ATR-FTIR Spectrum Analysis of Saliva Samples from COVID-19 Positive Patients

Adriana Martinez-Cuazitl
Escuela Militar de Medicina
Gustavo Vazquez-Zapien
Escuela Militar de Medicina
Miguel Sanchez-Brito
Instituto Tecnologico de Aguascalientes
Jorge Limon-Pacheco
Escuela Militar de Medicina
Melissa Guerrero-Ruiz
Escuela Militar de Medicina
Francisco Garibay-Gonzalez
Escuela Militar de Medicina
Raul Delgado-Macuil
CIBA-Tlaxcala del Instituto Politécnico Nacional
Maria Garcia de Jesus
Universidad Politecnica de Cuautitlan Izcalli
Miguel Corona-Perezgrovas
Hospital Central Militar
Armando Pereyra-Talamantes
Hospital Central Militar
Monica Maribel Mata-Miranda (mmcmaribel@gmail.com)
Escuela Militar de Medicina

Research Article

Keywords: COVID-19, SARS-CoV-2, FTIR Spectroscopy, ATR, Saliva

Posted Date: June 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-646875/v1

Abstract

The coronavirus disease 2019 (COVID-19) is the latest biological hazard for the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Numerous diagnostic tests for SARS-CoV-2 have been used, which are expensive and require specialized personal. So, new diagnosis strategies are being developed, looking for less expensive methods which could be used as screening for better spread control. Many researchers have described the use of saliva as a potential indicator of COVID-19, and even the same patient could carry out its collection. In this sense, this study aimed to establish specific salivary vibrational modes analyzed by attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectroscopy to detect COVID-19 biological fingerprints that allow the discrimination between COVID-19 and healthy patients.

Previous written informed consent, clinical dates, laboratories, and saliva samples of COVID-19 patients (n = 255) and healthy persons (n = 1209) were obtained and analyzed through ATR-FTIR spectroscopy. Then, a multivariate linear regression model (MLRM) was developed.

The COVID-19 patients showed low SaO_2, cough, dyspnea, headache, and fever principally. Obesity was the main comorbidity. Various laboratory blood tests were altered. In the FTIR spectra analysis, changes in amide I and immunoglobulin regions were evidenced, and the MLRM showed clear discrimination between both groups.

Specific salivary vibrational modes employing ATR-FTIR spectroscopy were established; moreover, the COVID-19 biological fingerprint in saliva was characterized, allowing the detection for COVID-19 using an MLRM, once it helps to reduce the number of variables, which could be helpful in the future development of diagnostic devices in a faster and cheaper way.

1 Introduction

The coronavirus disease 2019 (COVID-19) is the latest biological hazard provoked for the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is wreaking havoc in the health, political, economic, and social sectors around the world [1]. More than 180 countries are affected by this pandemic, and despite all containment efforts, the number of COVID-19 infected people is rising above 160 million, with over three million deaths accounting for the global fatality rate of ~ 2.07% [2, 3]. For COVID-19, the mean of incubation period and serial intervals are 5.2 and 7.5 days, respectively, while the basic reproduction number (R_0) is reported to be 2.2 persons on average [4].

Coronaviruses are single-stranded enveloped RNA viruses that cause diseases in mammals and birds. They are composed of the nucleocapsid (N) and spike (S) proteins, which participate in viral genome assembly, transcription, replication, or mediating viral entry provoking a cytopathic effect [5].

Some diagnostic tests for SARS-CoV-2 infection, such as the use of nucleic acids amplification test, antigen test, and serological antibody immunoassays, have been proposed and used. However, it is

essential to mention that even though the reverse transcriptase-polymerase chain reaction (RT-PCR) is considered the gold standard for detecting SARS-CoV-2, false-negative results can occur due to there is a window period of up to 5 days after exposure before viral nucleic acids can be detected. Moreover, the RT-PCR testing procedure requires high-level personnel training and a high level of laboratory expertise and usually takes 4 to 6 hours to complete the test, highlighting that this technique is expensive, which is not used for screening [3, 6]. In addition, it is essential to mention that the sample collection procedures for RT-PCR (nasopharyngeal and oropharyngeal swabs) require health personnel and close contact between health workers and patients, which generate a risk of virus transmission to health workers. In contrast, even the same patient could carry out saliva collection.

On the other hand, antigen-based diagnostic tests employing monoclonal antibodies against SARS-CoV-2 antigens, such as N protein and the S1 or S2 domains of the S protein, can also be used to detect viral infection. Nevertheless, these are less sensitive than RT-PCR-based tests, but an advantage is that they are available in rapid test kits (can provide results in 20–60 minutes) and also are easy to use. Finally, serological immunoassays or antibody tests use immobilized recombinant viral proteins/domains of viral proteins to which human antibodies bind if present in the blood, intended to identify persons with recent or prior SARS-CoV-2 infection. However, it is a low-cost alternative that can be widely used at the community level, and that can provide results in 15 to 20 minutes, it may take 21 days or longer after symptom onset for seroconversion or detection of antibodies to SARS-CoV-2, like immunoglobulin (Ig) M and/or IgG, which have been reported with an 88.66% of sensitivity and 90.63% of specificity in blood samples [3, 6].

So, new diagnosis strategies are being developed, looking for less expensive methods that could be used as screening for better spread control and greater sensitivity and specificity. Moreover, even though nasopharyngeal testing specimens remain the current standard for SARS-CoV-2 diagnosing, lower respiratory tract samples have a higher yield than upper tract samples, but they are often not obtained due to aerosolization of the virus during sample collection procedures. For this reason, some other tests suggest the use of saliva specimens [3].

Considering the above mentioned, in this research, we propose using the Fourier transform infrared (FTIR) spectroscopy technique, which analyzes the components of saliva samples emitting a "biological fingerprint," allowing the discrimination between positive and negative COVID-19 patients. FTIR is a technique based on the measurement of the absorption of electromagnetic radiation with wavelengths within the mid-infrared (IR) region (4000–400 cm^{-1}) that identifies variations in functional groups through the measure of the vibration and rotation of molecules influenced by IR [7]. Moreover, FTIR can effectively provide information concerning biological samples' structure and chemical composition at a molecular level and characterize lipids, proteins, carbohydrates, and nucleic acids. Besides, it is also sensitive to detect changes in molecular compositions according to a diseased state, providing fingerprints of biological samples, like tissues, cells, and biological fluids [8].

The reason for using saliva is due to the easy way of sampling, once the same patient can recollect it, without the complications (bruising, mucosal erosions, and bleeding) showed in nasopharyngeal and oropharyngeal swabs. It is essential to mention that saliva shows viral shedding of both the salivary glands and the upper and lower respiratory tract. In addition, previous studies have demonstrated that saliva has a high concordance rate with more than 90% with nasopharyngeal samples in the detection of respiratory viruses, including coronaviruses it has been reported that, in some patients, the coronavirus was detected only in the saliva, but not in nasopharyngeal aspirate samples [9]. In this sense, the use of saliva as a potential indicator of COVID-19 has been described by many researchers [10–15]. Considering the saliva advantages mentioned, we analyzed specific salivary vibrational modes that can be used to discriminate between positive and negative patients with COVID-19, which may prove that salivary spectral biomarkers are appropriate for detecting COVID-19.

However, the people characterization from the FTIR spectra analysis is a complex task due to the spectra differences, even from the same group. Moreover, saliva has a significant content of components, which show an intersection of its links, making it challenging to describe particular affectations of pathologies [16]. Currently, the problem of characterizing populations from FTIR spectra has been addressed using techniques commonly used in the area of machine learning; Santos et al. have suggested the use of different methodologies that allow distinguishing between two or more groups of spectra based on the results obtained through them in different investigations [17]. Since there is no way to know which method would be the most appropriate to tackle a particular problem, some authors have suggested using regression models followed by more complex models such as vector support machines (SVM) or artificial neural networks (ANN).

In this sense, the present study aimed to establish specific salivary vibrational modes analyzed by ATR-FTIR spectroscopy to detect COVID-19 biological fingerprints that are suitable for diagnosis or that can be used to identify distinct regions with fewer variables through a multivariate linear regression model (MLRM), allowing the discrimination between COVID-19 and healthy patients.

2 Material And Methods

2.1 Ethical aspects and study subjects

In this study, we discriminated between positive and negative COVID-19 patients through ATR-FTIR spectroscopy. For that purpose, the healthy group samples were used as a control, and it was integrated by 1209 healthy ambulatory volunteers who were recruited from February 2019 to February 2020 (a period in which the COVID-19 has not reached Mexico borders), 496 (41%) men, and 713 (59%) women with an average age of 60.5 ± 8.6 years. For positive COVID-19 samples, 255 hospitalized patients positive to SARS-CoV-2 diagnosticated through RT-PCR technique were recruited from May 2020 to March 2021, 160 (62.7%) men, and 95 (37.2%) women with an average age of 54.3 ± 14.7 years. Written informed consent for the obtention of 1 mL of saliva and participation in this study was obtained from the participants. The diagnosis of COVID-19 was developed in the Hospital Central Militar, Mexico.

