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

Luciane Eichelt

Thesis to obtain the Master of Science Degree in

Pharmaceutical Engineering

Supervisors: Prof. Dr. João Pedro Martins de Almeida Lopes
Prof. Ana Margarida Nunes da Mata Pires de Azevedo

Examination Committee

Chairsperson: Prof. Dr. José Monteiro Cardoso de Menezes
Supervisor: Prof. Dr. João Pedro Martins de Almeida Lopes
Members of the Committee: Dra. Clara Cecília Santana de Sousa

November 2021

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Preface

The work presented in this thesis was performed at the Molecular Biology Laboratory at Faculty of Farmacy at University of Lisbon (Lisbon, Portugal), during the period February-August 2021, under the supervision of Prof. João Pedro Martins de Almeida Lopes. The thesis was co-supervised at Instituto Superior Técnico by Prof. Ana Margarida Nunes da Mata Pires de Azevedo.

Acknowledgments

To my mother and father, my aunt and my uncle, and all family members who had supported me in this journey.
To my twin sister.
To Gilson.
To nne and Nordika, companies that allow(ed) me to flourish and thrive on my professional life.
To the University of Lisbon and Instituto Superior Técnico. A special thank you to Prof. Dr. João Pedro Martins de Almeida Lopes, for guiding me through these challenging times.

I am grateful.

“Sinto-me nascido a cada momento
Para a eterna novidade do Mundo…”
Fernando Pessoa

“Life can only be understood backwards; but it must be lived forwards”
Søren Kierkegaard

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 showed 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 reducing 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.

Resumo

A disseminação da COVID-19 durante os últimos dois anos mudou drasticamente nossa vida diária, tornando-se o período mais instável enfrentado pela sociedade contemporânea. O alto contágio e razoável potencial de agravamento da doença tem colocado grande pressão nos sistemas globais de saúde. Portanto, o desenvolvimento de métodos de diagnóstico confiáveis e rápidos tornou-se imprescindível no controle da pandemia.

O presente trabalho visa testar a capacidade de um método de diagnóstico para a presença de SARS-CoV-2 no sangue humano, com base em espectometria portátil FTIR-ATR e análise de dados multivariada (nomeadamente PCA e PLS-DA). Além disso, foi avaliada a capacidade do método em diferenciar amostras IgG positivas e IgG negativas e a repetibilidade da metodologia.

O método demonstrou ser capaz de diferenciar amostras positivas e negativas para SARS-CoV-2, muito embora com valores consideráveis de RMSEC (0,34) e RMSECV (0,39). Blocos contíguos foi a metodologia de validação cruzada mais eficiente, fornecendo erros preditivos menores (RMSEP = 0,35) e reudizindo efeitos de overfitting no modelo. Considerando a capacidade de diferenciação entre amostras IgG positivas e IgG negativas, o modelo também apresenta consideráveis valores de RMSEC (0,38) e RMSECV (0,42), embora seja observado overfitting devido ao pequeno número de amostras disponíveis para validar / calibrar o modelo. FTIR-ATR mostrou-se um método altamente sensível a diferentes metodologias de tratamento das amostras de sangue analisadas. Conclui-se que a técnica mostra-se promissora e futuros estudos, elaborados com maior número de amostras e com técnicas de processamento padronizadas, devem ser realizados.

Palavras-chave: SARS-Cov-2; espectroscopia FTIR-ATR; quimiometria; diagnóstico de doenças; discriminação.

Contents

List of tables ..... **ix**
List of Figures ..... **x**
Acronyms ..... **xiii**
Chapter 1. ..... **1**
Introduction **1**
1.1 Coronaviruses 1
1.2 COVID-19 2
1.3 Cytokine Release Storm (CRS) 3
1.4 Detection tests for COVID-19 4
1.5 Current laboratory techniques diagnosis of COVID-19 5
1.5.1 Immunoassays 5
1.5.2 Bioassays 5
1.5.2.1 Flow Cytometry 6
1.5.2.2 Nanoparticle-modified aptamers 6
1.6 Infrared Spectroscopy 6
1.7 Attenuated Total Reflectance 8
1.8 Portable Spectrometers 9
1.9 Chemometrics 10
1.9.1 Principal Component Analysis (PCA) 12
1.9.2 Partial Least-Squares Discriminant Analysis (PLS-DA) 14
1.10 Currently on the realm of COVID-19 diagnosis using sensors 14
Chapter 2. ..... **17**
Objective **17**
Chapter 3. ..... **18**
Material **18**
3.1.2 Bench Centrifuge 18
3.1.3 FTIR ATR sensor 19
3.1.4 Solo Software® 21
Chapter 4. 23
Methods 23
4.1 Sample Collection 23
4.2. Sample Processing 23
4.2.1 Groups A and B 23
4.2.2 Group C 24
4.3 Spectroscopic analysis 24
4.4 Analytical Method 28
Chapter 5. 29
Results 29
5.1 Spectra Interpretation 29

5.2 General 30
5.3 Repeatability Analysis 33
5.4 Differentiation on positive and negative samples for SARS-CoV-2 – Groups A, B and C 35
5.5 Samples IgG Positive and Negative 41
5.6 Differentiation on Group A (Beatriz Ângelo Hospital) compared to group B (Curry Cabral Hospital) 47
Chapter 6. 50
Conclusion 50
6.1 Future Perspectives 51
References 52
Annexes 55
Table 01: Complete profile of samples 55

List of tables

Table 3.1 - Equipment, material and software adopted during the process and analysis of blood samples. 18

Table 4.1 - Groups A, B and C from where blood samples had been collected. 23

Table 4.2 - Samples collected on the current study. Group A, collected at Beatriz Ângelo Hospital, group B at Hospital Curry Cabral and group C is the Control Group. 26

Table 5.1 - Samples for repeatability analysis for each group - A, B and C 34

Table 5.2 - Samples for repeatability analysis three samples on group B. 35

Table 5.3 - Principal Components and variance captured for the entire dataset (groups A, B and C) . 37

Table 5.4 - RMSEC, RMSECV and RMSEP values for PLS-DA models, applying two different cross-validation methods: Venetian Blinds and Contiguous Block 41

Table 5.5 - Principal Components and variance captured for group C (IgG Positive and IgG Negative samples) 42

Table 5.6 - RMSEC, RMSECV and RMSEP values for PLS-DA models, applying two different cross-validation methods: Venetian Blinds and Contiguous Block 47

List of Figures

Figure 1.1 - Environmental and Clinical samples of 2019-nCoV. β coronavirus 2

Figure 1.2 - Electromagnetic spectrum.. 7

Figure 1.3 - Schematic representation of Near-Infrared Spectroscopy apparatus 8

Figure 1.4 - Requirements that must be considered to a successful portable spectroscopy platform.. 10

Figure 1.5 - Flowchart for chemometrics 10

Figure 1.6 - Graphical outlook of Principal Components Analysis 13

Figure 3.1 - Centrifuge Eppendorf® model 5810 R 19

Figure 3.2 - Centrifuge Eppendorf® Rotor A-4-81 19

Figure 3.3 - Agilent 4300 Handheld FTIR with DTGS detector 20

Figure 3.4 - Solo® Software Browse Graphical User Interface. Version 8.9.2 22

Figure 5.1 - Tentative band assignment of some bands frequently observed in bacterial MIR spectra (*Staphylococcus aureus* (strain SG 511)). 29

Figure 5.2 - Raw FTIR-ATR spectra for the entire set of collected samples. Wavenumbers from 4000cm^{-1} to 600 cm^{-1} 31

Figure 5.3 - Raw FTIR-ATR spectra of a positive (green curve) and a negative (blue curve) samples for COVID-19. Wavenumbers from 4000cm^{-1} to 600 cm^{-1} 31

Figure 5.4 - 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} 32

Figure 5.5 - Raw FTIR-ATR spectra of a positive (green curve), a negative (blue curve) samples for COVID-19 - Excluding wavenumbers from 4000 to 2000 cm^{-1} and 900 to 6000 cm^{-1} 33

Figure 5.6 - 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}. 34

Figure 5.7 - 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} 35

Figure 5.8 - RMSEC and RMSECV for groups A, B and C 36

Figure 5.9 - Score plots of group A and B (PCR Positive samples in red) and group C (Control in blue). Principal Components 2 and 3 37

Figure 5.10 - Loadings for Principal Component 2 – Groups A, B and C. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1} 38

Figure 5.11 - Average Cross-Validation Error for Classification and Average Calibration Classification Error - Groups A, B and C. Cross Validation method: Venetian Blinds 39

Figure 5.12 - Average Cross-Validation Error for Classification and Average Calibration Classification Error - Groups A, B and C. Cross Validation method: Contiguous Blocks..... 39

Figure 5.13 - Histogram for PLS-DA Cross-Validation predictions of Positive (groups A and B – red) x Negative (group C - blue) samples - Cross Validation method: Venetian Blinds..... 40

