLS results presented as plots showed, that in the FTIR spectra, the peaks corresponding to CH_2 scissoring vibrations and amide II, amide I vibrations could be used as a potential marker, allowing the separation of the COVID-19 samples from control.

### 3.5. Correlation between laboratory and FTIR results

The Pearson correlation test was performed to obtain the information about the correlation of clinical laboratory results with the FTIR data, i. e. correlation between lymphocytes, WBC, triglycerides, ferritin, D-dimer, albumin, total proteins concentrations and both: proteins and lipid absorbances from the FTIR spectra, Table 6.

The correlation test showed, that in all three analyzed groups of patients (T2, T3, T2s), a positive correlation between lipid vibrations and lymphocytes was noticed. Moreover, a correlation between proteins and lymphocytes, WBC, CRP, D-dimer and total protein amount was visible in T2 group. In these groups, correlation between lipids and triglycerides, ferritin, albumin and total protein amount was observed. In the T3 group, correlation between the proteins measured by FTIR and WBC, as well as total proteins was noticed, while lipids correlated with albumin in T3 group. In the T2s group correlation between proteins and ferritin, and also between lipids and CRP, as well as total protein amount

Table 6

Correlation between laboratory index and FTIR results.

	T2 group		T3 group		T2s group	
	Proteins FTIR range	Lipids FTIR range	Proteins FTIR range	Lipids FTIR range	Proteins FTIR range	Lipids FTIR range
Lymphocytes	0.64	**0.66**		0.90		0.93
WBC	0.66		0.90			
Triglycerides		0.75				
CRP	0.75					0.77
Ferritin		0.75			0.77	
D-dimer	0.75					
Albumin		0.66		0.90		
Total protein	0.64	0.66	0.90			0.93

Table 7

Classification results obtained by three machine learning methods for 8 datasets.

Datasets	Random forest		C5.0		DNN	
	Accuracy	Error	Accuracy	Error	Accuracy	Error
4 categories with all 453 attributes	64.86%	35.14%	78.38%	21.62%	62,16%	37,84%
4 categories with 13 attributes selected (confirmed + tentative)	75.68%	24.32%	75.68%	24.32%	62,16%	37,84%
4 categories with 7 attributes selected (confirmed)	78.38%	21.62%	75.68%	24.32%	62,16%	37,84%
3 categories with all 453 attributes	75.68%	24.32%	78.38%	21.62%	81,08%	18,92%
3 categories with 3 attributes selected (confirmed)	67.57%	32.43%	72.97%	27.03%	64,86%	35,14%
2 categories with all 453 attributes	97.30%	2.70%	97.30%	2.70%	100,00%	0,00%
2 categories with 223 attributes selected (confirmed + tentative)	97.30%	2.70%	97.30%	2.70%	100,00%	0,00%
2 categories with 115 attributes selected	97.30%	2.70%	97.30%	2.70%	100,00%	0,00%

Table 8
Classification results of sick and healthy patients (two-class dataset) with additional classification quality parameters.

	Datasets	Accuracy	Error	Selectivity	Specificity	F1	MCC
RF	2 categories with 453 attributes	97.30%	2.70%	96.30%	100%	98.11%	93.56%
C5.0	2 categories with 453 attributes	97.30%	2.70%	100%	91.67%	98.04%	93.88%%
DNN	2 categories with 453 attributes	100%	0%	100%	100%	100%	100%
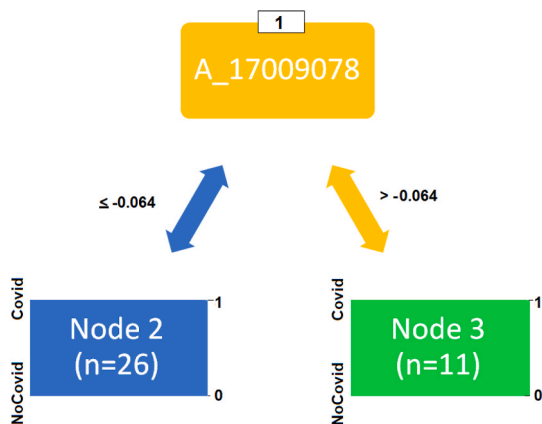
Fig. 5. Diagram of a decision tree built from a full two-class set.

was visible.