The inclusion criteria were patients that accepted to participate in this study, aged over 18 years, and fasting at least 8 hours. The exclusion criteria were patients who had brushed or rinsed the oral cavity with mouthwash before sampling and patients with orthodontic treatment or other dental treatment. The healthy patients were informed that their samples would be used for different diagnostic assays as a reference control, and the COVID-19 patients were informed that their samples would be used to try other types of diagnosis.

The Clinical Research Ethics Committee of the Unidad de Especialidades Medicas and the Hospital Central Militar of the Secretaria de la Defensa Nacional approved the protocol and informed consent.

2.2 COVID-19 patients clinical data compilation

The 72.2% of the COVID-19 population were sampled from the first day when they showed symptoms to the third week of clinical evolution; most samples were collected at day 9.24 after RT-PCR diagnosis. At the sampling moment, vital signs were evaluated, symptoms like cough, dyspnoea, headache, fever, myalgia, arthralgia, among others, were interrogated, and comorbidities such as diabetes, obesity, hypertension, smoking, and other of importance in this disease. In the same way, the blood type was questioned. Finally, laboratory blood tests (hematic biometry, blood chemistry test, serum electrolytes, hepatic-function test, blood gas test, and others) were evaluated.

2.3 Sample pre-processing

For FTIR spectral analysis, we develop the methodology that has been carried out for the analysis of biological samples [18]. Therefore, this pre-processing was conducted in the spectral range between 4000 and 400 cm^{-1} (mid-infrared), using an FTIR spectrometer (6600, Jasco) in the attenuated total reflection (ATR) sampling mode. The instrument has a fixed spectral resolution of 4 cm^{-1}. Three µL of each sample was deposited onto the surface of the ATR crystal and dried at room temperature for about 15 minutes to eliminate excess water. The IR radiation propagated along the crystal to obtain the corresponding spectra that were the average of 120 data acquisitions. Each sample was analyzed three times, and all data were obtained in triplicate.

2.4 Spectral analysis

After normalizing through standard normal variate (SNV) and calculate the second derivative of the spectra, the analysis was performed in the biological fingerprint (1800 – 800 cm^{-1}). The second derivative was obtained using Unscrambler X. The mean of each population was obtained to identify relevant differences (absorbance differences and displacements). The graphs were obtained employing the Origin software (version 6.0, OriginLab Corporation).

2.5 Immune response and DNA and nucleic acids content through FTIR

Like other pathogens that cause infections, SARS-CoV-2 infection causes IgM, IgG, and IgA antibodies, making it mandatory to evaluate and compare the concentration of these antibodies between COVID-19

and healthy groups. For which purpose, the integrated areas were assessed in the spectral regions at 1420 – 1289 cm^{-1} and 1160 – 1028 cm^{-1} to evaluate IgM, 1560 – 1464 cm^{-1} which correspond to IgG, and finally the area 1285 – 1237 cm^{-1} corresponded to IgA [19]. Once the integrated areas were calculated, a Mann-Whitney test was developed to determine any significant differences between COVID-19 and the healthy group.

Moreover, the integrated areas of IgA, IgM, and IgG bands were compared between them in the COVID-19 group, employing a Kruskal-Wallis test.

The DNA and nucleic acid content were also compared between COVID-19 and healthy groups, analyzing the ratio (A966/991) for DNA content and the integrated area at 1237 cm^{-1} for nucleic acid content [20]. After that, a T-student test was developed to determine any significant differences.

2.7 Classification model

For the classification of our population, we carry out a multivariate linear regression model (MLRM), which is similar to the simple regression model except for having more coefficients for the variables considered (1) where *b* is the interceptor, m is the slope, and *x* an absorbance value:

$$Y_i = b + m_1 x_1 + \ldots + m_n x_n \ldots(1)$$

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. This analysis is called a multivariate or multiple linear regression model because more than one regressor is involved.

We used the leave-one-out (LOOCV) cross-validation methodology to evaluate the classification model because it has less bias than a validation set and produces the same results each time you run it [21]. The LOOCV methodology considers segmenting the database into two subsets: training and evaluation. The training subset comprises N-1 samples, and the evaluation subset considers the omitted spectrum in the training process. These partitions are performed as many times as spectra make up the database; in this way, all spectra contribute N-1 times to the regression model, and all spectra are evaluated once.

3 Results

3.1 Study population description

As previously mentioned, the healthy group was integrated by 1209 patients, 496 (41%) men, and 713 (59%) women with an average age of 60.5 ± 8.6 years. Furthermore, the COVID-19 group was integrated by 255 patients, 160 (62.7%) men and 95 (37.3%) women with an average age of 54.3 ± 14.7 years. Focusing on the COVID-19 group, the signs and symptoms, comorbidities, blood type, and laboratory blood tests were evaluated.

Table 1 shows the signs and symptoms that each patient presented and referred. The average of the obtained vital signs showed that the patients presented a low SaO_2; however, the rest of the vital signs were within normal parameters. Likewise, the ponderal state was determined through body mass index (BMI), evidencing overweight. The main symptoms that the patients mentioned were cough, dyspnoea, headache, and fever in more than 50% of the patients, followed by myalgias and arthralgias in more than 45% of the patients.

Table 1
Signs and symptoms

Vital Signs and Ponderal State	**median (IQR)**
SaO_2 (%)	90(83, 94)
Heart Rate (beats per minute)	87 (77, 102)
Respiratory Rate (breaths per minute)	22 (20, 25)
Systolic (mmHg)	120 (108, 131)
Diastolic (mmHg)	72 (65, 80)
Mean arterial pressure (mmHg)	94 (86, 103)
BMI (Kg/m^2)	28.3 (25.5, 35.1)
Symptoms	**n (%)**
Cough	194 (76.1)
Dyspnoea	189 (74.1)
Headache	145 (56.9)
Fever	145 (56.9)
Myalgia	125 (49)
Arthralgia	120 (47.1)
Sputum	67 (26.3)
Pharyngeal Pain	80 (31.4)
Emesis	30 (11.8)
Conjunctivitis	31 (12.2)
Asymptomatic	11 (4.3)

Table 2 resumes the comorbidities showed in the COVID-19 group, where it can be observed that 186 patients (72.9%) presented comorbidities, and some of them showed more than two comorbidities; even more, 2.4% of the patients presented more than four comorbidities. Obesity was the main comorbidity, the

38% of the COVID-19 patients showed this condition followed by diabetes and hypertension. Table 3 shows that the main blood type of the population that integrated the COVID-19 group was O+ (61.6%).

Table 2

Comorbidities

Comorbidities	n (%)
Obesity	97 (38)
Diabetes	84 (32.9)
Hypertension	70 (27.5)
Smoking	45 (17.6)
HIV	3 (1.2)
Cardiovascular disease	4 (1.6)
Others (cancer, Parkinson, infertility)	3 (1.2)
Asthma	2 (0.8)
Tuberculosis	2 (1.2)
None	69 (27.1)

Table 3

Blood Type

Blood Type	n (%)
O +	157 (61.6)
O -	2 (0.8)
A +	32 (12.5)
A -	2 (0.8)
B +	7 (2.7)
Unknown	55 (21.6)

The laboratory blood tests findings are described in Table 4, where it can be observed that the median values of neutrophils, N/L ratio, glucose, blood pH, C-reactive protein (CRP), lactate dehydrogenase (LDH), fibrinogen, D-dimer, and ferritin were increased; contrary PaO_2 and $PaCO_2$ were diminished.

Table 4
Laboratory blood tests

Hematic Biometry	**n**	**Median (IQR)**
Leukocytes (10^3/µl)	252	8.4 (6.3, 11.8)
Neutrophils (10^3/µl)	238	9.9 (5.6, 6.7)
Lymphocytes (10^3/µl)	249	1.2 (0.75, 3.7)
N/L Ratio	243	8.1 (4.4, 14.7)
Hemoglobin (g/dl)	251	15.3 (14, 16.4)
Platelets (10^3/µl)	251	238 (168.8, 318.5)
Blood Chemistry Test		
Creatinine (mg/dL)	250	0.8 (0.6, 0.96)
Urea (mg/dL)	240	34.6 (27.0, 49.2)
Glucose (mg/dL)	241	126 (103.3, 189.8)
Serum Electrolytes		
Na (mmol/L)	240	139 (137, 141)
K (mmol/L)	240	4.4 (4, 5)
Mg (mmol/L)	174	2.1 (2.0, 2.4)
Hepatic-Function Test		
ALT (U/L)	244	38 (25, 64)
AST (U/L)	243	39 (29, 55)
ALP (U/L)	146	91 (69, 124)
Bilirubin, Total (mg/dL)	189	0.8 (0.6, 1.1)
Albumin (g/dL)	239	3.3 (3, 3.7)
Blood Gas Test		
pH	150	7.46 (7.43, 7.50)
PaO_2 (mmHg)	149	66 (54, 77)
$PaCO_2$ (mmHg)	148	31.1 (28, 36)
HCO^{3-} (mmol/L)	149	22.7 (20.6, 24.7)
Lactate (mmol/L)	149	1.4 (0.9, 1.9)