Figure 5.14 - Histogram for PLS-DA Cross-Validation predictions of Positive (groups A and B – red) x Negative (group C - blue) samples - Cross Validation method: Contiguous Blocks. 40

Figure 5.15- RMSEC and RMSECV for group C (IgG Positive and IgG Negative samples) 42

Figure 5.16 - Score plots of group C. IgG Positive samples in green and IgG Negative samples in red. Principal Components 1 and 3 43

Figure 5.17 - Loadings for Principal Component 1 – Group C. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1} 43

Figure 5.18 - Loadings for Principal Component 3 – Group C. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1} 44

Figure 5.19 - Average Cross-Validation Error for Classification and Average Calibration Classification Error – Group C. Cross Validation method: Venetian Blinds 45

Figure 5.20 - Average Cross-Validation Error for Classification and Average Calibration Classification Error – Group C. Cross Validation method: Contiguous Blocks..... 45

Figure 5.21 - Histogram for PLS-DA Cross-Validation predictions of IgG Positive (green) x IgG Negative (red) samples - Cross Validation method: Venetian Blinds..... 46

Figure 5.22 - Histogram for PLS-DA Cross-Validation predictions of IgG Positive (green) x IgG Negative (red) samples - Cross Validation method: Contiguous Blocks. 46

Figure 5.23 - 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 48

Figure 5.24 - Loadings for Principal Component 2 – Groups A and B. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1}..... 48

Figure 5.25 - 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. 49

Acronyms

ACE2 Angiotensin-converting enzyme 2
ATR Attenuated Total Reflectance
AUC Area under the curve
CRS Cytokine Release Storm
CV Cross Validation
DGTS Deuterated triglycine sulfate
DNA Deoxyribonucleic Acid
EDA Exploratory Data Analysis
ELISA Enzyme-linked immunosorbent assay
FDA Food and Drug Administration
FTIR Fourier-transform infrared
GUI Graphical User Interfaces
ICU Intensive Care Units
IgG Immunoglobulin G
IgM Immunoglobulin M
IL-6 Interleukin 6
IR Infrared Region
LFIA Lateral flow immunoassay
LV Latent Variables
NDT Non-destructive tool
NIR Near Infrared
PC Principal Components
PCA Principal Components Analysis
RT-PCR Real-time reverse transcription - Polymerase chain reaction
PCR Polymerase chain reaction
PLS Partial Least Squares
PLS-DA Partial Least Squares – Discriminant Analysis
POC Point-of-care
R&D Research & Development
RFID Radio-frequency identification
RMSEC Root Mean Square Error of Calibration
RMSECV Root Mean Square Error of Cross Validation
RMSEP Root Mean Square Error of Prediction.
RNA Ribonucleic Acid
ROC Receiver Operator Curves
RSMECV Root Mean Square Error of Cross Validation
RT- PCR Real Time Polymerase Chain Reaction

SARS-CoV Severe Acute Respiratory Syndrome CoV

SIMCA Soft Independent Modeling of Class Analogy

SOP Standard Operational Procedure

WHO World Health Organization

Chapter 1.

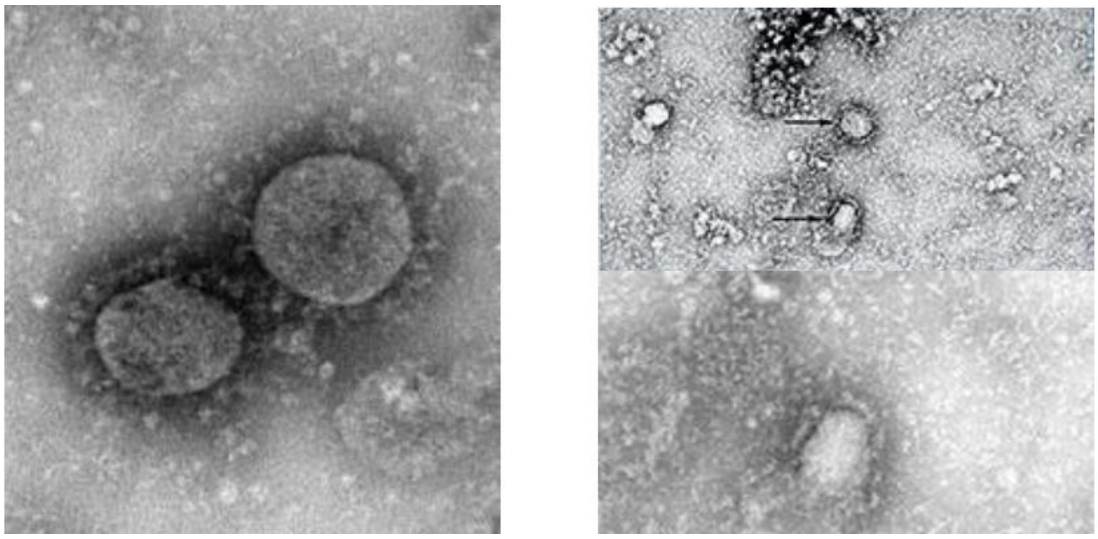
Introduction

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 (order Nidovirales, family Coronaviridae, genus Coronavirus) 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. They have the largest genome found in RNA viruses, with 30,000 nucleotides. Coronaviruses can be mammalian or avian viruses, classified into distinct species by host range, antigenic relationships, and genomic organization. [01] They are currently grouped into four genotypes: alpha-coronavirus and beta-coronavirus: Highly infectious in bats and rats; delta-coronavirus and gamma-coronavirus: Both present in birds. [02] 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. Their envelope possesses roughly spherical or mildly pleomorphic shape with diameter within 60 – 140 nm. Furthermore, the S proteins on the envelope forming a crown shape around the virus, which gives the name "corona". In general terms, the S protein is divided into the S1 domain, responsible for receptor binding and S2 domain, responsible for cell membrane fusion. [03]. Two images of the virus, analyzed through electronic microscopy, can be seen on figure 1.1.



(A) Isolation source: Environmental samples (B) Isolation source: Clinical patients

Figure 1.1 - Environmental and Clinical samples of 2019-nCoV**.** β coronavirus (Source: National Institute for Viral Disease Control and Prevention under Chinese Center for Disease Control and Prevention).

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 [04]. Besides, the pandemic generates unprecedented pressure over national health systems around the world. The economic wreckage caused by the recurrent lockdowns and decreased economic activity in diverse fields, such as tourism, is incalculable. 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. SARS-CoV-2 genome consists of a single strand of positive-sense RNA5 surrounded by a lipid-bilayer membrane studded with transmembrane proteins [05]. Similarly to SARS-CoV viruses, SARS-CoV-2 uses the cell receptor Angiotensin-converting enzyme 2 (ACE2) to connect and invade the host cells. Lung cells show a high expression of ACE2 in alveolar type II cells, making these cells susceptible to the SARS-CoV-2 virus, facilitating the spread of the virus in human populations.

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. According to the severity of symptoms, patients can be classified as mild, severe or critical. Critical patients have severe conditions, such as respiratory failure, septic shock and/or multiple organ dysfunction or failure [03]. The severity and outcome of the disease is highly influenced by the so called Cytokine Release Storm (CRS) [05].

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. Respiratory droplets appear mainly during coughing, sneezing, or talking. Biggest particles (> 5 μm diameter) are limited and generally spread less than 1 m. Droplets originated from COVID19 patients can contaminate the surface of objects. If a health person touches these contaminated surfaces, their hands can contact the nasal cavity, oral cavity or face, leading to the transmission of the disease.

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. The SARS-CoV-2 activates the innate immune system. Macrophages and other innate immune cells capture the virus and release a large number of cytokines and chemokine. Adaptive immunity is also activated by antigen presenting cells (mainly dendritic cells). T and B cells play an antiviral role and directly or indirectly promote the secretion of inflammatory cytokines. 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. [05] In addition to elevated serum cytokines, high concentrations of ferritin are characteristic of the syndrome caused by the CRS. The syndrome leads to an increased risk of vascular hyperpermeability, multiorgan failure, and eventually death when the high cytokine concentrations are unabated over time.

Much effort has been applied worldwide to better understanding of the cytokine storm and it is already known that the virus replicates very quickly during the early stages of the infection. Therefore, it leads to a strong yet delayed proinflammatory response. These pro-inflammatory cytokines and chemokines attract macrophages and neutrophils that also release pro-inflammatory agents. This amplifies the inflammation, in a chain reaction that can lead to acute respiratory distress syndrome and multiple organ failure, which are associated with poor outcomes for the disease. Identifying the distinct cytokine profile and immune signature of COVID-19 has practical implications. It can predict the requirement for hospitalization/intubation and mortality.

1.4 Detection tests for COVID-19

Since the beginning of the COVID-19 pandemic, in March 2019, quick, cheap, easy and reliable tests for the early detection of the disease have been investigating. 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. Among all the known and developed techniques with this purpose, the more worldwide applicable methods are hereby listed and briefly described [06].

