SARS-CoV-2 infection testing by infrared spectroscopy and multivariate data analysis using a portable spectrometer

Luciane Eichelt
luciane.eichelt@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2021

Abstract

The spread of COVID-19 during the past two years has dramatically changed our daily lives, resulting in the most unstable and uncertain period our society faces in this century. Being a disease that can progress into severe outcomes and put massive pressure on global healthcare systems, it is mandatory the develop reliable, fast and, not labor-intensive methods of diagnosis.

The capacity of a diagnostic method for the presence of SARS-CoV-2 in human blood, based on a portable FTIR-ATR spectrometer and multivariate data analysis (namely PCA and PLS-DA) had been challenged on the current work. Besides, the capacity on differentiate between IgG positive and IgG negative samples and the repeatability of the technique had been evaluated.

The method proved difficulties to differentiate positive and negative samples for SARS-CoV-2 due to the RMSEC (0.34) and RMSECV (0.39) observed. Contiguous Blocks is a more efficient cross validation technique, providing smaller RMSEP (0.35) and eliminating overfitting effects on the model. Considering the differentiation capacity among IgG positive and IgG negative samples, the model also shows considerable RMSEC (0.38) and RMSECV (0.42) values, even though overfitting is observed due the small number of samples available to validate/calibrate the model. FTIR-ATR evidences high sensitivity to different methodologies for sample treatment (e.g. distinct methodologies for blood collection and processing). Therefore, further studies are required with more extensive datasets and standardized methodologies for the samples processing.

Keywords: SARS-Cov-2; FTIR-ATR spectroscopy; chemometrics; infection diagnosis; discrimination.

.

1. Introduction

The spread of COVID-19 has dramatically changed our daily lives, resulting in the most unstable and uncertain period our society is facing in this century. Infections have been coming in waves, with high contagion and increasing deaths. Many governments are struggling with competing demands for protecting the economy and healthcare systems.

Meanwhile, researchers around the world are dedicating significant effort to find strategies capable of containing the pandemic, reducing hospitalization, finding new diagnosis methods and, mostly, developing vaccines. By our previous experience of almost two years, we have learned that our society will probably coexist with the virus for the years to come and that new viral pandemics are highly likely.

1.1. Coronaviruses

Coronaviruses are a diverse group of large, enveloped, positive stranded RNA viruses that cause diseases in humans and other animals. They can compromise hepatic, enteric, respiratory and neurological systems in different ranges of severity. Four main structural proteins comprise these viruses: Spike (S), Membrane (M), Envelope (E) and the Nucleocapsid (N), that act during host cell entry and virion morphogenesis and release. S proteins on the envelope forming a crown shape around the virus, which gives the name "corona". [01]

Regarding respiratory diseases in humans caused by coronaviruses, three recent outbreaks caused by Coronaviruses (CoV) can be highlighted: Severe Acute Respiratory Syndrome CoV (SARS-CoV) in China in 2003, Middle East Respiratory Syndrome CoV (MERS-CoV) in Saudi Arabia in 2012 and Severe Acute Respiratory Syndrome CoV-2 (SARS-CoV-2) or Coronavirus Disease (COVID-19), the latest being further explored during this work.

1.2. COVID-19

COVID-19 is the disease caused by the SARS-CoV-2 coronavirus. A cluster of pneumonia cases caused by the virus was first identified in December 2019 in Wuhan, China. The World Health Organization declared it as a public health emergency of international concern on the 3rd of January 2020. As of 31th of October 2021, a total of 245,373,039 cases of COVID-19 have been confirmed throughout the world, including 4,979,421 deaths [02]. Currently, there are no cure or preventive therapies for the disease, which increases the need for current understanding of the virus and the epidemics to develop therapeutics.

The clinical presentations of COVID-19 vary from completely asymptomatic infection to severe respiratory failure. A wide variety of symptoms can be observed, from sore throat to conjunctivitis, from diarrhea to skin rash, being the more prevalent ones: fever, dry cough and fatigue. The severity and outcome of the disease is highly influenced by the so called Cytokine Release Storm (CRS) [03].

The disease is transmitted through airborne infected respiratory droplets and direct contact with infected surfaces, when, after touching the surface the person touches its nose, mouth or eyes.