3.6. Machine learning methods applied to COVID data

The results obtained (Table 7) using Random Forest, C5.0, and Deep Neural Networks algorithms indicate that the proposed methods can effectively classify the studied groups.

The study was conducted on eight information databases containing cases grouped into 2, 3, or 4 categories. These were three databases containing all 453 attributes (wavelengths) describing each of the 37 patients. In addition, sets were used after performing significant attribute selection with Boruta algorithm. The restricted attribute space contained attributes that were either confirmed as significant or identified as tentative.

The accuracy of the classification was in the range from 64.86% to 97.30%. Correspondingly, the classification error is between 2.70% and 35.14%. The obtained results clearly show that the proposed method allows for 97.30% correct differentiation between sick and healthy patients. Only one of the 37 cases was diagnosed incorrectly.

In Table 8 of additional values of parameters characterizing the quality of classification are presented. The parameters used include sensitivity, specificity, F1 score, and Matthews correlation coefficient. The results obtained for the 2 class sets are very good. The value of the sensitivity parameter (96.30%–100%) shows that almost all sick people are correctly identified as having the condition. The specificity parameter value (91.67%–100%) shows that almost all healthy people are identified as not having the condition.

In addition, a decision tree diagram was constructed with the C5.0 algorithm on the full two-class set which is shown in Fig. 5. From this simple diagram it can be seen that the wavenumber at 1700.9078 perfectly distinguishes between sick (26 Covid cases) and healthy (11 No Covid cases) patient groups.

3.7. Analysis of FTIR absorbance dynamics

To find the differences between the IR spectrum of the COVID-19 patients and the spectrum of healthy persons, we developed a method based on the spectral absorbances dynamic as a function of the wavenumber. In this approach, we exploit the fact that in IR spectra, for

instance A, for a carefully selected range of wavenumbers k, two types of absorbance dynamics can be distinguished. First, we have $\frac{dA}{dk} > 0$, and for the second type, $\frac{dA}{dk} < 0$. In the case when $\frac{dA}{dk} > 0$ the absorption is increasing with k, while if $\frac{dA}{dk} < 0$ the absorption 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. Here the first IR spectrum, is the reference spectrum calculated as an average of the control group (ControlSpec) and the second one is the spectrum of the COVID-19 patient from the group – T2 + T3 + T2s (CovidSpec). We are looking for wavenumbers, for which the opposite dynamics of absorption in ControlSpec and CovidSpec spectra take place. As result, we obtain the set of wavenumbers, which indicates differences in absorption dynamics between ControlSpec and CovidSpec spectra. This set can be used as potential set of markers, which distinguish COVID-19 patients from the group of healthy individuals. In Table 8 we present the results for the CovidSpec group.

4. Discussion

We reported, that using of FTIR spectroscopy markers, which correspond to the way of passing COVID-19 by pregnant women could be identified. Additionally, in the literature we found information about successful using of FTIR spectroscopy to detect other type of viruses [8, 29,30]. Furthermore, to validate FTIR results, we show laboratory indexes associated with the count of peripheral blood cells, biochemical parameters, and coagulation indicators. Indeed, in our study, similarly to other works [31,32], the levels of D-dimer, CRP, ferritin and procalcitonin were higher in symptomatic T2s group (Table 3). Consequently, our medical data can be correlated with characteristic parameters in people suffering from symptomatic and asymptomatic COVID-19. Furthermore, in pregnant women with symptomatic COVID-19, the highest level of triglycerides was noticed (Table 3), which also agrees with the available literature [33]. Given that medical data coincide with COVID-19 findings from other studies, we are confident that we have carefully and representatively selected the investigated groups.