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 RT- PCR test. This test is widely used to detect gene expression levels and facilitate the rapid diagnosis of acute respiratory viral infections. A methodology for RT-PCR for detecting SARS-CoV-2 was first developed at the Charité Institute of Virology in Germany and introduced by the WHO on January 13, 2020. It targets SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) gene, as well as the E and N genes. [07]

Sampling methodology generally consists of collecting samples through a swab from the nose and/or throat region. In short, the analytical methodology resides in: Extraction of viral RNA; RT-PCR (mixing of RT-PCR reagents and primers to viral RNA samples previously extracted, performing of RT-PCR; results acquisition). This methodology presents more than 90% of sensitivity and false-positives results are rare.

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. Testing methodology consists in: Collection of blood sample from the finger or swab from the throat/nose; one drop of blood or one drop of prepared sample is placed on the testing strip; adding one drop of buffer solution to the strip. The results are available after 10 to 20 minutes [08]. This test provides a very fast output, even though its high rates of false-positive or false negatives. It´s a cheaper methodology and provides faster results than Molecular Nucleic Acid Amplification Tests.

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. Testing methodology is basically the same as for rapid antigen tests: Collection of sample using a swab from the throat/nose; the swab is inserted in a small tube with buffer solution; one drop of prepared sample is placed on the testing strip; adding two drops of buffer solution to the strip. The results are available after 10 to 20 minutes. Even though subject to high uncertainties regarding sample collection, it allows the tracking of a major amount of population [14].

1.5 Current laboratory techniques diagnosis of COVID-19

1.5.1 Immunoassays

The advantages of immunoassays are: Specificity, sensitivity, rapid turnaround time, convenience, the ease of performance and a relatively low cost.

As drawbacks immunoassays are strictly dependent on the quality of the capture and the detection; requires the transport of samples to a laboratory; execution of the technique is time consuming; It can detect biologically inactive cytokines and their fragments; may return a negative result, if the primary structure of the antibody is altered; is often influenced by "matrix effects," which refers to nonspecific binding to the assay matrix of proteins present in complex biologic fluids that may interfere in the result interpretation. [09]

ELISA (Enzyme-linked immunosorbent assay) is the most commonly used form of immunoassay.

In this technique antigens from the samples are attached to a surface. Then a primary antibody is applied to the same surface, binding to the antigen. This antibody is linked to an enzyme, and during the analysis, unbound antibodies are removed. A secondary antibody conjugated to an enzyme or radioisotope for the detection. It is often used because it is quantitative, easy to use and relatively inexpensive. Besides, possess high sensitivity. As a main drawback for this technique, the ability of detecting only one cytokine in a given time can be highlighted.

In order to overcome this problem, multiplexed systems have been introduced to the technique. They can simultaneously detect multiple cytokine expressions in one assay at the protein levels in any body fluid. They also show high sensitivity, being able to identify concentrations as of pg/ml.

1.5.2 Bioassays

Biological activity of cytokinesis determined in cellular assays. They are mainly used for validation purposes rather than routine cytokine assessments

These assays have as a positive side being highly sensitive.

Nonetheless, they are not specific and neutralization with cytokine-specific antibodies is necessary to confirm measured cytokine identity. They are also time-consuming and labor-intensive, requiring a tissue culture facility. They are also not easily automated and are not appropriate for a large-scale screening of specimens.

Bioassays can be divided into the following categories: Assays that measure proliferation or growth inhibition; assays of cytotoxicity; assays dependent on the induction of a specific cell function; assays estimating the quantity of a protein induced in the target cell.

Cellular functions other than proliferation or death have also been used for measuring biological effects of cytokines. [09]

1.5.2.1 Flow Cytometry

Flow cytometry is a laboratory technique applied to detect and measure physical and chemical characteristics of a population of cells or particles. A sample is suspended in a fluid and injected into the Flow cytometer equipment. The sample is focused on ideally flow one cell through a laser beam, where the light scattered is characteristic. Particles on the sample are often labeled with fluorescent markers so light is absorbed and then emitted in a band of wavelengths.

This technique can be used for the detection of intracellular cytokines. They are detected on blood and body fluids. This technique is a rapid analysis for the identification of cells. Besides, tens of thousands of cells or particles can be quickly examined and the data gathered are processed by a computer.

On the other hand, the complexity of identifying and quantifying intracellular cytokines and related molecules, background induced autofluorescence, the automatic gating of cells and the requirement for negative control for cutoff levels are the drawn backs of this technique [10].

1.5.2.2 Nanoparticle-modified aptamers

Short synthetic nucleotide sequences of DNA or RNA, known as aptamers, can be used to mimic the properties of antibodies. Ultrasensitive densitometric methods for cytokine-detection that use gold nanoparticle-modified aptamers have been developed. In such assays, aptamer pairs bind to the target protein to form a sandwich complex fixed onto a micro-plate. After amplification of a signal of gold nanoparticles by silver-enhancement technology, and following several steps, a micro-plate reader measures the absorbance. So far, a number of aptamers with good specificity and good affinity have been characterized for a host of cytokines such as vascular endothelial growth factor.

1.6 Infrared Spectroscopy

Infrared spectroscopy is a method that applies the infrared bands of the electromagnetic spectrum: From 12500 to 40 cm^{-1}. It is positioned beyond the red limit of the visible spectrum, as represented on the Figure 1.2. The IR region is often referred to as the "optical window" due to its minimal absorption by water compared to other wavelengths. 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.

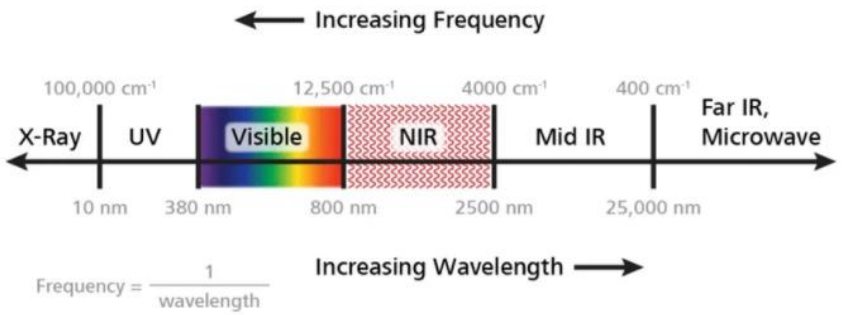


Figure 1.2 - Electromagnetic spectrum. Image credit: Metrohm AG.

Infrared light has been the first kind of non-visible electromagnetic radiation discovered in the year 1800, when the astronomer William Herschel discovered the infrared region of the electromagnetic spectrum. Later, during the years 1950 – 1960, instrumentation working in the IR region was still being used only as an add-on unit to the major optical devices. It happened due to the fact that, among other reasons, the absorption occurring in the IR region of the vast majority of organic and inorganic samples is relatively weak. This changed with the advent of Fourier-transform (FT) optical spectrometers [11]. Eventually, the advantages of the technique, as its wide applicability, non-invasive, non-destructible analysis, uncomplicated instrumentation, low time-to-result and low cost superpassed the drawbacks [11]. Nowadays analytical IR sensors is an active research field with a considerably wide scientific and professional community involved.

The principle of IR spectroscopy is established on the interaction of the chemical bonds of a sample with the radiation of a light source to generate a signature fingerprint in the form of a spectrum. The technique allows to probe a sample in order to acquire quantitative and qualitative information that comes from the interaction between the near-infrared electromagnetic waves and the sample´s elements. It measures overtones and combination tones of molecular vibrations in the infrared range, i.e. stretch vibrations involving hydrogen bonds (e.g. C-H, O-H and N-H). The overtones that different molecule bonds absorb at specific frequencies are characteristic of their structure.

The Infrared spectra can provide characteristic signatures that may be useful to characterize complex biological materials (for example, microorganisms, blood samples, swabs). Functional groups present on these samples – OH, NH and CH – can be a valuable asset in order to allow the identification and information acquisition from the analyzed material. While exact allocation to specific structures is still too complex a task using IR, many spectral features of biological samples can be visualized by applying resolution enhancement techniques, such as derivative or Fourier self-deconvolution techniques [12].

In general, a sensor that uses IR consists of a source of infrared radiation, a detector of the radiation and a dispersive element. The studied sample is irradiated with a broad spectrum of infrared light. The light interacts with the sample and the detector measures its transmittance and absorbance. Transmittance refers to the amount of light that passes completely through the sample and hits the detector. Absorbance is a measurement of light that is absorbed by the sample. The detector measures the light being transmitted through the sample and converts this information into a digital display. A schematic representation of the system is shown on Figure 1.3.