1.3. Cytokine Release Storm (CRS)

Cytokine Release Storm (CRS) induced by pathogens seems to play a role in aggravating clinical symptoms. A possible mechanism of cytokine release syndrome in severe COVID-19 patients happens when SARS-CoV-2 infects alveolar epithelial cells through the ACE2 receptor. Viruses replicate inside the cell leading to its destruction. This process increases cell permeability leading to the release and consequent spread of viruses units. In addition, under the stimulation of inflammatory factors, a large number of inflammatory exudates and erythrocytes enter the alveoli, resulting in dyspnea and respiratory failure. [03]

1.4. Detection tests for COVID-19

Many strategies and techniques for the population mass testing for the presence of the virus SARS-CoV-2, or even for the antibodies presence – demonstrating previous or current contamination. More worldwide applicable methods are hereby listed and briefly described [04].

a) Molecular Nucleic Acid Amplification Tests

These are the reference method (gold standard) for diagnosis and confirmation of SARS-CoV-2 virus in the organism. The most applicable molecular test in this category is the PCR-RT test. This test is widely used to detect gene expression levels and facilitate the rapid diagnosis of acute respiratory viral infections. [05]

b) Rapid Antigen Tests

Rapid tests can either apply small blood drops or collect swabs from the nose and/or throat. Those are immunochromatographic tests, more precise after 14

days symptoms had initiated, when the titration of IgG and IgM antibodies against SARS-CoV-2 are higher. It´s a cheaper methodology and provides faster results than Molecular Nucleic Acid Amplification Tests. [06]

c) Rapid antigen tests for self-testing

These are rapid antigen tests with low complexity to be performed. They allow normal untrained people to perform the tests, instead of health professionals or other qualified professionals. Even though subject to high uncertainties regarding sample collection, it allows the tracking of a major amount of population.

1.5. Infrared Spectroscopy

Infrared spectroscopy is a method that applies the infrared bands of the electromagnetic spectrum: From 12500 to 400 cm^{-1} / from 800 to 25000 nm. It is positioned beyond the red limit of the visible spectrum. Through the reading of spectra in this region, users can interpret peaks and infer details about the composition of a sample, for example, human blood. It measures overtones and combination tones of molecular vibrations in the infrared range, especially the asymmetric vibrations that are intensive in the near-infrared range, i.e. stretch vibrations involving hydrogen bonds (e.g. C-H, O-H and N-H).

A sensor of IR consists of a source of infrared radiation, a detector of the radiation and a dispersive element. In the last decade, the increased development in both instrumentation and chemometric methods positioned the IR qualitative and quantitative analytical technique as the most widespread method for determining the chemical structures of molecules. Also, in this period, a milestone has been set on the evolution path of IR sensors: The introduction of portables spectrometers.

1.6. Attenuated Total Reflectance

Attenuated Total Reflectance technique deals with the concerns previously cited by providing high-quality spectra combined with enhanced reproducibility. ATR measures the changes that occur in a totally internally reflected infrared beam when the beam comes into contact with the sample through the surface of a crystal made of zinc selenide, germanium or diamond. The sample is intact and unmodified since no other components are needed. Therefore, ATR provides valuable data that cannot be obtained with any other method. [08]

1.7. Portable Spectrometers

Portable spectrometers are analytical instruments whose size, weight and dimensions allow their moving to the sample location, generating precise and reliable answers for the operator on the occasion of sample reading. They are generally utilized on the field, which means in a pandemic situation, inside of hospitals and healthcare centers in general. Therefore, the expectation for this equipment is to be small, lightweight and capable of running on battery power for a reasonable duration of time. They also must be able to perform well outside the environment of a laboratory, which is usually very well controlled. The need of reduced dimensions oftentimes results in sacrifice of features, reliability and signal-to-noise [09].

1.8. Chemometrics

An important aspect for spectroscopy users, besides the spectra itself, is the quantification and classification of components in a sample using more than one variable, through multivariate analysis techniques. Given the molecular complexity of biological samples, several common techniques such as chemometrics that combine mathematical and statistical procedures are used to provide chemo-physical evidence from spectroscopic data [08].

Chemometrics is a multidisciplinary approach referring to the analysis of large, complex chemical data obtained from spectroscopic measurements, extracting valuable and useful information through mathematical tools and multivariate statistics. Among the different methods of

chemometrics, they hold the following steps in common (each one further detailed on this work):

1.8.1. Principal Component Analysis (PCA)

PCA (principal component analysis) is a method for dimensionality-reduction of large data sets, which transforms a large set of variables into a smaller one. This small data set still contains most of the information available in the large set. PCA produces linear combinations of the original variables to generate new variables, also known as principal components, or PCs. Mathematically, PCA relies on an eigenvector decomposition of the covariance or correlation matrix of the process variables.