Hematic Biometry	n	Median (IQR)
Other		
CRP (mg/dL)	228	90 (43.2, 202.4)
LDH (UI/L)	144	397 (266.3, 595.8)
Fibrinogen (mg/dL)	121	693 (467.3, 825.8)
D-dimer (ng/mL)	140	777 (438, 1675)
Ferritin (ng/mL)	143	489.5 (287.8, 954)

3.2 FTIR analysis

Figure 1A shows saliva FTIR spectra of COVID-19 and healthy groups, where diverse absorption bands related to different biomolecules are evidenced, such as lipids, proteins, carbohydrates, and nucleic acids, usually present in biological samples. The average saliva spectrum of both groups showed characteristics of biological samples, peaks of proteins at 1644 cm^{-1} (Amide I, C = O stretching), 1545 cm^{-1} (Amide II, N-H bending), and 1240 cm^{-1} (Amide III) were evidenced. Besides, glycosylated molecules (1200 – 1000 cm-^{1}), P = O asymmetrical and symmetrical stretching vibrations of PO_2 phosphodiester groups from phosphorylated molecules (1240 cm^{-1} and 1076 cm^{-1}) were also observed, as well as C-O stretching vibration coupled with C-O bending of the C-OH groups of carbohydrates (including glucose, fructose, and glycogen) at 1030 cm^{-1}. Moreover, a band in the interval 992 – 982 cm^{-1} associated with the ribose phosphate main chain and a band at 968 cm^{-1} from the DNA backbone stretching vibration are also demonstrated. Finally, it is essential to mention that the sugar moieties of glycosylated proteins, including α-amylase, were shown in the spectral range of 1080 – 950 cm^{-1}.

Additionally, the intervals between 1560 – 1464 cm^{-1}, 1420 – 1289 cm^{-1} and 1160 – 1028 cm^{-1}, and 1285 – 1237 cm^{-1} related to IgG, IgM, and IgA, respectively, are shown in Fig. 1B. These regions related to immunoglobulins showed an absorbance increment in the COVID-19 group and the region associated with glycosylated molecules (1200 – 1000 cm^{-1}) (Fig. 1).

When comparing the spectra between healthy and COVID-19 groups in the amide I and amide II regions (1700–1500 cm^{-1}), a slight displacement, as well as a decrease in the absorbance in the band of amide I at 1644 cm^{-1} (C = O stretching vibration) in the COVID-19 group was shown (Fig. 2A).

Figure 2B depicts the region of nucleic acids (1100–850 cm^{-1}), the band at 1076 cm^{-1} is associated with symmetrical stretching vibrations of PO_2 phosphodiester groups, and the interval between 1080 – 1030 cm^{-1} which is associated with glycogen, evidenced a higher absorbance in the COVID-19 group.

Furthermore, to analyze biomolecular changes with greater precision, we compared the spectra in the second derivative. Figure 3A shows the second derivative of the FTIR spectra of the COVID-19 and healthy groups depicted in the amide I region (1700–1600 cm^{-1}), where the bands related to the components of the secondary structure of proteins are shown, such as β-turns in the interval between 1695 – 1660 cm^{-1}, α-helices (1660 – 1650 cm^{-1}), and intramolecular native β-sheets (1640 – 1630 cm^{-1}), which are sensitive to structural and conformational changes [22]. Notably, an increment in the intensity of these bands in the healthy group compared with the COVID-19 group is observed.

In the same way, Fig. 3B shows the second derivative of the region of the nucleic acids (1100–850 cm^{-1}), where a significant increasement in the intensity of the band corresponding to symmetrical stretching vibrations of PO_2 phosphodiester groups at 1076 cm^{-1}, and other bands at 1037 cm^{-1} and 1028 cm^{-1} associated with glycogen on COVID-19 group compared to the healthy group is evidenced. Likewise, we noted that the bands at 992 cm^{-1} (ribose phosphate main chain) and 986 cm^{-1} (stretching vibration C-C of DNA backbone) changed their conformation in the COVID-19 group; on the other hand, the bands at 968 cm^{-1} and 924 cm^{-1} associated with the stretching vibration C-C of DNA backbone and the vibration of ribose ring increased on COVID-19 group.

3.3. Comparison of immunoglobulins, DNA, and nucleic acids content

Figure 4 shows the content of IgA (1285 – 1237 cm^{-1}), IgM (1420 – 1289 cm^{-1}, 1160 – 1028 cm^{-1}), and IgG (1560 – 1464 cm^{-1}) in healthy and COVID-19 groups. The IgA was slightly higher expressed in the COVID-19 group (40.95) than the healthy group (40.42), not showing statistical significance. Likewise, IgM (134.10 and 144.03) in the two studied regions and IgG (174.80) were more expressed in the COVID-19 group compared to the healthy group, IgM (131.69 and 130.24) and IgG (169.14), showing statistical significance.

When comparing the expression of the immunoglobins in the COVID-19 group, it can be observed that the IgA was the lowest expressed immunoglobulin, and the most expressed in this research group was de IgG, showing a statistical significance between each immunoglobulin (Fig. 5).

Figure 6 depicts nucleic acids and DNA content, where it can be observed that the DNA content was similar in both groups (0.66). In the same way, the nucleic acid content did not show a difference between the COVID-19 group (33.30) and the healthy group (33.00).

3.5. Multivariate analysis

Figure 7 shows the score plots obtained employing the MLRM with the spectra of healthy and COVID-19 groups. This MLRM was done in four spectral regions: 1700 – 1600 cm^{-1} (amide I of proteins), and the regions related to IgG (1560 – 1464 cm^{-1}) and IgM (1420 – 1289 cm^{-1}, 1160 – 1028 cm^{-1}).

Concerning the MLRM in these four regions (Fig. 7), the sensibility, specificity, and exactitude are shown in Table 5.

Table 5
Sensibility, specify, and exactitude of the multiple linear regression model (MLRM).

Region	Interval (cm^{-1})	Sensibility (%)	Specificity (%)	Exactitude (%)
Amide I of proteins	1700 – 1600	99.2	100	99.6
IgG	1560 – 1464	99.2	100	99.6
IgM	1420 – 1289	99	99.9	99.5
	1160 – 1028	77.1	93.2	87

Equally or even more important than the values presented in Table 5 is the analysis of the behavior of the outputs of the classification model (MLRM). In Table 6, we provide a brief analysis of these values.

Table 6
MLRM output analysis.

Region	Interval (cm^{-1})	HEALTHY		COVID-19	
		Output range	S.D.*	Output range	S.D.*
Amide I of proteins	1700 – 1600	0.62 to 1.39	0.06	0.95 to 2.15	0.09
IgG	1560 – 1464	0.44 to 1.46	0.07	0.94 to 2.28	0.10
IgM	1420 – 1289	0.49 to 1.53	0.12	0.79 to 2.37	0.14
	1160 – 1028	0.25 to 1.86	0.21	0.81 to 3.63	0.32
*Standard deviation					

Considering Table 6 and Figs. 7A and 7B, we determined that it is possible to discriminate both populations by analyzing the region associated with amide I. However, it was also possible to misidentify six COVID-19 patients.

4 Discussion

Considering the previous knowledge that virus infections provoke changes in the structures of biomolecules, in this research, we examined the FTIR spectra of COVID-19 and healthy patients, seeking the discrimination between these two populations through the analysis of FTIR spectra and an MLRM. Although ATR-FTIR is not used as a diagnosis technique, several authors have reported the use of FTIR for virus detection; for example, Erukhimovitch et al. in 2005, stated that it is possible to apply FTIR microscopy as a sensitive and effective assay for the detection of cells infected with various members of the herpes family of viruses and retroviruses [23]. Lee-Montiel et al., in 2011, evaluated the utility of FTIR

spectroscopy for rapid detection of infective virus particles poliovirus in cell cultures [24], and Santos et al. in 2020, reported several spectral features changes for hepatitis infected patients [17]. Therefore, in the search to propose new techniques that allow detecting the SARS-CoV-2 virus, FTIR spectroscopy has been considered in this research.

One of the main reasons for new diagnosis strategies was that although RT-PCR tests have shown high sensitivity for detecting the virus (97.2% [90.3%-99.7%]), false-negative results are expected within seven days of infection. Additionally, the RT-PCR process is time-consuming, and shortages in test kit supplies are common worldwide. On the other hand, serological testing of IgM and IgG production in response to viral infection is usually detected only 1–3 weeks after symptoms [25]. However, the employed technique used in this research allowed us to discriminate patients within the first day of clinical symptomology up to the third week of clinical evolution.