Fourier Transform InfraRed spectra of women, who had COVID-19, show a shift of peak originating from C=O vibrations from lipids, in comparison with the control group (Fig. 2). These vibrations were observed in carbohydrate fractions [22]. It is known, that one part of our immune system is correlated with carbohydrates. Furthermore, the surface of the COVID-19 is heavily glycosylated, with pre-existing antibodies to glycans. Therefore, antibody responses to carbohydrates could be induced, affecting disease severity and clinical outcome. Moreover, some studies showed, that pre-existing aluminum antibodies have the potential to recognize the virus and influence the progression of the disease [22]. Viruses, e.g., COVID-19 use the host's glycosylation machinery as a camouflage strategy, hiding hypothetical immunogenic epitopes, and also using host carbohydrate-binding receptors as entry mechanisms [23,24]. Moreover, it was found that antibodies with low affinity to some virus glycoproteins increase viral infection [25]. Our results showed, that carbohydrates could play a very important role in the infection of COVID-19. Structural changes, which were observed in FTIR spectra indicated, that these structures may be involved in our immune system's defense against the virus.

Our immune system produces antibodies, which have a structure very similar to proteins [34]. Therefore, to control the COVID-19

Table 9
Wavenumbers, which can be used as potential markers to discriminate COVID-19 patients by IR spectrum. The second column shows the corresponding discrimination probabilities.

Wavenumbers [cm^{-1}]	Probability of differentiation
1392	88%
1421	80%
1460	80%
1590	80%
2925	84%
2954	96%

infection, it is important to rapidly generate multiple high-affinity antibodies or antibody-like proteins (ALPs) against the virus proteins [35]. Moreover, activation of specific proteins was observed during the disease. Consequently, we noticed differences in the protein fraction and its structure in woman with COVID-19 (Fig. 3). Interestingly, the PLS plot (Fig. 4) showed, that the infrared region, which differentiates our samples, is placed in the range corresponding to amide II and amide I vibrations (proteins structure) and CH deformation vibrations. To investigate the accuracy of our results, three different machine learning and neural network, we done, Table 7. Importantly, we obtain accuracy values from 65% to 97% depending on the method.

We also tested new methods to determine infrared regions, which identify COVID-19 and its course. These methods were used for the first time for diagnostics of pediatric precursor B lymphoblastic leukemia [35], where the Authors showed, which blood parameter was correlated with structural changes visible form the spectra. In our study, we showed, that the phase shift is equal to π, consequently, it is significant, for the infrared region corresponding to proteins and lipids, Table 9. Consequently, we confirmed the data obtained from the PLS analysis, as well as machine learning and neural networks based on data from FTIR spectra. Importantly, using dynamics of the absorbance spectra and Lissajous curves we also showed, that amides and lipids levels obtained from laboratory are an important factor in the course of COVID-19.

5. Conclusion

In this study we report that the COVID-19 has effects on peripheral blood cells, biochemical parameters and coagulation indicators of both second trimester and third trimester pregnant women independently on the disease course, which is consistent with the literature. Furthermore, we report that the albumin level decreases both in third trimester and severe second trimester women with COVID-19. The obtained FTIR spectra showed significant differences between COVID-19 women with severe and light symptoms. In the first ones, shifts of peaks originating from asymmetric stretching vibrations of CH3 groups from lipids were noticed in comparison with results obtained for COVID-19 women with light symptoms. This could mean that these chemical fractions play a critical role in the course of COVID-19. Machine learning methods estimated the accuracy of FTIR results to be around 90%. The results of the absorbance spectra dynamics and Lissajous curves clearly showed, that 1392 cm^{-1}, 1421 cm^{-1}, 1460 cm^{-1}, 1590 cm^{-1}, 2925 cm^{-1} and 2954 cm^{-1} IR wavenumbers differed for COVID-19 and non-COVID-19 women.

Credit author statement

ZG, JD designed the study and planned the experiments, EÖ; conceived clinical diagnosis and collected blood samples, ZG and HB undertook the experiments, PJ, PK, PW undertook the machine learning analysis, ZG and JD undertook spectral analysis. ZG, JD wrote and edited the manuscript. All authors revised and agreed about the manuscript.