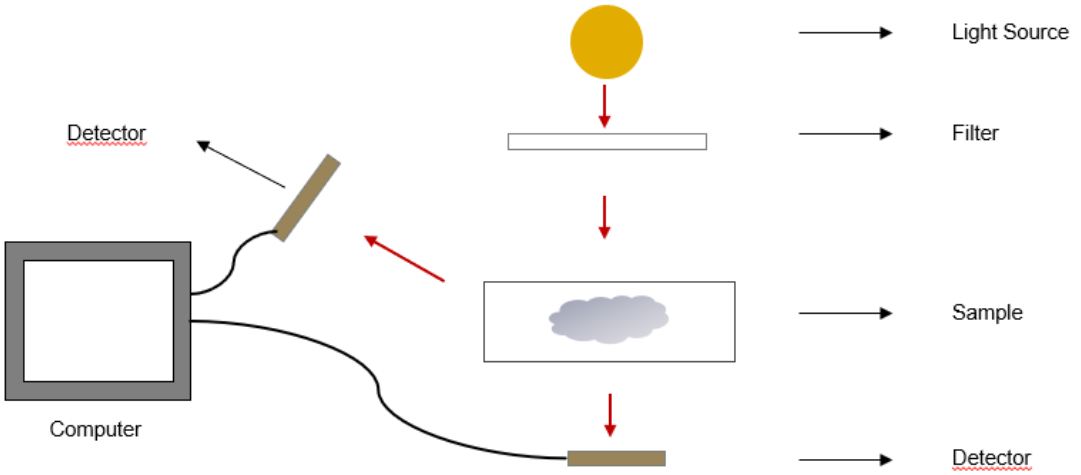


Figure 1.3 - Schematic representation of Near-Infrared Spectroscopy apparatus

The differences in the reflected signal can be correlated to the different concentrations of the chemical groups. Applying chemometric tools, a correlation between reference spectra and reference analysis results can be determined. The resulting calibration model is used for prediction of unknown samples.

The advantages of this techniques are: Analysis are fast, non-destructive, non-invasive, suitable for real-time use, applicable for any molecule containing C-H, N-H, S-H or O-H bonds (including those part of biological molecules), with high penetration of the probing radiation beam and requires minimum sample pre-preparation. It´s also an easy to apply technique, doesn´t present any kind of chemical waste from reagents used in other analysis and is cost-effective. The drawbacks of the method are related to its low sensitivity, low absorption coefficients and requires the development of a multivariate calibration model against a suitable reference method.

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, marking the beginning of on-site capability for this analytical technique.

In IR spectroscopy, the route to the measurement solution depends on a high-quality instrument with the correct configuration, the right sampling, preset calibrations, standard operating procedures, analytical methods, measurements, software platform, data processing, and validation [13].

1.7 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. [13]

FTIR ATR uses a property of total internal reflection resulting in an evanescent wave. A beam of infrared light is passed through the ATR crystal and reflects at least once off the internal surface in contact with the sample. This reflection forms the evanescent wave which extends into the sample. This evanescent effect works as the crystal is made of an optical material with a higher refractive index than the sample being studied. [13]

1.8 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 [15].

Portable instruments have been dramatically improving their capacities on the last 20 years while getting smaller and lighter. This happened due to all big developments in electronics on computer power and the evolution of the companies that manufacture the devices related to R&D, innovation and manufacturing.

Therefore, the spectrometer must be able to read the sample, generate spectra, and process it to generate a reliable and actionable result. Consequently, a combination of spectral libraries or databases is demanded, together with a suitable matching algorithm [15]. A validation protocol for the technique must be also be available.

Portable spectrometers can be divided in different classes: vehicle-mounted and transported, vehicle transportable, person portable, handheld, wearable and smartphone-based [15]. For the applications on COVID-19 pandemics and health assistance environments the category is portable spectrometer applied is "Person portable". The technique studied o the current work, though, can be further adjusted to other devices, such as: handheld and smartphone-based platforms. Related to molecular assays on a field environment, the use of portable spectroscopy devices can be a game-changer: They are simple to operate, give fast output, requires minimum or absolutely no consumables and can provide near-instant feedback. This is crucial in places and situations, as the current COVID-19 pandemics, where resources as laboratory reagents, power and even analysis time are scarce. A summary of the main requirements to a successful portable spectroscopy platform is shown on Figure 1.4.

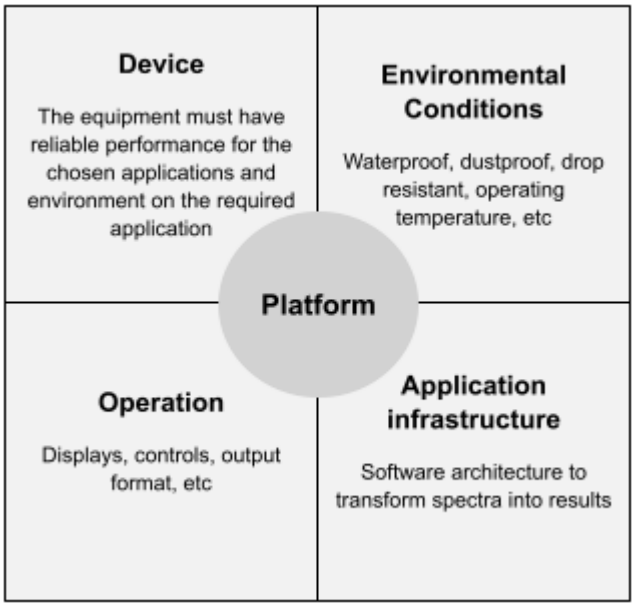


Figure 1.4 - Requirements that must be considered to a successful portable spectroscopy platform.

1.9 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 [13]. A summary on how chemometrics is related to the analysis can be evidenced on Figure 1.5.

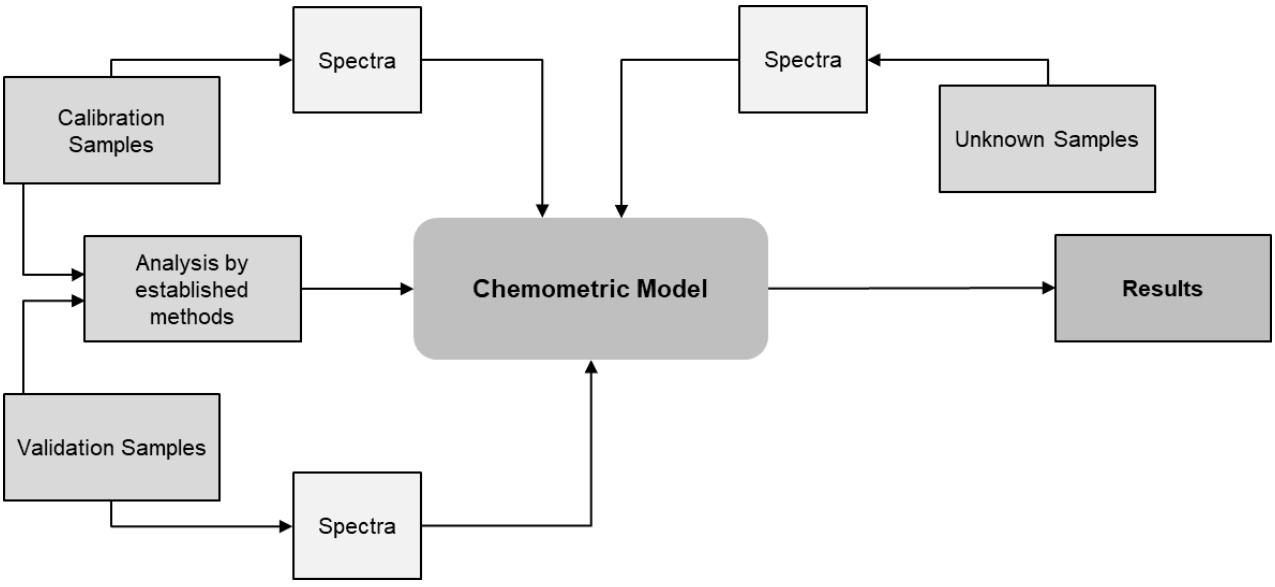


Figure 1.5 - Flowchart for chemometrics [15]

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):

- Measuring and collecting data:

The collection of data and its measurement are the most critical parts of chemometric tools. In general, there are diverse approaches in order to collect chemical data from a system, being the most common ones:

- physical measurements: concentration, temperature, melting point, viscosity, etc.
- spectroscopic measurements using near-infrared, Raman and fluorescence radiation.
- chromatographic measurements through gas chromatography and high-performance liquid chromatography

- Preprocessing data

After data collection, the second step is preprocessing. The main purposes of this step are to correct for signal unwanted features and to identify outliers. There are several established mathematical ways to do this. The following are the most common ones used in chemometrics:

- Mean Centering
- Normalization
- Baseline Correction
- Detrending
- Orthogonal Signal Correction
- Savitsky-Golay Derivatisation
- Multiplicative Signal Correction

The multivariate approach of chemometrics is one of its most important features. Different chemometric analysis methods are needed to extract information from the chemical data. As of Figure 1.5, the result of the analysis is a model, which will be chosen to make future predictions [16].