In the analysis of the COVID-19 population characteristics, although Peckham et al. have demonstrated that there is no difference in the proportion of males and females infected with SARS-CoV-2 [26], in this research, we documented that 160 (62.7%) men integrated the COVID-19 population, and 95 (37.3%) women, this probably due to the samples were obtained from hospitalized patients. The same authors declared that males face higher odds of intensive therapy unit (ITU) admission and death than females.

About the age, even though Hu et al. reported that it appears that all ages of the population are susceptible to SARS-CoV-2 infection, the median age of infection is around 50 years [27], which was also observed in this research, once the average age was 54.3 ± 14.7 years.

Concerning vital signs in the COVID-19 group, as previously mentioned, the only altered vital sign was the SaO_2 is showing a mean of 90%. Nevertheless, it is mandatory to remember that these patients were hospitalized; one of the main criteria for hospitalization besides evidence of pulmonary affection through CT, the low PO_2, which entails a low SaO_2. Hu et al. have reported that the most common symptoms in COVID-19 patients are fever, dry cough, and fatigue in patients less than 50 years, adding dyspnea in patients over 60 years (27). Likewise, in this research, we found that the main reported symptoms were cough, dyspnea, headache, and fever.

About comorbidities, as previously mentioned, obesity, diabetes, and hypertension were the most reported entities in this study. These results agree with Ortiz-Brizuela et al., Berumen et al., and Petrova et al., who declared that the pathologies above are the main risk of infection and hospitalization by COVID-19 [28–30].

Regarding the blood group, even though Zhao et al. have reported that blood group O is associated with a lower risk for the infection compared with non-O blood groups [31], in this research, the main blood type was O, probability due to this blood type is the most common in Mexico [32], country where this research took place.

About the laboratory blood tests, Velavan and Meyer have declared that CRP, D-dimers, ferritin, cardiac troponin, and IL-6 could be used in risk stratification to predict severe and fatal COVID-19 in hospitalized patients [33]. In our study, we observed that the values of neutrophiles, glucose, CRP, LDH, fibrinogen, D-dimer, and ferritin were increased, i.e., the patients that integrated our study presented three of the laboratory risks mentioned by Velavan, probably due to these patients were hospitalized because they required specialized medical attention. As expected, we detected neutrophilia, as it is known the primary function of the neutrophils is clearance of pathogens and debris through phagocytosis, the liberation of neutrophil extracellular traps is needed for viral infection inactivation and restriction of virus replication, been the neutrophils the first cell recruitment in COVID-19 [34]. In addition, hypoxia and hypocapnia are seen in severe COVID-19 cases; Wang et al. reported a median PaO2 of 68 mmHg and a median of PaCO2 of 34 mmHg in 138 COVID-19 patients [35], results that are similar to the ones obtained in this research (PaO2 66 mmHg, and PaCO2 31.1 mmHg).

On the other hand, the obtained spectra were similar to those reported by Caetano et al., showing characteristics of biological samples [16]. However, it is essential to mention that the population evaluated by Caetano et al. was informed to abstain from food and caffeine products for at least two h before the saliva collection and rinse out their mouths with distilled water. Contrary, in this study, a fasting period of at least 8 hours was required, and an exclusion criterion was patients who had brushed or rinsed the oral cavity with mouthwash before sampling.

As previously mentioned, in the FTIR spectra analysis, a slight displacement, as well as a decrease in the absorbance in the regions of amide I and amide II, were exhibited in the COVID-19 group, which may be attributed to a decrease in protein production, which corresponds to that reported by Denisa Bojkova et al., who observed a decrease in the expression of proteins, especially those related to cholesterol metabolism in CaCo-2 cells infected by SARS-CoV-2 [36]. In the same way, Bouhaddou et al. reported a decrease in the abundance of host proteins and a predominance of viral proteins, which is consistent with the mechanisms reported by other viruses in the inhibition of protein translation of the guest [37]; similarly to that found in Vero cells infected by herpes viruses, in which, protein synthesis and cellular metabolism decrease in the initial stages of infection consuming cellular metabolites such as nucleotides, amino acids, and cellular enzymes [36–38]. Highlighting that Barauna et al. reported a decrease in the peak related to amide I in saliva combined with inactive SARS-CoV-2 virus compared to saliva without infection [39].

In the same way, it is important to mention that the peak at 1240 cm^{-1}, which is related to phosphorylated molecules, is increased in COVID-19 patients, respect to healthy patients. About this, Bouhaddou et al. reported an increase in phosphorylated proteins with a decrease in protein abundance, as well as hyperphosphorylation of the CK2 and p38 MAPK pathways related to cytokine production [37], which is also consistent with that reported by Diamond et al. [40]. Moreover, Erukhimovitch et al. reported an increase in the peak at 1240 cm-1 in cells infected with the herpes virus [38].

About the immune response, it has been declared that the combination of IgG and IgM achieves an overall sensitivity of 87.8% and specificity of 98.9% for detecting SARS-CoV-2; nevertheless, the complexity of the humoral response in COVID-19 is not fully elucidated, and the relevance of the SARS-CoV-2 antibody response for the long-term clinical outcome of viral clearance is still lacking. Some authors have declared that the reported time to IgM positivity ranges from 5 to 10 days following disease onset, whereas IgG positivity occurs between 13–21 days. Moreover, some others have stated that the earliest detection of IgM was at five days post symptom onset, and the earliest detection of IgG was at seven days post symptom onset [41–43].

In the same way, it has been reported that IgA plays an essential role in mucosal immunity, being the most crucial immunoglobulin to fight infectious pathogen in the respiratory system [44]. Furthermore, it has been stated that salivary testing is the most convenient way to measure IgA, the reason by which it has been used to characterize mucosal immune responses to many viral infections such as SARS, MERS, influenza, HIV, and RSV. Serum IgA has been detected in COVID-19 patients and appears to be detectable earlier than IgM or IgG antibodies, possibly as early as two days after onset of symptoms, suggesting that IgA may be the first antibody to appear in response to SARS-CoV-2 infection [45]. In this research, changes in absorbance in the areas related to IgG (1560 – 1464 cm^{-1}), IgM (1420 – 1289 cm^{-1}, 1160 – 1028 cm^{-1}), and IgA (1285 – 1237 cm^{-1}) were observed, noticing a higher absorbance in the spectra of COVID-19 group, which is concordant with all those mentioned above.

On the other hand, the second derivative spectra of SARS-CoV-2 patients at 1695 cm^{-1}, 1682 cm^{-1}, 1660 cm^{-1}, 1652 cm^{-1}, 1646 cm^{-1}, 1639 cm^{-1}, 1631 cm^{-1}, and 1625 cm^{-1} showed an absorbance decrease and a displacement, suggesting changes in the protein structures. About this, Diamond et al. declared a decrease in the expression of the mRNA of ACE2 and IL-6 in saliva samples, which would correspond to the decrease in the secondary structures reported by Meirson et al., who through a bioinformatic analysis described that the main secondary structure between the union of SARS-CoV-2 and ACE is the κ-helix structure (polyproline II), followed by the α-helix and β-strand, changing the disulfide bonds [40, 46]. Moreover, Giubertoni et al. assigned the peak at 1619 ± 2 cm^{-1} as helical conformation similar to that of a polyproline II helix, and 1659 ± 2 cm^{-1} as α-helix, which are also diminished in our study [47].

As expected, the immunoglobulins content showed that the COVID-19 group expressed a higher IgA, IgM, and IgG content than the healthy group. Moreover, when comparing the expression of these in the COVID-19 group, it can be observed that the IgA was the least immunoglobin expressed, followed by the IgM; being the IgG the most expressed immunoglobulin, which may be attributed to that most of the samples were collected at day 9.24 after PCR diagnosis, and according to the aforementioned the IgM is detected five days post symptom onset. The earliest detection of IgG is at seven days post symptom onset. Nevertheless, it has to be considered that some samples were obtained the first day when the patients showed symptoms, so that IgA was also detected in this population.

When comparing DNA and nucleic acid content, the COVID-19 group showed a higher content of these molecules. Zaling et al. have declared that in necrotic cell death, the DNA is completely unwound, the

reason by which 100% of the DNA is visible to IR at this stage, observing an increase of ∼65% in DNA absorbance in necrosis compared to the control. Moreover, they also reported a decrease in the random coil structure of the total protein, similarly to the COVID-19 group of this research [48].

On the other hand, as previously mentioned, the characterization of two or more populations from the analysis of the FTIR spectra of their individuals is not an easy task; in a more complex sample, it will be more complicated to find characteristic patterns of the population. This because the links of the different components could overlap with the characteristic component links of each sample. Moreover, the nature of the samples (fluid or tissues, cells, among others) has its particularities.

Different methodologies have been proposed to identify populations from the analysis of FTIR spectra, facilitating the adoption of a classification method by allowing experimentation to focus only on the most promising. In this sense, in another work, we first experimented with linear classification models to discriminate COVID-19 patients, although these models were affected by the overlap of the spectra due to the variances of the absorbances/transmittances of the populations; this problem can be overcome by having a large population thanks to the central limit theorem. In this work, we discriminated against our groups employing an MLRM, which was validated employing a LOOCV according to our previous research.