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] S. Ellington, P. Strid, V.T. Tong, K. Woodworth, R.R. Galang, L.D. Zambrano, J. Nahabedian, K. Anderson, S.M. Gilboa, Characteristics of women of reproductive age with laboratory-confirmed SARS-CoV-2 infection by pregnancy status — United States, Morb. Mortal. Wkly. Rep. 69 (2020) 769–775, https://doi.org/10.15585/mmwr.mm6925a1.
[2] K.G. Pringle, M.A. Tadros, R.J. Callister, E.R. Lumbers, The expression and localization of the human placental prorenin/renin- angiotensin system throughout pregnancy: roles in trophoblast invasion and angiogenesis? Placenta 32 (2011) 956–962, https://doi.org/10.1016/j.placenta.2011.09.020.
[3] P. Dashraath, J.L.J. Wong, M.X.K. Lim, L.M. Lim, S. Li, A. Biswas, M. Choolani, C. Mattar, L.L. Su, Coronavirus disease 2019 (COVID-19) pandemic and pregnancy, Am. J. Obstet. Gynecol. 222 (2020) 521–531, https://doi.org/10.1016/j.ajog.2020.03.021.
[4] L. Zhang, M. Xiao, Y. Wang, S. Peng, Y. Chen, D. Zhang, D. Zhang, Y. Guo, X. Wang, H. Luo, Q. Zhou, Y. Xu, Fast screening and primary diagnosis of COVID-19 by ATR-FT-IR, Anal. Chem. 93 (2021) 2191, https://doi.org/10.1021/acs.analchem.0c04049. –2199.
[5] B. Udugama, P. Kadhiresan, H.N. Kozlowski, A. Malekjahani, M. Osborne, V.Y. Li, C. Chen, H. Mubareka, S. Gubbay, W.C.W. Chan, Diagnosing COVID-19: the Disease and Tools for Detection, vol. 14, 2020, https://doi.org/10.1021/acsnano.0c02624, 3822–3835.
[6] M.C.D. Santos, Y.M. Nascimento, J.M.G. Araújo, K.M.G. Lima, ATR-FTIR spectroscopy coupled with multivariate analysis techniques for the identification of DENV-3 in different concentrations in blood and serum: a new approach, RSC Adv. 7 (2017) 25640–25649, https://doi.org/10.1039/c7ra03361c.
[7] L. Sitole, F. Steffens, T.P.J. Krüger, D. Meyer, Mid-ATR-FTIR spectroscopic profiling of HIV/AIDS sera for novel systems diagnostics in global health, OMICS A J. Integr. Biol. 18 (2014) 513–523, https://doi.org/10.1089/omi.2013.0157.
[8] S. Roy, D. Perez-Guaita, S. Bowden, P. Heraud, B.R. Wood, Spectroscopy goes viral: diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy, Clin. Spectrosc. 1 (2019) 100001, https://doi.org/10.1016/j.clispe.2020.100001.
[9] A. Salman, V. Erukhimovitch, M. Talyshinsky, M. Huleihil, M. Huleihel, FTIR spectroscopic method for detection of cells infected with herpes viruses, Biopolym. - Biospectroscopy Sect. 67 (2002) 406–412, https://doi.org/10.1002/bip.10171.
[10] A.H. Agbaria, G. Beck Rosen, I. Lapidot, D.H. Rich, M. Huleihel, S. Mordechai, A. Salman, J. Kapelushnik, Differential diagnosis of the etiologies of bacterial and viral infections using infrared microscopy of peripheral human blood samples and multivariate analysis, Anal. Chem. 90 (2018) 7888–7895, https://doi.org/10.1021/acs.analchem.8b00017.
[11] R. Eccles, Understanding the symptoms of the common cold and influenza, Lancet Infect. Dis. 5 (2005) 718–725, https://doi.org/10.1016/S1473-3099(05)70270-X.
[12] Z. Guleken, H. Bulut, G.İ. Gültekin, S. Arıkan, İ. Yaylım, M.T. Hakan, D. Sönmez, N. Tarhan, J. Depciuch, Assessment of structural protein expression by FTIR and biochemical assays as biomarkers of metabolites response in gastric and colon cancer, Talanta 231 (2021), https://doi.org/10.1016/j.talanta.2021.122353.
[13] Z. Guleken, B. Ünübol, R. Bilici, D. Sarıbal, S. Toraman, O. Gündüz, S. Erdem Kuruca, Investigation of the discrimination and characterization of blood serum structure in patients with opioid use disorder using IR spectroscopy and PCA-LDA analysis, J. Pharmaceut. Biomed. Anal. 190 (2020), https://doi.org/10.1016/j.jpba.2020.113553.
[14] F. Bonnier, F. Petitjean, M.J. Baker, H.J. Byrne, Improved protocols for vibrational spectroscopic analysis of body fluids, J. Biophot. 7 (2014) 167–179, https://doi.org/10.1002/jbio.201300130.
[15] Deep learning, the MIT press (n.d.), https://mitpress.mit.edu/books/deep-learning. (Accessed 4 May 2021). accessed.
[16] H.J. Byrne, F. Bonnier, J. McIntyre, D.R. Parachalil, Quantitative analysis of human blood serum using vibrational spectroscopy, Clin. Spectrosc. 2 (2020) 100004, https://doi.org/10.1016/j.clispe.2020.100004.
[17] J. Steinier, Y. Termonia, J. Deltour, Comments on smoothing and differentiation of data by simplified Least square procedure, Anal. Chem. 44 (11) (1972) 1906–1909, https://doi.org/10.1021/ac60319a045.
[18] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[19] W.R. Rudnicki, M. Wrzesien, W. Paja, All relevant feature selection methods and applications``, in: U. Sta´nczyk, L. Jain (Eds.), Feature´Selection for Data and Pattern Recognition, Studies in Computational Intelligence 584, Springer-Verlag, Germany, 2015, pp. 11–28.