Multivariate Analysis englobes the following steps, further detailed on this work:

a) Exploratory Data Analysis (EDA): Includes techniques of data mining (e.g.: Cluster Analysis, Principal Component Analysis) which are used for gaining deeper insights into high-volume complex data such as a large set of IR spectra.

b) Regression analysis: It groups the methods used for the prediction/quantification of chemical content (predictive models). The most utilized techniques include Multiple Linear Regression, Artificial Neural Networks, Principal Component Regression and Partial Least Squares Regression.

c) Classification techniques: used for the separation, sorting and grouping of samples in accordance to a selected property. Classification approaches include supervised (e.g., SIMCA, Soft Independent Modeling of Class Analogy; PLS-DA, Partial Least Squares Discriminant Analysis or

Support Vector Machine Classification) and unsupervised approaches (e.g., K-mean and K-median methods, Hierarchical Cluster Analysis or PCA, this time in its classification role).

In a well designed analysis the initial spectrum will be analyzed regarding principal components analysis and a cluster-analysis algorithm. This will show if the acquired spectrum is similar to the calibration samples. Thus, a model will be developed (calibrated) with a set of the collected samples. Lately, the model will be validated, confirming that the model works properly and that the correlations are not random [17].

1.9.1 Principal Component Analysis (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. PCA finds lines, planes and hyper-planes in the dimensional space that approximate, statistically, the data as well as possible in the least squares sense. A line or plane that is the least squares approximation of a set of data points makes the variance of the coordinates on the line or plane as large as possible.

Mathematically, PCA relies on an eigenvector decomposition of the covariance or correlation matrix of the process variables. For a given data matrix X with m rows and n columns, the covariance matrix of X is defined as equation I, if columns of X have been "mean-centered" by subtracting off the original mean of each column. If the columns of X have been adjusted to zero mean and unit variance by dividing each column by its standard deviation), the correlation is shown by equation II.

$$\text{(I)} \qquad \mathbf{COV}\,(\mathbf{X}) = \frac{\mathbf{X^T X}}{\mathbf{m-1}}$$

Equation 1.1 - Equation for covariance matrix for mean-centered columns of X. [18]

$$\text{(II)} \qquad \boldsymbol{X = T_1 p^T 1 + t_2\, pT_2 + \ \ldots + \ t_k pT_k + \ \ldots + t_r pT_r}$$

Equation 1.2 - Equation for covariance matrix for “auto-scaled” columns of X. [18]

The data set should be in standard matrix form, with n rows of samples and p columns of variables. There should be no missing values: every variable should have a value for every sample – it is allowed to have zero values. Generally, the first step is to center the data on the mean of each variable, which is accomplished by subtracting the mean of a variable from all values. This procedure ensures that the cloud of data is centered on the origin of our principal components, but it does not affect the spatial relationships of the data nor the variances along our variables. Graphically, the data set is represented in standard matrix form – N rows of samples (observations) and P columns (variables). To every variable should have a value for every sample, and this value may be zero. The length has been standardized according to a scaling criterion [19].

Following, each observation of the X-matrix is placed in the K-dimensional variable space.

After mean-centering and scaling to a certain unit, the data set will be placed on of the first summary index, the first principal component (PC1). This component is the line in the K-dimensional variable space that best approximates the data in the least squares sense. This line goes through the average point. Each observation is projected onto this line in order to get a coordinate value along the line of the Principal Component. Usually, one principal component is not sufficient to capture the systematic variation of a data set. Hence, a second summary index – a second principal component (PC2) – is calculated. The second PC is also represented by a line in the K-dimensional variable space, which is orthogonal to the first PC. This line also passes through the average point and improves the approximation of the X-data as much as possible.

If two principal components are not enough to represent the variation of the observed data set, more principal components can be added. Each added principal components is always orthogonal to the preceding principal components [19].

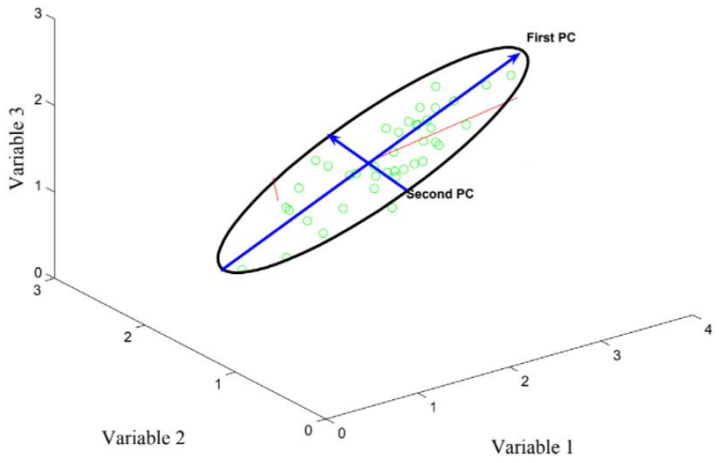


Figure 1.6 - Graphical outlook of Principal Components Analysis [18].

The calculation of principal components can be performed until the total amount of P variables had been calculated (number of PC = number of variables). With this calculation completed, the total variance on all principal components will be equal to the total variance amongst all original variables. Therefore, all information contained in the original data is preserved.

Loadings are interpreted as the coefficients of the linear combination of the initial variables constructed from the principal components. Large loadings values (positive or negative) demonstrate that a particular variable strongly relates to a particular principal component. The sign of a loading indicates whether a variable and a principal component are positively or negatively correlated. Loadings are comparable to the covariances/correlations observed between variables and unit-scaled components [20].

1.9.2 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. As the input (X) most likely does not belong to well defined classes, thresholds between 0 and 1 must be set (applying the Bayes theorem). Receiver Operator Curves (ROC) are part of the PLS-DA. It sets the model´s sensitivity (or its true positive rate) against the specificity – in short, the false positives rate. ROC allows to access how successful (or not) models are on discriminate different samples. It compares the calibration curve of the model (applying the whole data set) to the cross-validation curve providing overfitting information of the model. The area under the curve (AUC) provides information about the model's ability to predict the samples, comparing the calibration and cross validation results (CV).

Mathematically, the PLS-DA weight vector (w) is estimated by maximizing the covariance between inputs and outputs. Then X, load X and load Y scores are sequentially established. A regression coefficient (b) is estimated using the results w, load X and load Y. Following, the first set of PLS-DA and loading components is established. Thenceforth, the residuals X (resX) and y (resy) of the first PLS component become the input data (X) and the output data (y), respectively, to build the second PLS component. This procedure is repeated all needed times to build the desired predicting model. Eventually, the PLS components are constructed using the training samples, and a regression coefficient matrix, B, is also prepared for subsequent prediction purposes.

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.

The perfect class membership which is able to predict values must be ‘+’ or ‘0’ to indicate ‘inside the group’ or outside the group’. However, in practice, the resulting predicted values generally take any values between 0 and 1 instead of an integer. For this reason, several decision rules have been proposed to translate the predicted value into a meaningful class association [21].

1.10 Currently on the realm of COVID-19 diagnosis using sensors

According to Russel M. et al, one alternative in order to accelerate the diagnosis and the probability of different outcomes for COVID-19 is the use of biosensors. A growing trend on the use of biosensors for the detection of diseases has been observed over the past decade. One of the main positive points of the biosensors is the fact they hold an interface with different mobile devices, which facilitates its use and penetration to the market. Related to COVID-19, mobile biosensors able to identify and quantify IL-

6 have been researched. They consist of immunosensor for colorimetric detection using gold nanoprobes, able to detect variations as small as 12.5 pg mL^{-1} with an assay time of 18 min [22].

Another approach is integrating the sensors with a cannula within the bloodstream, enabling real-time and continuous monitoring of IL-6 levels.

Wearables have been also proposed as a platform for fast detection and real-time monitoring of COVID-19. It is based on the intradermal delivery of biocompatible near-infrared (NIR) quantum dots. An array of these nanosensors is delivered using dissolvable microneedles. Fluorescence emission of the quantum dots was fine-tuned so that they would be invisible to the naked eye, but detectable upon illumination with NIR light. The pattern of emitted light generated by the array can be detected with a modified smartphone and evaluated with a machine learning algorithm [23].