The absorbance variations and principally the peak displacement associated with viral infections shown in Figs. 3A and 3B contributed to the excellent performance of MLRM. As we note in (1), the slope performs an essential role in MLRM models because a displacement in any peak means that one population has reached its maximum absorbance level while the other continues growing, so its sign is the opposite. Our results presented in Fig. 7 suggest that the best region to identify possible virus carriers is the amide I of proteins (1700 – 1600 cm^{-1}) to compact the outputs between the predictions of the same populations and the separation to the other one.

Enthought the spectra analysis allowed us to detect the molecular components that characterize a positive patient to SARS-CoV-2, and the data analysis through MLRM let us discriminate these patients from healthy persons, more assays need to be done, one of them should consider the time elapsed from the symptoms to the diagnosis and categorize this population. Another one should consider the diagnosis corroboration through the serological test (IgA, IgM, and IgG), correlating these results with the FTIR spectra.

5 Conclusions

In the present research, we established specific salivary vibrational modes employing ATR-FTIR spectroscopy and characterizing the COVID-19 biological fingerprint. These specific spectra can be used to detect possible carriers of the virus or patients who have presented the disease and retain some immunological respect. In any case, it is necessary to analyze and continue investigating the spectra in their different regions to determine their meaning with greater precision.

In addition, these spectra have allowed us to identify a suitable region for COVID-19 detection. By performing the MLRM, the number of variables decreased considerably, which would help us think about viable techniques or devices for diagnosing diseases faster and cheaper.

Declarations

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Written informed consent for the obtention of 1 mL of saliva and participation in this study was obtained from the participants. The Clinical Research Ethics Committee of the Unidad de Especialidades Medicas and the Hospital Central Militar of the Secretaria de la Defensa Nacional approved the protocol and informed consent. All experiments were examined and approved by the appropriate ethics committee and were performed following the ethical standards laid down in the 1964 Declaration of Helsinki.

CONSENT FOR PUBLICATION

The authors give consent for publication.

AVAILABILITY OF DATA AND MATERIAL

All the generated data and the analysis developed in this study are included in this article.

COMPETING INTERESTS

The authors declare no competing financial interest.

FUNDING

We appreciate the support of the A022-2021 (SEDENA) budgetary program.

AUTHORS' CONTRIBUTIONS

AMC, GJVZ, and MMMM designed the project, planned the experiments, and wrote the manuscript. MSB, MGR, and MGG performed the experiments. MACP and APT recollected the saliva samples. JHLP, FGG, and RJDM provide important suggestions about manuscript writing. All authors reviewed the manuscript.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Enrique David Jimenez Acosta and the people who kindly agreed to participate in the project and provided a saliva sample, also to the “Unidad de Especialidades Médicas” (UEM) and the “Hospital Central Militar” of the Secretaría de la Defensa Nacional, especially to the Laboratory Areas for providing all the necessary guarantees to carry out the sample collection.

References

1. Carvalho, L. F. D. C. E. S. & Nogueira, M. S. Optical techniques for fast screening - Towards prevention of the coronavirus COVID-19 outbreak. *Photodiagnosis Photodyn Ther*, **30**, 101765 https://doi.org/https//10.1016/j.pdpdt.2020.101765 (2020).
2. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organ 2021. https://covid19.who.int (accessed May 17, 2021).
3. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. Available at https://www.covid19treatmentguidelines.nih.gov/. Accessed [accessed May 17, 2021].
4. Afzal, A. Molecular diagnostic technologies for COVID-19: Limitations and challenges. *J Adv Res*, **26**, 149–159 https://doi.org/10.1016/j.jare.2020.08.002 (2020).
5. Chang, F. Y. *et al.* Immunologic aspects of characteristics, diagnosis, and treatment of coronavirus disease 2019 (COVID-19). *J Biomed Sci*, **27**, 72 https://doi.org/10.1186/s12929-020-00663-w (2020).
6. Santaella-Tenorio, J. SARS-CoV-2 diagnostic testing alternatives for Latin America. *Colombia medica*, **51**, 2 https://doi.org/10.25100/cm.v51i2.4272 (2020).
7. Yang, H., Yang, S., Kong, J., Dong, A. & Yu, S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nature protocols*, **10**, 382–396 https://doi.org/10.1038/nprot.2015.024 (2015).
8. Ferreira, I. *et al.* Attenuated Total Reflection-Fourier Transform Infrared (ATR-FTIR) Spectroscopy Analysis of Saliva for Breast Cancer Diagnosis. *Journal of Oncology*, https://doi.org/10.1155/2020/4343590 (2020).
9. Melián, R., Calcumil, H. & Boin-Bakit, Carrasco, R. Detección de COVID – 19 (SARS-CoV-2) Mediante la Saliva: Una Alternativa Diagnóstica poco Invasiva. *International journal of odontostomatology*, **14**, 316–320 https://doi.org/10.4067/S0718-381X2020000300316 (2020).
10. Farshidfar, N. & Hamedani, S. The Potential Role of Smartphone-Based Microfluidic Systems for Rapid Detection of COVID-19 Using Saliva Specimen. *Mol. Diagn. Ther*, **24**, 371–373 https://doi.org/10.1007/s40291-020-00477-4 (2020).
11. Wei, Z. *et al.* Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerging Microbes & Infections*, **9**, 386–389 https://doi.org/10.1080/22221751.2020.1729071 (2020).
12. To, K. *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis*, **20**,

565–574 https://doi.org/10.1016/S1473-3099(20)30196-1 (2020).

13. To, K. *et al.* Consistent Detection of 2019 Novel Coronavirus in Saliva. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, **71**, 841–843 https://doi.org/10.1093/cid/ciaa149 (2020).
14. Wang, A., Wang, C., Tu, M. & Wong, D. Oral biofluid biomarker research: current status and emerging frontiers. *Diagnostics*, **6**, 45 https://doi.org/10.3390/diagnostics6040045 (2016).
15. Wang, W. *et al.* Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA*, **323**, 1843–1844 https://doi.org/10.1001/jama.2020.3786 (2020).
16. Caetano, J., Paulo, C., Strixino, J. & Raniero, L. Analysis of saliva by Fourier transform infrared spectroscopy for diagnosis of physiological stress in athletes. *Research on Biomedical Engineering*, **31**, 116–124 https://doi.org/10.1590/2446-4740.0664 (2015).
17. Santos, M., Morais, C. & Lima, K. ATR-FTIR spectroscopy for virus identification: A powerful alternative. *Biomedical Spectroscopy and Imaging*, **3**, 103–118 https://doi.org/10.3233/bsi-200203 (2021).
18. Mata-Miranda, M. M. *et al.* (2019) Characterization of the Biological Fingerprint and Identification of Associated Parameters in Stress Fractures by FTIR Spectroscopy. BioMed Research International, 2019:1241452. https://doi.org/10.1155/2019/1241452
19. Benezzeddine, L., Cazorla, G. & Melin, A. Validation for quantification of immunoglobulins by Fourier transform infrared spectrometry. *Clinical chemistry and laboratory medicine*, **47**, 83–90 https://doi.org/10.1515/CCLM.2009.007 (2009).
20. Gautam, R. *et al.* Molecular profiling of sepsis in mice using Fourier Transform Infrared Microspectroscopy. *J Biophotonics*, **9** (1–2), 67–82 https://doi.org/https:// (2016).
21. Raschka, S. (2020) Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. ArXiv:1811.12808v3.
22. Geinguenaud, F., Militello, V. & Arluison, V. (2020) RNA Spectroscopy Methods and Protocols. Application of FTIR Spectroscopy to Analyze RNA Structure (ed) Humana Press 119–133.
23. Erukhimovitch, V., Talyshinsky, M., Souprun, Y. & Huleihel, M. FTIR microscopy detection of cells infected with viruses. *Methods in molecular biology*, **292**, 161–172 https://doi.org/10.1385/1-59259-848-x:161 (2005).
24. Lee, F., Reynolds, K. & Riley, M. Detection and quantification of poliovirus infection using FTIR spectroscopy and cell culture. *J Biol Eng*, **5**, 16 https://doi.org/10.1186/1754-1611-5-16 (2011).
25. Böger, B. *et al.* Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *American journal of infection control*, **49**, 21–29 https://doi.org/10.1016/j.ajic.2020.07.011 (2021).
26. Peckham, H. *et al.* Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat Commun*, **11**, 6317 https://doi.org/10.1038/s41467-020-19741-6 (2020).