1.8.2. Partial Least-Squares Discriminant Analysis (PLS-DA)

PLS-DA is discriminant method and a tool for the reduction of multivariate dimensionality. In a certain sense, PLS-DA allows to convert data to a lower dimensional space, leading to full awareness of the class labels. For the classification of unknown samples, it is possible to correlate an input (X) and output (y). Therefore, the algorithm employs a "dummy" and "not dummy" variables. PLS-DA allows this classification taking benefit of (0,1) in the output (Y-block) in order to indicate if the input (X) belongs to certain a class or not. For prediction purposes, the test set is reduced to the new dimensions to produce the predicted values. Given a set of training data that belonged to classes G, the PLS-DA model would have produced predicted G values for each test sample [10].

2. Objective

This present work aims to evaluate the reliability of a diagnostic method, based on a portable FTIR-ATR spectrometer and multivariate data analysis, of the presence of SARS-CoV-2 in human blood samples. It also aims to verify if the same methodology may be applied for the differentiation of IgG positive and negative samples.

3. Material and Methods

3.1. Material

The main material used in the collection and analysis of blood samples during the achievement of the current work is detailed on the table 3.1 below.

Equipment	Material	Software
Bench centrifuge	Micropipettes	Solo analysis software® Eigenvector - PLS Toolbox
Laboratory Oven	Microscope Slides	
Freezer -80°C	EDTA tubes	
FTIR ATR Agilent sensor® 4300 handheld FTIR with DTGS detector	Laboratory tubes	
	Laboratory reagents	

Table 3.1 - Equipment, material and software adopted during the process and analysis of blood samples.

3.1.2. FTIR ATR sensor

Agilent® Handheld FTIR is a sensor specially designed for field use and deployment into non-laboratory situations.

The instrument is a portable battery-operated analyzer designed to measure a variety of solid and liquid samples in locations difficult or impossible to analyze in a laboratory. It can be used for quantitative or qualitative analysis of materials. Its main advantages are: Small size, lightweight ergonomics, ease of use, ruggedness and flexibility into one system Instrument is provided with a dedicated sampling probe interface - attenuated total reflectance (ATR) probe [11].

3.1.3. Solo Software®

Solo Software® was applied, in the current study, for the chemometrics analysis of collected samples.

Solo® is a software developed by Eigenvector Research Inc., capable of rendering the PLS_Toolbox (part of the MATLAB® computational environment),

although the download and acquisition of MATLAB® is not required. Solo applies MATLAB® Compiler™ to compile all of the PLS_Toolbox GUI interfaces into a standalone application. To the current study version 8.9.2 had been applied.

4. Methods

4.1. Sample Collection

Samples were collected from three distinct groups: A, B and C. Group A and B correspond to patients admitted in the intensive care unit (ICU) of Curry Cabral hospital, in Lisbon (further designed in this work as group A) and patients on the ICU of Beatriz Ângelo hospital, in Loures (group B). Third group, denominated control group (Group C), consists on people without any symptoms of COVID-19, employees from the University of Lisbon. A summary of the collected data is shown on table 4.1.

	Location of samples collection	Sample Description	Nr. total of samples	PCR Testing?	IgG Testing?	Processed with Ficoll®?
A	Beatriz Ângelo Hospital	All severe patients.	9	Yes	No	Yes
B	Curry Cabral Hospital	Severe and moderate patients.	13	Yes	Yes	Yes
C	University of Lisbon	Control Group. Health people.	21	No	Yes	No

Table 4.1 - Groups A, B and C from where blood samples had been collected

.4.2. Sample Processing

4.2.1. Groups A and B

Collected blood was transferred to EDTA tubes and 15 ml of blood were diluted with 15 ml of PBS pH 7.4. Next, 15 ml de Ficoll Paque Plus® was added to SepMate and the diluted blood was slowly poured over the Ficoll®, originating two layers. After that, the samples were centrifugated at 700 x g for 10 minutes. After centrifugation, the SepMate was poured to other 50 ml tube and centrifugated again at 400 x g for 10 min. The supernatant was poured into another 50 ml tube and centrifugated once again at 3000 x g for 10 minutes Afterwards, plasma samples were collected and stored at -80 °C for subsequent analysis.