Maddali et. al (2020) state that optical can be developed into point-of-care (POC) diagnostics for viral detection and are promising alternatives to detect COVID-19. They also show optical biosensors as an alternative method for virus detection due to their safe, straight-forward use, and cost-effective technology, including eliminating the need for nucleic acid amplification. The race to design a novel biosensor that would optimize COVID-19 tests in terms of cost, testing rate, and sensitivity is underway. In their work, they present different POC diagnosis methods for COVID-19, detailed below. [24]

Abbott™, BioFire™, and Cepheid™ are some of POC diagnostic systems that have received emergency use authorization from the U.S. Food and Drug Administration (FDA). They use nucleic acid amplification followed by fluorescent labeling to detect the presence of SARS-CoV-2 quickly and with accurate results. Canon™ uses loop-mediated isothermal amplification (LAMP), a similar technique to PCR. These companies have produced nucleic acid amplification instruments that are portable and do not require trained operators. This makes an easy platform for SARS-CoV-2 virus detection. [24]

Chen et al. developed a lateral flow immunoassay (LFIA) to test for the presence of anti-SARS-CoV-2 Immunoglobin G (IgG) antibodies in human serum. In this work, a test line was coated with recombinant SARS-CoV-2 nucleocapsid phosphoprotein and a control line with goat anti-rabbit IgG. As human serum was introduced onto the flow assay, human anti-SARS-CoV-2 IgG antibodies conjugated with M-HIgG@L-NPs and attached to the test line material causing a change in the fluorescent signal at the test line. The ratio of the test line to control line fluorescence signal was calculated to determine the concentration of anti-SARS-CoV-2 IgG in the human serum sample. These methods offer an affordable, user friendly and fast test for the presence of anti-human SARS-CoV-2 antibodies. However, the delayed emergence of antibodies could make these methods non-viable for recently infected patients. Also, the biophysical property of the SARS-CoV-2 virus to recognize and bind to angiotensin-converting enzyme 2 (ACE-2) at a high affinity has been exploited. A LFIA can be designed with AuNPs conjugated with SARS-CoV-2 spike antibody, which has shown promise as a field-effect transistor detection method for SARS-CoV-2 to bind to SARS-CoV-2 virus particles. By modifying the test line with recombinant ACE-2 protein, a simple, no-sample-preparation colorimetric method could be prepared to detect for SARS-CoV-2. [25]

There has been extensive research on the optical detection of virus particles by sandwiching them between primary and secondary antibodies or between a peptide sequence and an antibody. Recent studies have provided evidence for antibodies that specifically bind to SARS-CoV-2. Designing and

fabricating optical biosensors based on the receptors of SARS-CoV-2 could be an effective method for COVID-19 testing.

Chapter 2.

Objective

In the face of COVID-19 pandemics, methods for the diagnosis of contaminated patients, healthcare workers and the general public are essential and latent. Such methods must be reliable, fast and, preferably, not be very labor intensive.

In order to support the fulfillment of such requirements, 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.

Chapter 3.

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.

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

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	

Main laboratory equipment is specified on the following items:

3.1.2 Bench Centrifuge

For the separation of plasma and red cells on the blood samples centrifuge Eppendorf® model 5810 R - Rotor A-4-81 was adopted. This equipment combines a compact footprint to a good flexibility on laboratory working environment. The aerosol-tight rotors allow working with potentially hazardous materials.

The rotor allows the use of different tubes, bottles, flasks, and plates. Additionally, the equipment provides a high centrifugation speed up to 20,913 × g (14,000 rpm) for applications in tubes from 0.2 to 750 mL. It is also capable of operating in a temperature range from −9 °C to 40 °C, maintaining constant 4 °C, even at maximum speed. The device presents a function for quick pre-cooling [26]. Illustrative images both from the centrifuge as well from the rotor can be seen on figures 3.1 and 3.2.

Figure 3.1 - Centrifuge Eppendorf® model 5810 R – illustrative image [26]

Figure 3.2 - Centrifuge Eppendorf® Rotor A-4-81 – illustrative image [26]

3.1.3 FTIR ATR sensor

The sensor applied in the current experiments was Agilent® 4300 handheld FTIR with DTGS detector.

Agilent 4300 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. It is also highly accurate mid-IR analysis and requires no sample preparation. This makes this sensor ideal for reading blood samples in a hospital environment during a pandemic situation. In addition, it is a non-destructive tool (NDT) that allows the analysis to be performed without damaging the sample. The sensor is the type Fourier transform infrared (FTIR) spectroscopy.

The core of the Agilent FTIR spectrometer system is a patented Michelson interferometer design mounted on shock-dampening mechanisms to protect the components from shock and vibration in the field.

The equipment is equipped with an integrated. The spectral display is represented with status bars and report formats, which increase the ease of use of the device [27].

The presence of a DTGS detector offers a variety of sampling interfaces (Diffuse Reflectance, External Reflectance, Grazing Angle, Diamond ATR, Ge ATR) and allows an easy transition from one sample type to another, with no alignment or adjustments necessary. A device illustrative image is present on Figure 3.3.

The specifications for the system are:

- Size: 10 x 19 x 35 cm
- Weight: 2.22 kg
- Spectral range DTGS: 4500–650 cm-1
- Resolution: 4–16 cm^{-1}
- Warmup time: 10 min
- Response time: 2 min

Figure 3.3 - Agilent 4300 Handheld FTIR with DTGS detector

Besides, the sensor counts with a sampling interface Diamond ATR. It is ideal for the analysis of solids, liquids, pastes and gels. The interface is comprised of a diamond window, which makes it impervious to corrosion and scratching. ATR is a surface technique and only the top 2–3 µm are analyzed. The diamond is 2 mm in diameter with a 200 µm active area and provides an approximate 2 micron depth of penetration for IR energy at 1700 cm^{-1} [27].

Each sampling interface is designed for maximum throughput without additional alignment of the 4300 Handheld FTIR system. Each interface is also programmed with an RFID chip, which allows the system to immediately recognize which interface is installed and change conditions accordingly.

The instrument is provided with a dedicated sampling probe interface - attenuated total reflectance (ATR) probe. The ATR takes advantage of the physical properties of light when encountering two materials with differences in index of refraction. When coming into contact with a sample having a different index of refraction, IR light creates an evanescent wave that provides a very small and specific depth of penetration into the sample before reflecting back into the detector. This small and consistent pathlength provides the advantage of no sample preparation required to get good measurement.

The key to obtaining good results with an ATR is making good contact between the sample and the ATR crystal. The equipment on this study uses a type IIa synthetic diamond crystal as the interface between the sample and the IR energy. Diamonds are extremely hard as well as chemically resistant [27].

3.1.4 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.

PLS_Toolbox is a collection of different set of chemometric routines. It is an extensive suite of multivariate and machine learning tools for chemometric applications within the MATLAB® computational environment. It contains tools required to explore data and build predictive models. The software allows point-and-click data-discovery tools including PLS, PCA and other multivariate and machine learning methods. Version 8.9.2 was used on the current work in order to analyze obtained results [28].

PLS_Toolbox is organized around four main Graphical User Interfaces (GUI) tools. Each tool is dedicated to a specific step in the data analysis process.

The Browse GUI is the PLS_Toolbox desktop environment. Accessing all loaded data, available analysis methods, DataSet Editor, and other tools is possible. DataSet Editor is the primary data-handling GUI in PLS_Toolbox. It provides a variety of functions and tools for loading, editing, and saving data. Analysis GUI is the core interface with PLS_Toolbox data modeling and exploration functions.

PlotGUI interface is the principal data visualization tool for PLS_Toolbox. It provides a vast number of tools for labeling, manipulating and publishing plots generated using PLS_Toolbox [18].

Figure 3.4 shows the main graphical user interface for Solo®, version 8.9.2.

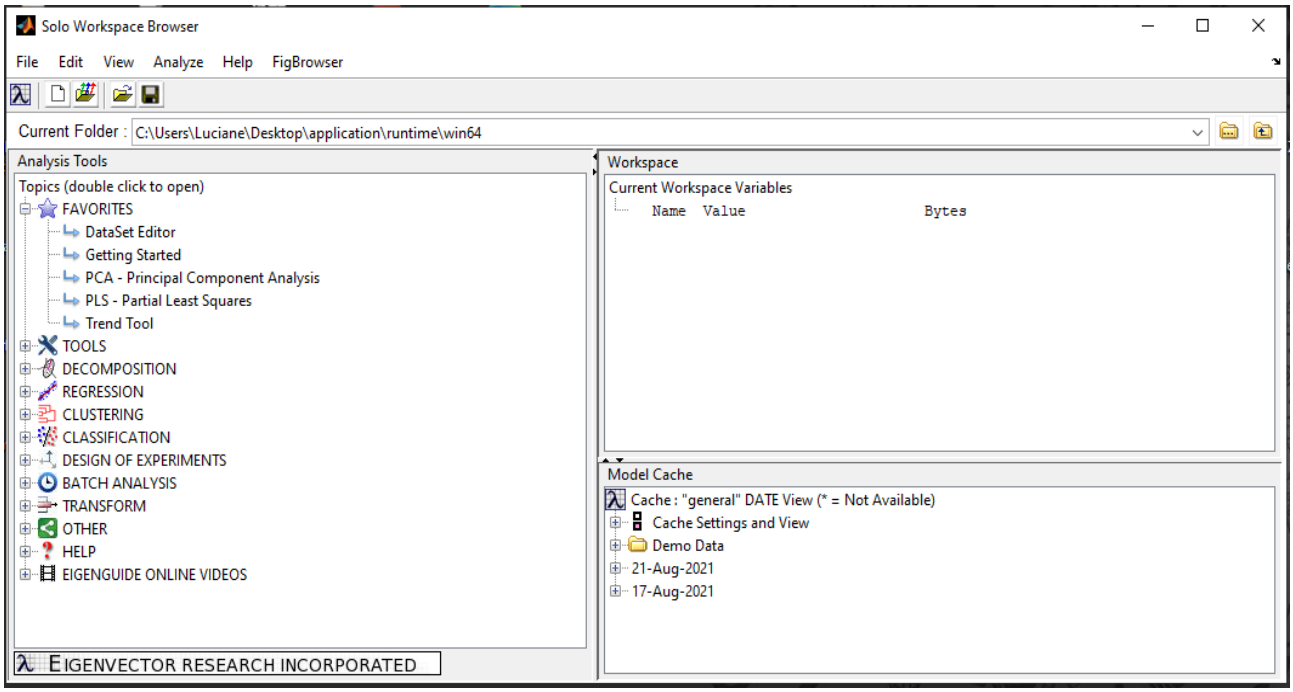


Figure 3.4 - Solo® Software Browse Graphical User Interface. Version 8.9.2

Chapter 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).