27. Hu, B. *et al.* Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*, **19**, 141–154 https://doi.org/10.1038/s41579-020-00459-7 (2021).
28. Ortiz, B. *et al.* Clinical and Epidemiological Characteristics of Patients Diagnosed with COVID-19 in a Tertiary Care Center in Mexico City: A Prospective Cohort Study. *Revista de investigación clínica*, **72**, 165–177 https://doi.org/10.24875/RIC.20000211 (2020).
29. Berumen, J. *et al.* Risk of infection and hospitalization by Covid-19 in Mexico: a case-control study. *medRxiv*, https://doi.org/10.1017/CBO9781107415324.004 (2020).
30. Petrova, D. *et al.* Obesity as a risk factor in COVID-19: Possible mechanisms and implications. *Aten Primaria*, **52**, 496–500 https://doi.org/10.1016/j.aprim.2020.05.003 (2020).
31. Zhao, J. *et al.* (2020) Relationship between the ABO Blood Group and the COVID-19 Susceptibility. medRxiv. https://doi.org/10.1101/2020.03.11.20031096
32. Canizalez, R. *et al.* Blood Groups Distribution and Gene Diversity of the ABO and Rh (D) Loci in the Mexican Population. *Biomed Res. Int*, https://doi.org/10.1155/2018/1925619 (2018).
33. Velavan, T. & Meyer, C. Mild versus severe COVID-19: Laboratory markers. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases*, **95**, 304–307 https://doi.org/10.1016/j.ijid.2020.04.061 (2020).
34. Borges, L. & Phithon, T. (2020) COVID-19 and Neutrophils: The Relationship between Hyperinflammation and Neutrophil Extracellular Traps. Mediators Inflammation. https://doi.org/10.1155/2020/8829674
35. Wang, D., Hu, B. & Hu, C. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA*, **323**, 1061–1069 https://doi.org/10.1001/jama.2020.1585 (2020).
36. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature*, **583**, 469–472 https://doi.org/10.1038/s41586-020-2332-7 (2020).
37. Bouhaddou, M. *et al.* The Global Phosphorylation Landscape of SARS-CoV-2 Infection., **182**, 685–712 https://doi.org/10.1016/j.cell.2020.06.034 (2020).
38. Erukhimovitch, V., Bogomoln, E., Huleihil, M. & Huleihel, M. Infrared spectral changes identified during different stages of herpes viruseinfection in vitro. *Analyst*, **136**, 2818–2824 https://doi.org/10.1039/c1an15319f (2011).
39. Barauna, V. *et al.* (2020) Ultra-rapid on-site detection of SARS-CoV-2 infection using simple ATR-FTIR spectroscopy and analysis algorithm: high sensitivity and specificity. medRxiv. https://doi.org/10.1101/2020.11.02.20223560
40. Diamond, G., Figgins, E., Robinson, T. & e al Examination of gene expression in saliva samples from COVID-19 patients to study the host defense response against SARS-CoV-2 in the oral cavity. *Molecular Oral Microbiology*, https://doi.org/10.1111/omi.12327 (2020).
41. Suhandynata, R. *et al.* Longitudinal Monitoring of SARS-CoV-2 IgM and IgG Seropositivity to Detect COVID-19. *J Appl Lab Med*, **5**, 908–920 https://doi.org/10.1093/jalm/jfaa079 (2020).

42. Secchi, M. *et al.* COVID-19 survival associates with the immunoglobulin response to the SARS-CoV-2 spike receptor binding domain. *The Journal of clinical investigation*, **130**, 6366–6378 https://doi.org/10.1172/JCI142804 (2020).
43. de Assis, R. *et al.* Analysis of SARS-CoV-2 antibodies in COVID-19 convalescent blood using a coronavirus antigen microarray. *Nature communications*, **12**, 6 https://doi.org/10.1038/s41467-020-20095-2 (2021).
44. Chao, Y., Rötzschke, O. & Tan, E. The role of IgA in COVID-19. Brain, behavior. *and immunity*, **87**, 182–183 https://doi.org/10.1016/j.bbi.2020.05.057 (2020).
45. Varadhachary, A. *et al.* (2020) Salivary anti-SARS-CoV-2 IgA as an accessible biomarker of mucosal immunity against COVID-19. medRxiv. https://doi.org/10.1101/2020.08.07.20170258
46. Meirson, T., Bomze, D. & Markel, G. (2020) Structural basis of SARS-CoV-2 spike protein induced by ACE2. Bioinformatics. https://doi.org/10.1093/bioinformatics/btaa744
47. Giubertoni, G., Meister, K., DeVries, A. & Bakker, H. Determination of the Solution Structure of Antifreeze Glycoproteins Using Two-Dimensional Infrared Spectroscopy. *Journal of Physical Chemistry Letters*, **10**, 352–357 https://doi.org/10.1021/acs.jpclett.8b03468 (2019).
48. Zelig, U., Kapelushnik, J., Moreh, R., Mordechai, S. & Nathan, I. Diagnosis of cell death by means of infrared spectroscopy. *Biophysical journal*, **97**, 2107–2114 https://doi.org/10.1016/j.bpj.2009.07.026 (2009).

Figures

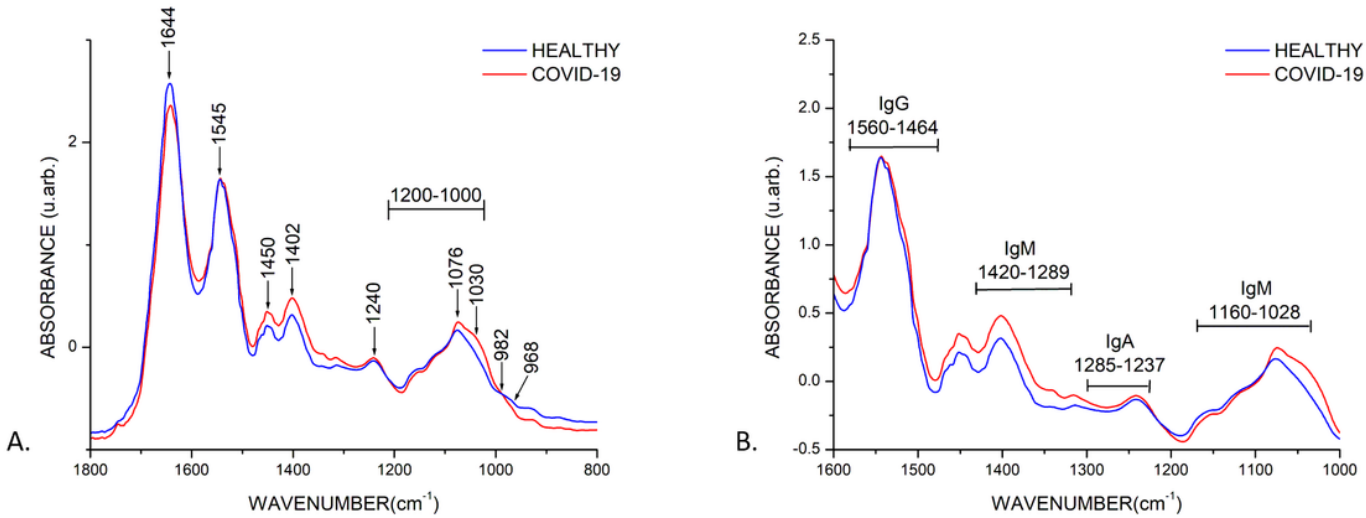

Figure 1

FTIR raw spectra of healthy (n=1209) and COVID-19 groups (n=255). A. Biological fingerprint region, B. Immunoglobulins regions.

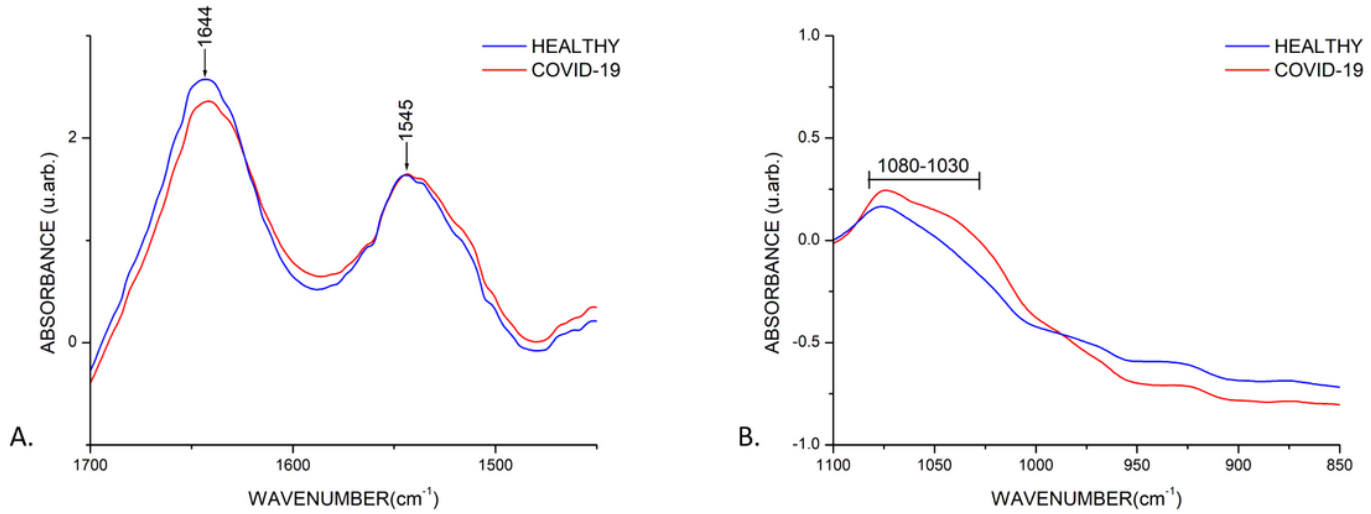

Figure 2

FTIR raw spectra of healthy (n=1209) and COVID-19 groups (n=255). A. Proteins amide I region, B. Nucleic acids region.