4.2.2. Group C

Blood was by venipuncture to EDTA blood collection tubes. Tubes were well mixed by invertion and spin down at 2000 × g on the centrifuge for 10 min at 4°C. Plasma was aliquoted and stored at -20°C.

4.3. Spectroscopic analysis

In the occasion of samples testing an Eppendorf tube containing the desired sample is collected from the freezer. Inside the laboratory, the sample is slowly thawed at room temperature – this procedure must last a minimum of 24 hours. Samples are then centrifuged for 10min at 2000 x g in a centrifuge After this procedure, 10 µL of plasma is placed on a microscope slide, using an automatic pipette on this task.

Following, a coverslip is placed over the microscopic slide with blood plasma-facing up the slide. The slides are placed in a laboratory oven at 24°C for 30 min. Afterwards samples were read by the FTIR ATR sensor Agilent® 4300 handheld FTIR with DTGS detector.

5. Results

5.1. Spectra Interpretation

Interpretation of spectra obtained through the methodology ATR FTIR had been done based on the results of Naumann et. all [07]. Even though figure 4.1 refers to the spectra obtained from microbiology material [07], the peaks obtained - due to the functional groups observed on biological material - will be applied on the reading and interpretation of results on this current work, as per figure 4.1.

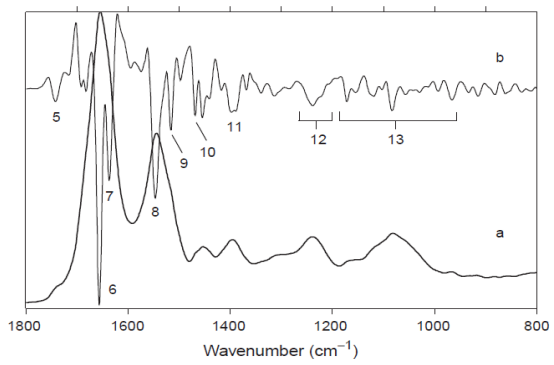

Figure 4.1 - Tentative band assignment of some bands frequently observed in bacterial MIR spectra (*Staphylococcus aureus* (strain SG 511)). (a) Original absorbance spectrum; (b) second derivative. (5) =C=O stretching; (6) amide I of a-helical structures; (7) amide I of b-sheet structures; (8) amide II; (9) tyrosine ring vibration band; (10) CH2 bending; (11) -COO- symmetric stretching; (12) PO2 asymmetric stretching; (13) spectral range dominated by complex ring vibrations of carbohydrates, P=O=P stretching, C-O=P stretching, PO2 □symmetric stretching. Technique: A/T; number of scans: 64; nominal physical resolution: 6 cm^{-1}; apodization function: Blackman-Harris, 3-term; IR cuvette used, see Figure 6(a); spectrometer: IFS 28/B (BrukerOptics, Germany).

5.2. General

Each sample was measured three times, obtaining three spectra of the same sample (Reading in triplicates), as figure 5.1 below.

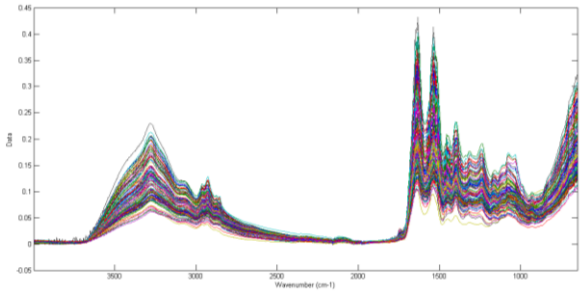

Figure 5.1 - Raw FTIR-ATR spectra for the entire set of collected samples. Wavenumbers from 4000cm^{-1} to 600 cm^{-1}

Aiming to check whether there is a significant difference between the raw spectra of a positive and a negative

Once FTIR-ATR is a very sensitive method, able to capture different functional groups on the molecules, it can also register the presence/absence of residual reagents applied to the processing of the samples. Group A and B samples had been treated with Ficoll® - same groups corresponding to samples positive to SARS-CoV-2. In opposition, samples from group C (negative) were not treated with Ficoll®..

No remarkable influence from the presence of Ficoll®, at least on the raw spectra of the samples, is captured by the reading with FTIR-ATR sensor, as shown in figure 5.2.