These were severe patients of COVID-19, tested positive for PCR and requiring intensive care for both groups. Blood samples were collected after approximately 7 days of hospitalization.

The third group, denominated control group (Group C), consists on people without any symptoms of COVID-19, employees from the University of Lisbon. For these blood samples, IgG tests had been performed. From a total of 21 samples collected on the control group, 10 were diagnosed as positive for SARS-CoV-2 on these testing. Yet not fully clarified (as no PCR for confirmation had been performed), these samples can be false-positives, detecting antibodies present from previous contact with other types of coronaviruses and even influenza viruses. They can also belong to asymptomatic patients. Therefore, to avoid confusion on the results interpretation, these IgG Positive samples had been excluded from the analysis of the entire dataset.

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

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

4.2. Sample Processing

4.2.1 Groups A and B

Once in the laboratories of the Faculty of Pharmacy of University of Lisbon, samples originated from Beatriz Ângelo and Curry Cabral Hospital were submitted to the following procedure: Collected

blood was transferred to EDTA tubes (#368856 K2E BD Vacutainer) and 15 ml of blood were diluted with 15 ml of PBS pH 7.4. Next, 15 ml de Ficoll Paque Plus® (Merck #GE17-1440-02) was added to SepMate (StemCell #85460) and the diluted blood was slowly poured over the Ficoll®, originating two layers. After that, the samples were centrifugated at 700 x g in Eppendorf 5810 R - Rotor A-4-81 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, in order to remove any residues like platelets, cell debris and protein aggregates. Afterwards, plasma samples were collected and stored at -80 °C for subsequent analysis.

4.2.2 Group C

After collection from volunteers at the University of Lisbon, blood was by venipuncture to EDTA blood collection tubes. Afterwards, tubes were well mixed by invertion and spin down at 2000 × g on the centrifuge for 10 min at 4°C. Finally, 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. The tube is then transported to the laboratory at the Faculty of Pharmacy of University of Lisbon, stored in a suitable recipient. 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 Eppendorf 5810 R - Rotor A-4-81. 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 (Previously described in this work, following the steps below:

Cleaning the ATR sample mounting window

Before using the sensor on a new sample, the sampling surface was cleaned using cleaning wipes. Visually inspect the probe tip after cleaning to make sure there is no residue present.

Collect a background spectrum

To ensure accuracy of the measurement, the system is configured in the software to collect a background spectrum before every sample is analyzed. This provides a baseline profile of the system conditions. By collecting a background automatically before each sample measurement, negative effects of changes in the environment can be avoided.

Collect a sample spectrum

Once the background is ready, a small amount of sample was pressured against the ATR crystal. The diamond is held on firmly into the sample. Tilt the instrument to ensure that the diamond surface is

flat to the sample, ensuring a good contact between the sample and the diamond. The software screen the sample and proceed with the analysis. [27]

Samples were read three times in triplicate (three replicate readings for each sample),

The entire set of samples, details and main characteristics are detailed on the table 4.2 below, as a reduced version. These are only illustrative images, complete information will be available on the Annex session of the current work.

Table 4.2 - Samples collected on the current study. Group A, collected at Beatriz Ângelo Hospital, group B at Hospital Curry Cabral and group C is the Control Group.

Samples			Methods data	Diagnosis Data		Date and Time (Readings)	
Sample ID	Group	Sequential #	Treatment with Ficoll	S1 IgG antibodies[a] (positive/negative borderline)	COVID-19 confirmed by PCR	Date	Time
20200522_111_1_2020-05-22T15-22-46	A	1	yes	NA	yes	22-05-2020	15:22
20200522_111_2_2020-05-22T15-24-40	A	2	yes	NA	yes	22-05-2020	15:24
20200522_111_3_2020-05-22T15-26-21	A	3	yes	NA	yes	22-05-2020	15:26
20200522_131_1_2020-05-22T13-33-58	A	4	yes	NA	yes	22-05-2020	13:33
20200522_131_2_2020-05-22T13-35-56	A	5	yes	NA	yes	22-05-2020	13:35
20200522_131_3_2020-05-22T13-38-29	A	6	yes	NA	yes	22-05-2020	13:38
20200522_132_1_2020-05-22T13-42-05	A	7	yes	NA	yes	22-05-2020	13:42
20200522_132_2_2020-05-22T13-43-58	A	8	yes	NA	yes	22-05-2020	13:43
20200522_132_2_2020-05-22T13-47-00	A	9	yes	NA	yes	22-05-2020	13:47
20200522_133_1_2020-05-22T13-50-22	A	10	yes	NA	yes	22-05-2020	13:50
20200522_133_2_2020-05-22T13-52-47	A	11	yes	NA	yes	22-05-2020	13:52
20200522_133_3_2020-05-22T13-54-51	A	12	yes	NA	yes	22-05-2020	13:54
20200522_136_1_2020-05-22T14-05-29	A	13	yes	NA	yes	22-05-2020	14:05
20200522_136_1_2020-05-22T14-14-02	A	14	yes	NA	yes	22-05-2020	14:14
20200522_136_2_2020-05-22T14-08-25	A	15	yes	NA	yes	22-05-2020	14:08
20200522_136_2_2020-05-22T14-15-53	A	16	yes	NA	yes	22-05-2020	14:15
20200522_136_3_2020-05-22T14-10-11	A	17	yes	NA	yes	22-05-2020	14:10
20200522_136_3_2020-05-22T14-17-43	A	18	yes	NA	yes	22-05-2020	14:17
20200522_137_1_2020-05-22T14-12-14	A	19	yes	NA	yes	22-05-2020	14:12
20200522_137_2_2020-05-22T14-19-48	A	20	yes	NA	yes	22-05-2020	14:19
20200522_137_3_2020-05-22T14-21-36	A	21	yes	NA	yes	22-05-2020	14:21
20200522_141_1_2020-05-22T14-43-10	A	22	yes	NA	yes	22-05-2020	14:43
20200522_141_2_2020-05-22T14-44-45	A	23	yes	NA	yes	22-05-2020	14:44
20200522_141_2_2020-05-22T14-46-15	A	24	yes	NA	yes	22-05-2020	14:46
20200522_143_1_2020-05-22T14-48-00	A	25	yes	NA	yes	22-05-2020	14:48
20200522_143_2_2020-05-22T14-49-48	A	26	yes	NA	yes	22-05-2020	14:49
20200522_143_3_2020-05-22T14-51-41	A	27	yes	NA	yes	22-05-2020	14:51
20200522_PC114_1_2020-05-22T15-33-24	C	28	No	negative	no	22-05-2020	15:33
20200522_PC114_2_2020-05-22T15-35-07	C	29	No	negative	no	22-05-2020	15:35
20200522_PC114_3_2020-05-22T15-36-55	C	30	No	negative	no	22-05-2020	15:36
20200522_PC115_1_2020-05-22T15-52-07	C	31	No	negative	no	22-05-2020	15:52
20200522_PC115_2_2020-05-22T15-53-58	C	32	No	negative	no	22-05-2020	15:53
20200522_PC115_3_2020-05-22T15-55-43	C	33	No	negative	no	22-05-2020	15:55
20200522_PC116_1_2020-05-22T15-57-30	C	34	No	negative	no	22-05-2020	15:57
20200522_PC116_2_2020-05-22T15-59-13	C	35	No	negative	no	22-05-2020	15:59
20200522_PC116_3_2020-05-22T16-01-33	C	36	No	negative	no	22-05-2020	16:01
20200522_PC117_1_2020-05-22T16-03-18	C	37	No	negative	no	22-05-2020	16:03
20200522_PC117_2_2020-05-22T16-04-51	C	38	No	negative	no	22-05-2020	16:04
20200522_PC117_3_2020-05-22T16-06-30	C	39	No	negative	no	22-05-2020	16:06
20200522_PC124_1_2020-05-22T16-26-45	C	40	No	negative	no	22-05-2020	16:26
20200522_PC124_2_2020-05-22T16-28-34	C	41	No	negative	no	22-05-2020	16:28
20200522_PC124_3_2020-05-22T16-30-28	C	42	No	negative	no	22-05-2020	16:30
20200522_PC125_1_2020-05-22T16-32-21	C	43	No	negative	no	22-05-2020	16:32
20200522_PC125_1_2020-05-22T16-33-58	C	44	No	negative	no	22-05-2020	16:33
20200522_PC125_2_2020-05-22T16-35-35	C	45	No	negative	no	22-05-2020	16:35
20200522_PC127_1_2020-05-22T16-48-22	C	46	No	negative	no	22-05-2020	16:48
20200522_PC127_2_2020-05-22T16-50-30	C	47	No	negative	no	22-05-2020	16:50
20200522_PC127_3_2020-05-22T16-52-22	C	48	No	negative	no	22-05-2020	16:52
20200527_PC101_1_2020-05-27T20-33-12	C	49	No	positive	no	27-05-2020	20:33
20200527_PC101_2_2020-05-27T20-34-35	C	50	No	positive	no	27-05-2020	20:34
20200527_PC101_3_2020-05-27T20-36-01	C	51	No	positive	no	27-05-2020	20:36
20200527_PC120_1_2020-05-27T19-07-19	C	52	No	positive	no	27-05-2020	19:07
20200527_PC120_2_2020-05-27T19-09-24	C	53	No	positive	no	27-05-2020	19:09
20200527_PC120_3_2020-05-27T19-11-21	C	54	No	positive	no	27-05-2020	19:11
20200527_PC122_1_2020-05-27T19-18-16	C	55	No	positive	no	27-05-2020	19:20
20200527_PC122_2_2020-05-27T19-20-01	C	56	No	positive	no	27-05-2020	19:21
20200527_PC122_3_2020-05-27T19-21-53	C	57	No	positive	no	27-05-2020	19:18
20200527_PC31_1_2020-05-27T20-28-37	C	58	No	positive	no	27-05-2020	20:28
20200527_PC31_2_2020-05-27T20-30-05	C	59	No	positive	no	27-05-2020	20:30
20200527_PC31_3_2020-05-27T20-31-42	C	60	No	positive	no	27-05-2020	20:31
20200527_PC35_1_2020-05-27T18-56-54	C	61	No	positive	no	27-05-2020	18:56
20200527_PC35_2_2020-05-27T18-58-40	C	62	No	positive	no	27-05-2020	18:58
20200527_PC35_3_2020-05-27T19-00-20	C	63	No	positive	no	27-05-2020	19:00