[20] S.L. Salzberg, C4.5: programs for machine learning by J. Ross quinlan. Morgan kaufmann publishers, inc., 1993, Mach. Learn. 16 (1994) 235–240, https://doi.org/10.1007/bf00993309.

[21] J. Bujok, M. Gąsior-Głogowska, M. Marszałek, N. Trochanowska-Pauk, F. Zigo, A. Pavl̇ak, M. Komorowska, T. Walski, Applicability of FTIR-ATR method to measure carbonyls in blood plasma after physical and mental stress, BioMed Res. Int. 2019 (2019), https://doi.org/10.1155/2019/2181370.

[22] D.L. Butler, J.C. Gildersleeve, Abnormal antibodies to self-carbohydrates in SARS-CoV-2 infected patients, BioRxiv Prepr. Serv. Biol. (2020), https://doi.org/10.1101/2020.10.15.341479, 2020.10.15.341479.

[23] G. Simmons, J.D. Reeves, C.C. Grogan, L.H. Vandenberghe, F. Baribaud, J.C. Whitbeck, E. Burke, M.J. Buchmeier, E.J. Soilleux, J.L. Riley, R.W. Doms, P. Bates, S. Pöhlmann, DC-SIGN and DC-SIGNR bind Ebola glycoproteins and enhance infection of macrophages and endothelial cells, Virology 305 (2003) 115–123, https://doi.org/10.1006/viro.2002.1730.

[24] A.N. Lekkerkerker, Y van Kooyk, T.B. Geijtenbeek, Piractwo wirusowe. HIV-1 celuje w komórki dendrytyczne w celu przeniesienia, Curr. HIV Res. 4 (2006) 169–176.

[25] D.J. Vigerust, V.L. Shepherd, Virus glycosylation: role in virulence and immune interactions, Trends Microbiol. 15 (2007) 211–218, https://doi.org/10.1016/j.tim.2007.03.003.

[26] Y. Zhang, W. Zhao, Y. Mao, Y. Chen, S. Wang, Y. Zhong, T. Su, M. Gong, D. Du, X. Lu, J. Cheng, H. Yang, Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike proteins, Mol. Cell. Proteomics 20 (2021) 100058, https://doi.org/10.1074/mcp.ra120.002295.