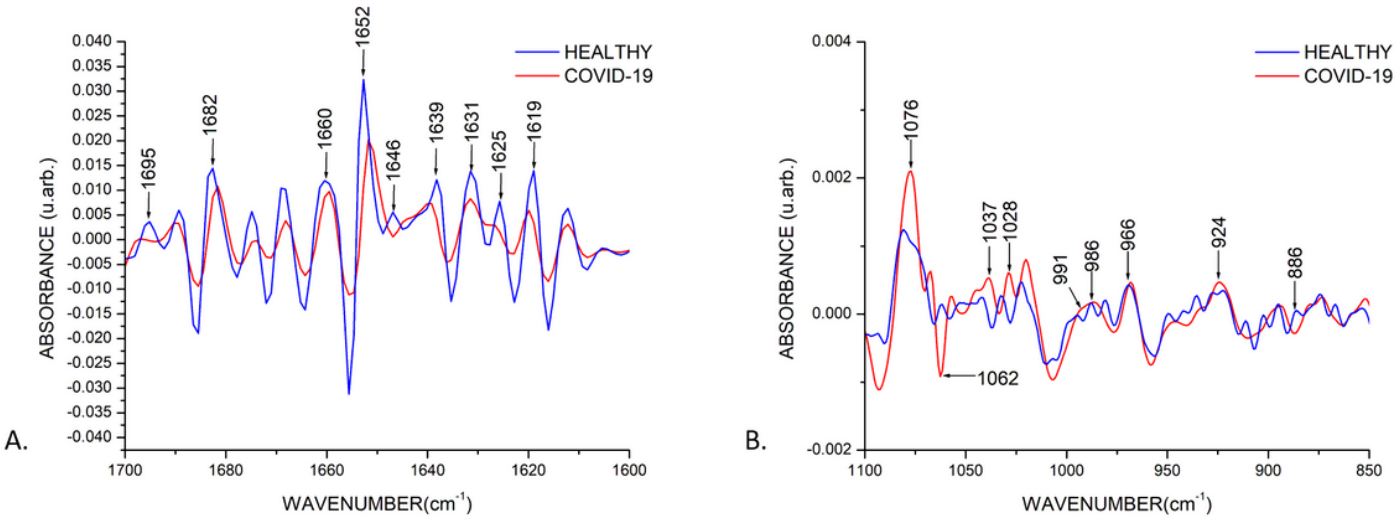

Figure 3

FTIR second derivative spectra of healthy (N=1209) and COVID-19 groups (N=255). A. Proteins amide I region, B. Nucleic acids region.

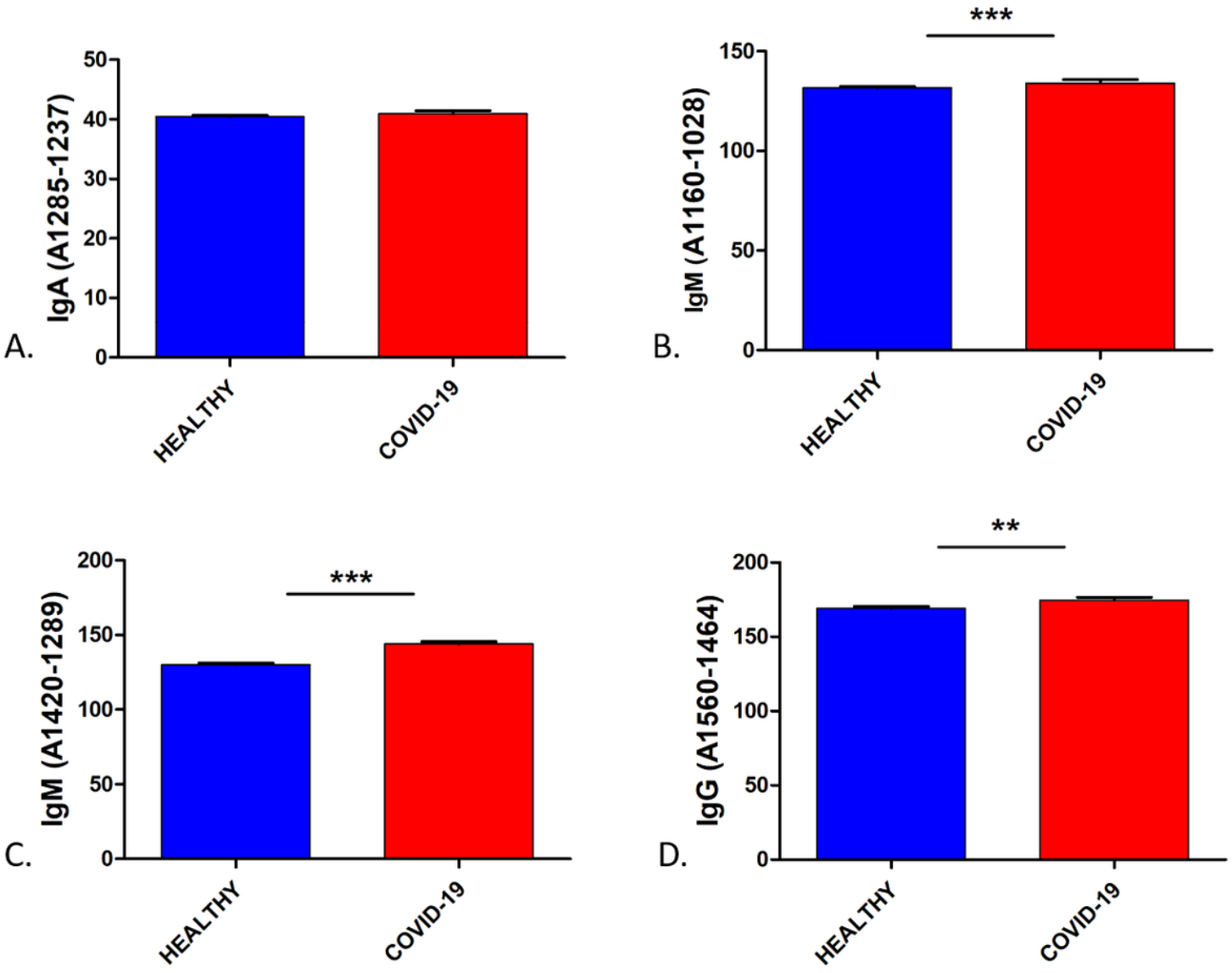

Figure 4

Immunoglobins comparison between healthy and COVID-19 groups. A. IgA (1285-1237 cm-1), B. IgM (1160-1028 cm-1), C. IgM (1420-1289 cm-1), D. IgG (1560-1464 cm-1). **p < 0.05 and ***p < 0.005, compared to the healthy group.