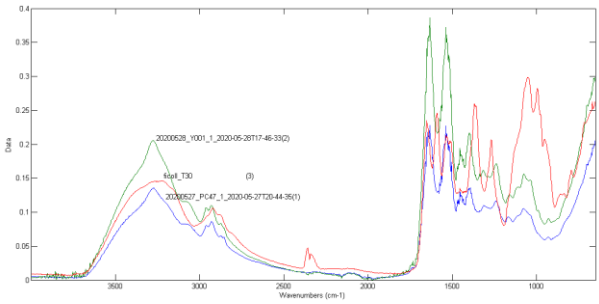

Figure 5.2 - Raw FTIR-ATR spectra of a positive (green curve), a negative (blue curve) samples for COVID-19 and Ficoll® (red curve). Wavenumbers from 4000cm^{-1} to 600 cm^{-1}

Wavenumbers from 4000 cm^{-1} to 2000 cm^{-1} were eliminated from the study, reducing the influence of intramolecular water on the spectra readings (-OH groups stretching over the entire wavenumber range, especially at approximately 3400 cm$^{-1)}$. Wavenumbers smaller than 900 cm^{-1} were also excluded from the analysis. The region between 900 and 600 cm^{-1} exhibits a variety of weak, but extremely characteristic, features overlapped on the basal broad spectral contour [07].

5.3. Repeatability Analysis

A repeatability analysis of the technique was performed, ensuring the confiability of the spectra acquisition. Two scenarios had been chosen to investigate the Repeatability of the analytical method:

a) Repeatability among different groups (A, B and C)

A PCA score plot considering three samples, each one belonging to one of the different groups was performed. Readings were executed in triplicate, each sample being read three times.

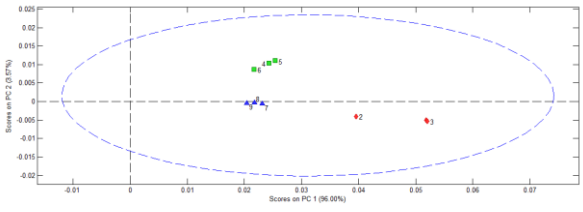

Figure 5.3 - PCA scores plot for Principal Components 1 and 2 showing three different samples read in triplicate: 20200522_111 (1, 2 and 3), 20200527_PC35 (4, 5 and 6) and 20200528_Y009 (7, 8 and 9). Wavelengths: 2000 cm^{-1} to 900 cm^{-1}.

Figure 5.3 indicates that repeatability can be ensured among different groups, once the three different readings of each sample converge to the same point on individual groups.

a) Repeatability samples same group - Same day and time

A PCA score plot of three samples from group B, all of them analyzed at the same day and consecutively was performed. This investigation was done to eliminate any possible inter operator variability that can influence the methodology repeatability. Readings were executed in triplicate, each sample being read three times. It is possible to evidence that replicate readings of the same sample converge to the same area (Figure 5.4).

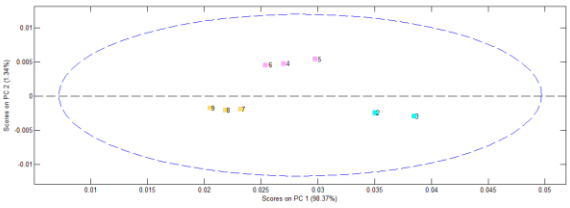

Figure 5.4 - PCA scores plot for Principal Components 1 and 2 showing three samples from group B read in triplicate: 20200528_Y007 (1, 2 and 3), 20200528_Y008 (4, 5 and 6) and 20200528_Y009 (7, 8 and 9). Wavelengths: 2000 cm^{-1} to 900 cm^{-1}.

5.4. Differentiation on positive and negative samples for SARS-CoV-2 – Groups A, B and C

An exploratory analysis of the entire dataset was performed using PCA. This aimed to understand if the chosen methodology is capable of segregating positives and negative samples. Also, it aims to identify, through loadings, which are the main factors influencing the segregation of samples.

Optimum number of factors was selected considering a correlation on the values of RMSECV and RMSEC. When the values start to diverge, the Principal Components are no longer considered. Therefore, five Principal Components are the best for this model.

Different plots had been performed comparing all Principal Components, observing which ones provide better visual segregation between different groups. The best results in terms of discrimination were achieved when comparing PC2 x PC3.