[27] A.P. Fellows, M.T.L. Casford, P.B. Davies, Spectral analysis and deconvolution of the amide I band of proteins presenting with high-frequency noise and baseline shifts, Appl. Spectrosc. 74 (2020) 597–615, https://doi.org/10.1177/0003702819898536.

[28] N.K. Bhatia, A. Srivastava, N. Katyal, N. Jain, M.A.I. Khan, B. Kundu, S. Deep, Curcumin binds to the pre-fibrillar aggregates of Cu/Zn superoxide dismutase (SOD1) and alters its amyloidogenic pathway resulting in reduced cytotoxicity, Biochim. Biophys. Acta Protein Proteonomics 1854 (2015) 426–436, https://doi.org/10.1016/j.bbapap.2015.01.014.

[29] L. Hansen, T. De Beer, K. Pierre, S. Pastoret, A. Bonnegarde-Bernard, R. Daoussi, C. Vervaet, J.P. Remon, FTIR spectroscopy for the detection and evaluation of live attenuated viruses in freeze dried vaccine formulations, Biotechnol. Prog. 31 (2015) 1107–1118, https://doi.org/10.1002/btpr.2100.

[30] S. Gamage, M. Howard, H. Makita, B. Cross, G. Hastings, M. Luo, Y. Abate, Probing structural changes in single enveloped virus particles using nano-infrared spectroscopic imaging, PLoS One 13 (2018), e0199112, https://doi.org/10.1371/journal.pone.0199112.

[31] I. Huang, R. Pranata, M.A. Lim, A. Oehadian, B. Alisjahbana, C-reactive protein, procalcitonin, D-dimer, and ferritin in severe coronavirus disease-2019: a meta-analysis, Ther. Adv. Respir. Dis. 14 (2020), https://doi.org/10.1177/1753466620937175.

[32] H. Keski, Hematological and inflammatory parameters to predict the prognosis in COVID-19, Indian J. Hematol. Blood Transfus. (2021) 1, https://doi.org/10.1007/s12288-021-01407-y.

[33] L. Masana, E. Correig, D. Ibarretxe, E. Anoro, J.A. Arroyo, C. Jericó, C. Guerrero, M.L. Miret, S. Näf, A. Pardo, V. Perea, R. Pérez-Bernalte, N. Plana, R. Ramírez-Montesinos, M. Royuela, C. Soler, M. Urquizu-Padilla, A. Zamora, J. Pedro-Botet, C. Rodríguez-Borjabad, N. Andreychuk, A. Malo, L. Matas, M. del Señor Cortes-Fernandez, M. Mauri, R.M. Borrallo, À. Pedragosa, P. Gil-Lluís, A. Lacal-Martínez, P. Barragan-Galló, G. Vives-Masdeu, C. Arto-Fernández, O. El Boutrouki, A. Vázquez-Escobales, M.C. Antón-Alonso, S. Rivero-Santana, A. Gómez, S. García, N. Rial-Lorenzo, L. Ruiz-Ortega, O. Alonso-Gisbert, A.I. Méndez-Martínez, H. Iglesias-López, E. Climent, R. Güerri, J. Soldado, M. Fanlo, A. Taboada, L. Gutierrez, Low HDL and high triglycerides predict COVID-19 severity, Sci. Rep. 11 (2021) 7217, https://doi.org/10.1038/s41598-021-86747-5.

[34] N. Chaffey, B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, Molecular biology of the cell. 4th edn, Ann. Bot. 91 (2003) 401, https://doi.org/10.1093/aob/mcg023, 401.

[35] T. Kondo, Y. Iwatani, K. Matsuoka, T. Fujino, S. Umemoto, Y. Yokomaku, K. Ishizaki, S. Kito, T. Sezaki, G. Hayashi, H. Murakami, Antibody-like proteins that capture and neutralize SARS-CoV-2, Sci. Adv. 6 (2020), https://doi.org/10.1126/sciadv.abd3916.