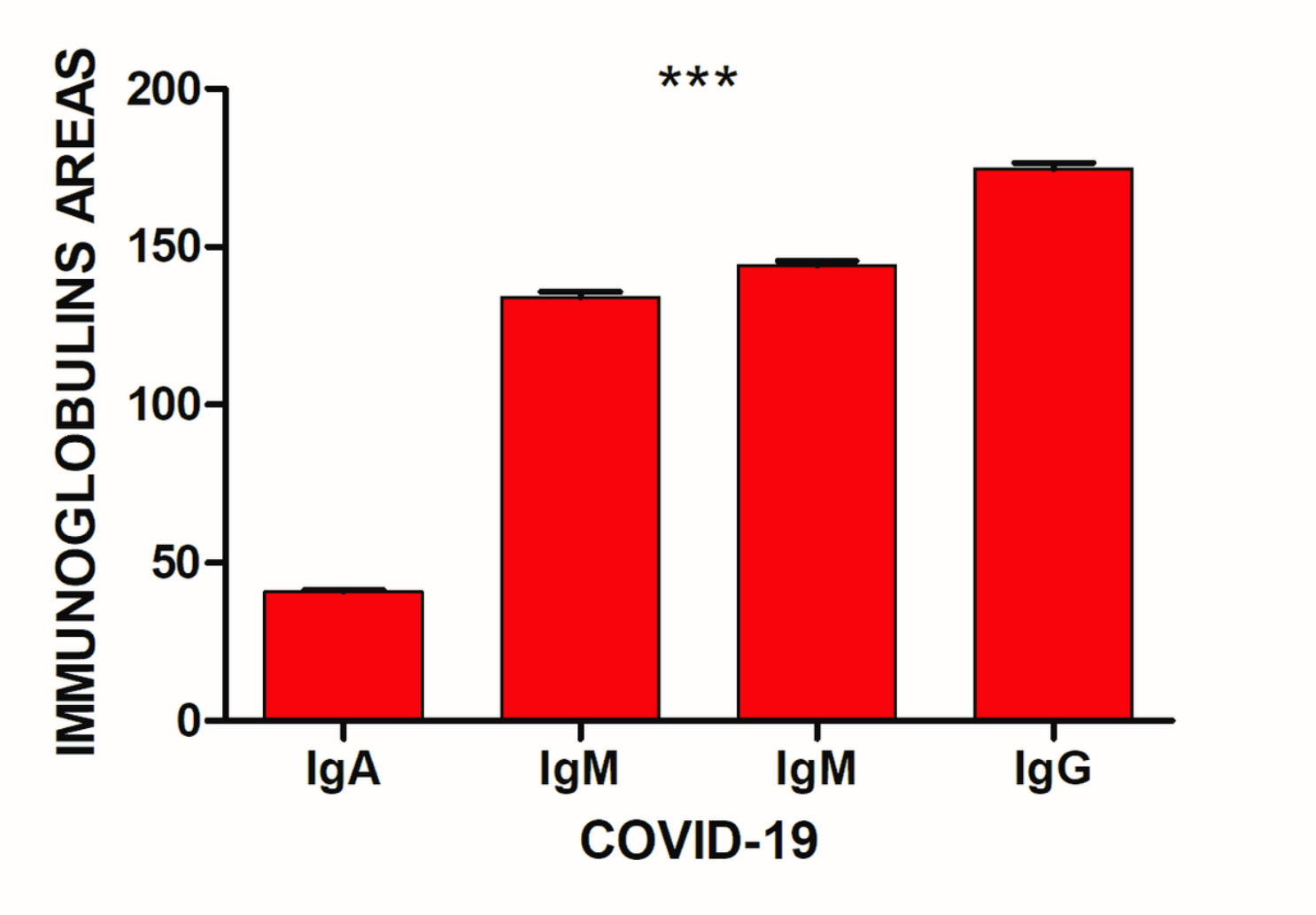

Figure 5

Immunoglobins expression in COVID-19 group. IgA (1285-1237 cm-1), IgM (1160-1028 cm-1, 1420-1289 cm-1), IgG (1560-1464 cm-1). ***p < 0.005.

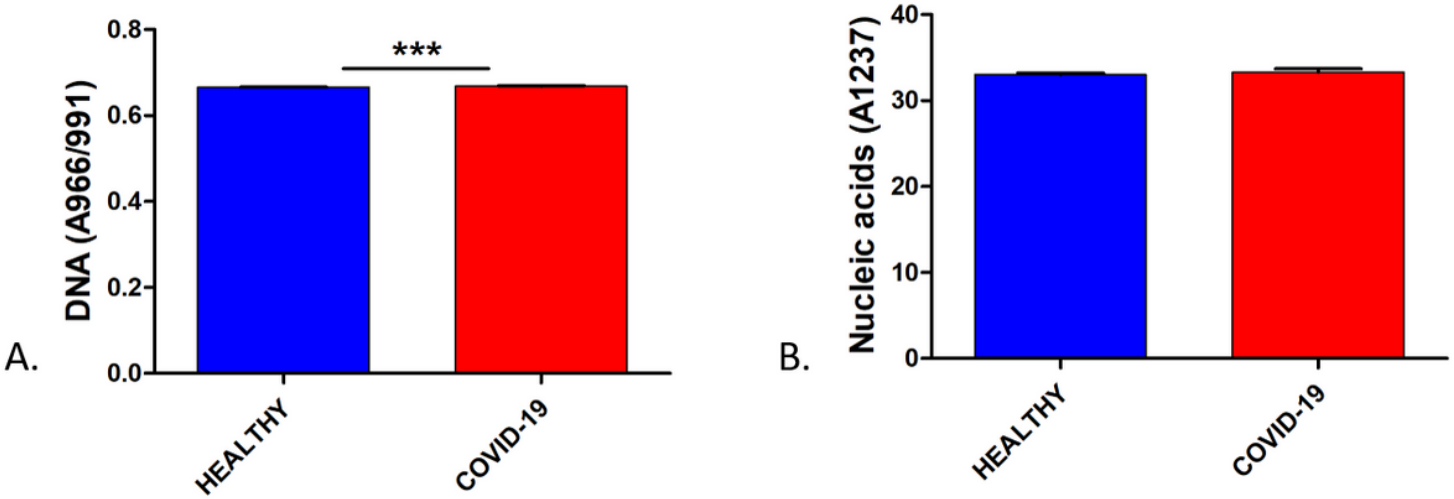

Figure 6

Comparison of the DNA and nucleic acid content between healthy and COVID-19 groups. A. DNA content, B. Nucleic acid content. ***p < 0.005 and ****p < 0.0005, relative to the healthy group.

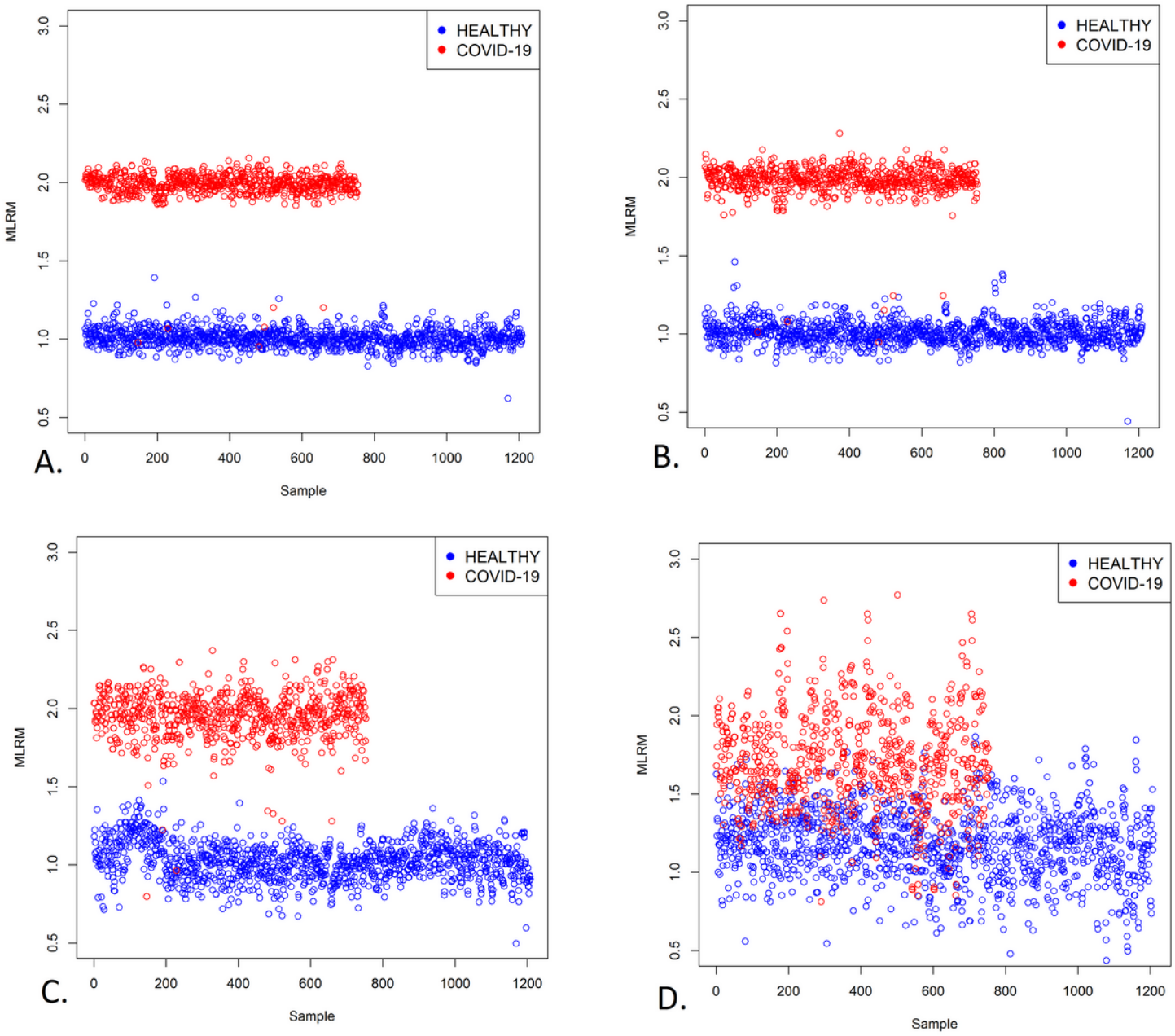

Figure 7

Multiple linear regression model (MLRM) between healthy (N=1209) and COVID-19 (N=255) groups in the regions A. Proteins amide I (1700-1600 cm-1), B. IgG (1560-1464 cm-1), C. IgM (1420-1289 cm-1), D. IgM (1160-1028 cm-1).