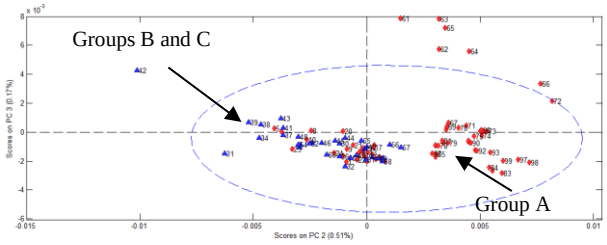

Figure 5.5 - Score plots of group A and B (PCR Positive samples in red) and group C (Control in blue). Principal Components 2 and 3

It is possible to evidence the formation of two clusters (Fig. 5.5): The first one, positive samples collected at the Curry Cabral hospital (Group A, in red, on the right side of the score plot) and samples collected at the Beatriz Ângelo hospital (Group B). The latter are completely mixed on the second cluster to control samples (Group C). An analysis from the loadings on the Principal Component 2 was carried through, aiming to explain the big differentiation on the PCA clusters, as per figure 5.6.

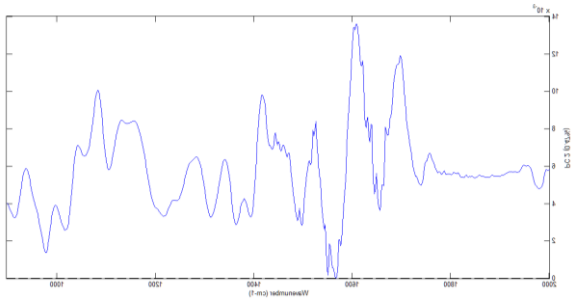

Figure 5.6 - Loadings for Principal Component 2 – Groups A, B and C. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1}

It is possible to notice stretching vibrations on the region between 1800 and 1500 cm^{-1}, originated by the presence of proteins at the samples. These proteins can be related to elements present in the plasma due to the immune system activation against SARS-CoV-2 virus, as antibodies, cytokines, etc.

PLS-DA was also applied, as a supervised pattern recognition technique used to classify the unknown samples to positive/negative predefined classes.

- Total of samples: 99
- Cross Validation methods: Venetian Blinds and Contiguous Blocks
- Calibration set: 70 samples

- External Validation set: 29 samples

A graph comparing the average Cross-Validation Error for Classification with average Calibration Classification Error was determined to establish the number of Latent Variables needed to build each model. Therefore, five Latent Variables had been chosen to Venetian Blinds cross-validation and three Latent Variables for Contiguous Blocks.

Both cross-validation methods have big calibration and cross-validation errors. Therefore, the model doesn't show good capacity of the model to differentiate PCR positive samples from groups A and B from negative ones from Control group C.

Considering the predictive capacity of the model when it is submitted to external samples, both models demonstrate certain capacity on differentiate positive and negatives, with considerable predictive erros. Contiguous Block possess a better predictability capacity, whereas Venetian Block tends to overfitting. This can be stated on table 5.1.

	Venetian Blinds	Contiguous Block
RMSEC	0.28	0.34
RMSECV	0.33	0.39
RMSEP	0.32	0.35

Table 5.1 - RMSEC, RMSECV and RMSEP values for PLS-DA models, applying two different cross-validation methods: Venetian Blinds and Contiguous Block

5.5. Samples IgG Positive and Negative

Immunoglobulin G (IgG) is an antibody that characterize previous exposure to an specific antigen and some degree of immunity to later reinfection. A discriminatory analysis was performed only for the control sample set (Group C), since there is no variance in the analytical methodology of extraction/processing of these samples.

The dataset comprises on a total of 21 samples (read in three replicates by the FTIR sensor), among which 11 are positive and 10 negative for IgG. Optimum number of factors was selected, considering a correlation on the values of RMSECV and RMSEC. Five Principal Components were selected for this model.

In order to determine which principal components have a stronger influence on the separation of IgG positive and IgG negative samples different plots had been performed comparing all Principal Components. Best separation is achieved when comparing PC1 and PC3, as evidenced on the scores plot Figure 5.7.

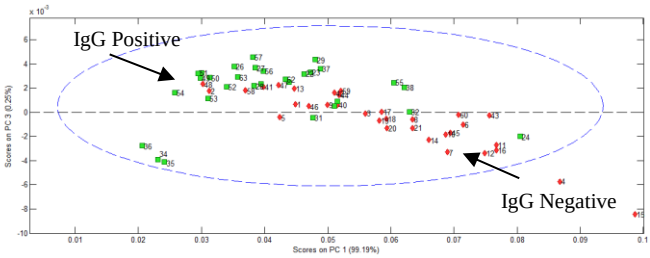

Figure 5.7 - Score plots of group C. IgG Positive samples in green and IgG Negative samples in red. Principal Components 1 and 3

Through loadings analysis is possible to evidence, in PC1, a great influence of wavenumbers 1800 and 1500 cm-1, confirming the presence of amide I and II bands – representing the presence of proteins, structural molecule of antibodies, at the sample.

A PLS-DA was performed with control samples (group C) only to differentiate IgG positive and IgG negative samples.

- Total of samples: 63
- Cross Validation methods: Venetian Blinds and Contiguous Blocks
- Calibration set: 45 samples
- External Validation set: 18 samples

Both cross-validation methods have considerable calibration and cross-validation errors, considering a threshold of 0.5. This confirms the good capacity of the model to differentiate IgG positive and IgG negative samples.

Regarding the model's predictive capacity when challenged by a new set of samples, it is possible to observe table 5.2. It shows a tendency of overfitting for both cross-validation methodologies, and therefore a high uncertainty on the prediction whether a sample is IgG positive or not. Besides, the predictive error is also considerable.

	Venetian Blinds	Contiguous Block
RMSEC	0.32	0.38
RMSECV	0.36	0.42
RMSEP	0.38	0.43

Table 5.2 - RMSEC, RMSECV and RMSEP values for PLS-DA models, applying two different cross-validation methods: Venetian Blinds and Contiguous Block

5.6. Differentiation on Group A (Beatriz Ângelo Hospital) compared to group B (Curry Cabral Hospital)

During the exploratory analysis of the different datasets (Figure 5.8), it was possible to evidence a clear separation between the samples of groups A and B. These samples were collected in different hospitals.

It is possible to observe an important discrimination between the two datasets, originating two clusters on the PCA (Figure 5.9).

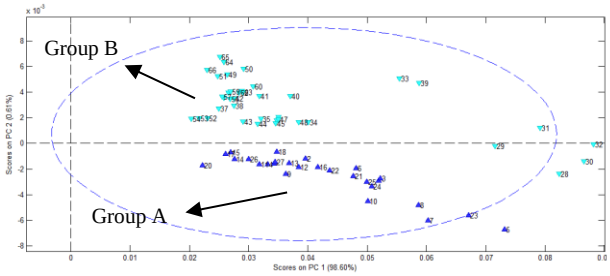

Figure 5.8 - Score plots of group A and B (PCR Positive samples). In light blue, group B – samples collected at Curry Cabral hospital. In blue, group A – samples collected at Beatriz Ângelo Hospital. Principal Components 1 and 2

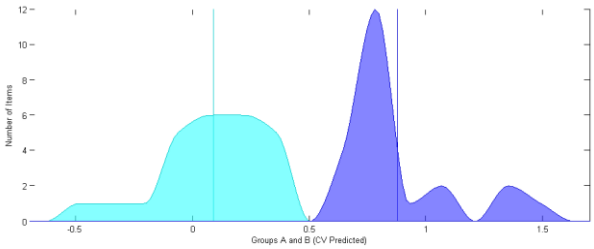

Figure 5.9 - Histogram for PLS-DA Cross-Validation predictions of samples collected in Curry Cabral (light blue) and Beatriz Ângelo (deep blue) hospitals - Cross Validation method: Contiguous Blocks.

The model is highly capable of separating which samples were collected in different hospitals.

5. Conclusion

The reproducibility capacity had been confirmed to the current work, both for samples belonging to different groups (A, B and C) or three samples read consecutively. The three readings of each one of the three samples tested converge to the same point, forming separated clusters.

Considering the results obtained during this work, it is possible to affirm that the method of analysis FTIR ATR can differentiate positive and negative samples for SARS-CoV-2, even though the current methodology still shows big RMSEC and RMSECV observed (0.28 / 0.33 for Venetian Blinds and 0.34/0.39 for Contiguous Block). Contiguous Blocks is a more efficient cross validation technique, providing smaller predictive errors (RMSEP 0.35) and reducing overfitting effects on the model. The influence of Ficoll® on the separation of samples cannot be confirmed in the current study as, when comparing samples submitted to the same analytical methodology (as those described below), the model is able to capture differences among samples.

Considering the differentiation capacity among IgG positive and IgG negative samples, the model also shows big RMSEC and RMSECV values (0.32 / 0.36 for Venetian Blinds and 0.38/0.42 for Contiguous Block). It evidences also a not strong capacity on differentiating IgG positive and IgG negative samples. On the other hand, the prediction errors obtained with the two cross-validation methodologies (0.39 and 0.43) show a tendency of the model to overfitting. When submitted to unknown samples, the model is not able to differentiate whether it is IgG positive or negative, due to the reduced number of samples available to validate / calibrate the model.

Besides, the model is highly able to differentiate samples collected in each one of the hospitals: Curry Cabral and Beatriz Ângelo. Possibly this happens due to distinct procedures and methodologies for collecting blood samples applied in different hospitals.

6. 1. Future Perspectives

The use of different reagents on the blood samples treatment (such as Ficoll®) can interfere with the

readings and identification of reliable and robust models. Moreover, one suggestion is to perform new studies with standard procedures of collecting and processing the samples.

Also, acquiring a bigger number of samples will support building a more reliable model, calibrating and validating them with a more substantial dataset. This can eliminate the tendency of overfitting.

The clear differentiation observed between samples collected at Hospital Beatriz Ângelo and Hospital Curry Cabral arises, very likely, due to procedures and methodologies for collecting blood samples applied in different hospitals This assumption must be confirmed in further studies

References

[01] He F, Deng Y, Li W. Coronavirus Disease 2019 (COVID-19): What we know? Journal of Medical Virology [Internet]. 2020 Mar 14 [cited 2021 May 10];92(7). Available from: https://www.ncbi.nlm.nih.gov/pubmed/32170865

[02] World Health Organization. WHO COVID-19 Dashboard [Internet]. covid19.who.int. World Health Organization; 2021 [cited 2021 Oct 31]. Available from: https://covid19.who.int

[03] Suarez Sanchez C. COVID-19 and cytokine storm syndrome [Internet]. www.mlo-online.com. 2020 [cited 2021 Oct 1]. Available from: https://www.mlo-online.com/continuing-education/article/21138224/covid19-and-cytokine-storm-syndrome

[04] Guias da Saúde - Teste COVID-19 [Internet]. SNS24. 2021 [cited 2021 Nov 4]. Available from: https://www.sns24.gov.pt/guia/teste-covid-19/

[05] Chung Y-S, Lee N-J, Woo SH, Kim J-M, Kim HM, Jo HJ, et al. Validation of real-time RT-PCR for detection of SARS-CoV-2 in the early stages of the COVID-19 outbreak in the Republic of Korea. Scientific Reports [Internet]. 2021 Jul 20 [cited 2021 Sep 5];11(1):14817. Available from: https://www.nature.com/articles/s41598-021-94196-3

[06] Panbio COVID-19 IgG/IgM Rapid Test [Internet]. www.globalpointofcare.abbott. Abbott Laboratories; [cited 2021 Nov 4]. Available from: https://www.globalpointofcare.abbott/pt/product-details/panbio-covid-19-igg-igm-antibody-test-br.html

[07] Lasch P, Naumann D. Infrared Spectroscopy in Microbiology. Encyclopedia of Analytical Chemistry [Internet]. 2015 Mar 12 [cited 2021 Sep 15];1–32. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9780470027318.a0117.pub2

[08] Balan V, Mihai C-T, Cojocaru F-D, Uritu C-M, Dodi G, Botezat D, et al. Vibrational Spectroscopy Fingerprinting in Medicine: from Molecular to Clinical Practice. Materials [Internet]. 2019 Sep 6 [cited 2021 Oct 1];12(18):2884. Available from: https://www.mdpi.com/1996-1944/12/18/2884/htm

[09] Crocombe RA, Leary PE, Kammrath BW. Portable spectroscopy and spectrometry. 2, Applications. Hoboken: John Wiley & Sons, Inc; 2021.

[10] Balan V, Mihai C-T, Cojocaru F-D, Uritu C-M, Dodi G, Botezat D, et al. Vibrational Spectroscopy Fingerprinting in Medicine: from Molecular to Clinical Practice. Materials [Internet]. 2019 Sep 6 [cited 2021 Oct 1];12(18):2884. Available from: https://www.mdpi.com/1996-1944/12/18/2884/htm

[11] Handheld FTIR, Mobile Nondestructive Testing [Internet]. www.agilent.com. Agilent Technologies, Inc; 2014 [cited 2021 Oct 1]. Available from: https://www.agilent.com/en/product/molecular-spectroscopy/ftir-spectroscopy/ftir-compact-portable-systems/4300-handheld-ftir