Samples			Methods data	Diagnosis Data		Date and Time (Readings)	
Sample ID	Group	Sequential #	Treatment with Ficoll	S1 IgG antibodies[a] (positive/negative borderline)	COVID-19 confirmed by PCR	Date	Time
20200527_PC38_1_2020-05-27T19-13-05	C	64	No	positive	no	27-05-2020	19:13
20200527_PC38_2_2020-05-27T19-14-41	C	65	No	positive	no	27-05-2020	19:14
20200527_PC38_3_2020-05-27T19-16-14	C	66	No	positive	no	27-05-2020	19:16
20200527_PC43_1_2020-05-27T19-49-16	C	67	No	negative	no	27-05-2020	19:49
20200527_PC43_2_2020-05-27T19-50-45	C	68	No	negative	no	27-05-2020	19:50
20200527_PC43_3_2020-05-27T19-52-20	C	69	No	negative	no	27-05-2020	19:52
20200527_PC44_1_2020-05-27T20-19-03	C	70	No	negative	no	27-05-2020	20:19
20200527_PC44_2_2020-05-27T20-20-31	C	71	No	negative	no	27-05-2020	20:20
20200527_PC44_3_2020-05-27T20-21-59	C	72	No	negative	no	27-05-2020	20:21
20200527_PC47_1_2020-05-27T20-44-35	C	73	No	negative	no	27-05-2020	20:44
20200527_PC47_2_2020-05-27T20-46-19	C	74	No	negative	no	27-05-2020	20:46
20200527_PC47_3_2020-05-27T20-47-54	C	75	No	negative	no	27-05-2020	20:47
20200527_PC72_1_2020-05-27T19-02-13	C	76	No	positive	no	27-05-2020	19:02
20200527_PC72_2_2020-05-27T19-03-49	C	77	No	positive	no	27-05-2020	19:03
20200527_PC72_3_2020-05-27T19-05-31	C	78	No	positive	no	27-05-2020	19:05
20200527_PC73_1_2020-05-27T19-38-49	C	79	No	positive	no	27-05-2020	19:38
20200527_PC73_2_2020-05-27T19-40-37	C	80	No	positive	no	27-05-2020	19:40
20200527_PC73_3_2020-05-27T19-42-14	C	81	No	positive	no	27-05-2020	19:42
20200527_PC81_1_2020-05-27T19-54-07	C	82	No	positive	no	27-05-2020	19:54
20200527_PC81_2_2020-05-27T19-55-54	C	83	No	positive	no	27-05-2020	19:55
20200527_PC81_3_2020-05-27T19-57-35	C	84	No	positive	no	27-05-2020	19:57
20200527_PC84_1_2020-05-27T19-59-40	C	85	No	negative	no	27-05-2020	19:59
20200527_PC84_2_2020-05-27T20-01-15	C	86	No	negative	no	27-05-2020	20:01
20200527_PC84_3_2020-05-27T20-02-43	C	87	No	negative	no	27-05-2020	20:02
20200527_PC96_1_2020-05-27T20-39-21	C	88	No	positive	no	27-05-2020	20:39
20200527_PC96_2_2020-05-27T20-40-55	C	89	No	positive	no	27-05-2020	20:40
20200527_PC96_3_2020-05-27T20-42-24	C	90	No	positive	no	27-05-2020	20:42
20200528_Y001_1_2020-05-28T17-46-33	B	91	Yes	positive	yes	28-05-2020	17:46
20200528_Y001_2_2020-05-28T17-48-15	B	92	Yes	positive	yes	28-05-2020	17:48
20200528_Y001_3_2020-05-28T17-49-50	B	93	Yes	positive	yes	28-05-2020	17:49
20200528_Y002_1_2020-05-28T17-51-34	B	94	Yes	positive	yes	28-05-2020	17:51
20200528_Y002_2_2020-05-28T17-53-18	B	95	Yes	positive	yes	28-05-2020	17:53
20200528_Y002_3_2020-05-28T17-54-53	B	96	Yes	positive	yes	28-05-2020	17:54
20200528_Y003_1_2020-05-28T18-01-04	B	97	Yes	positive	yes	28-05-2020	18:01
20200528_Y003_2_2020-05-28T18-02-37	B	98	Yes	positive	yes	28-05-2020	18:02
20200528_Y003_3_2020-05-28T18-04-07	B	99	Yes	positive	yes	28-05-2020	18:04
20200528_Y004_1_2020-05-28T17-56-32	B	100	Yes	positive	yes	28-05-2020	17:56
20200528_Y004_2_2020-05-28T17-58-04	B	101	Yes	positive	yes	28-05-2020	17:58
20200528_Y004_3_2020-05-28T17-59-32	B	102	Yes	positive	yes	28-05-2020	17:59
20200528_Y005_1_2020-05-28T18-07-27	B	103	Yes	Negative	yes	28-05-2020	18:07
20200528_Y005_2_2020-05-28T18-09-00	B	104	Yes	Negative	yes	28-05-2020	18:09
20200528_Y005_3_2020-05-28T18-10-28	B	105	Yes	Negative	yes	28-05-2020	18:10
20200528_Y006_1_2020-05-28T18-18-29	B	106	Yes	positive	yes	28-05-2020	18:18
20200528_Y006_2_2020-05-28T18-20-25	B	107	Yes	positive	yes	28-05-2020	18:20
20200528_Y006_3_2020-05-28T18-22-07	B	108	Yes	positive	yes	28-05-2020	18:22
20200528_Y007_1_2020-05-28T18-24-07	B	109	Yes	Negative	yes	28-05-2020	18:24
20200528_Y007_2_2020-05-28T18-25-36	B	110	Yes	Negative	yes	28-05-2020	18:25
20200528_Y007_3_2020-05-28T18-27-15	B	111	Yes	Negative	yes	28-05-2020	18:27
20200528_Y008_1_2020-05-28T18-28-52	B	112	Yes	positive	yes	28-05-2020	18:28
20200528_Y008_2_2020-05-28T18-31-37	B	113	Yes	positive	yes	28-05-2020	18:31
20200528_Y008_3_2020-05-28T18-33-03	B	114	Yes	positive	yes	28-05-2020	18:33
20200528_Y009_1_2020-05-28T18-41-58	B	115	Yes	positive	yes	28-05-2020	18:41
20200528_Y009_2_2020-05-28T18-43-31	B	116	Yes	positive	yes	28-05-2020	18:43
20200528_Y009_3_2020-05-28T18-45-26	B	117	Yes	positive	yes	28-05-2020	18:45
20200528_Y010_1_2020-05-28T18-47-33	B	118	Yes	positive	yes	28-05-2020	18:47
20200528_Y010_2_2020-05-28T18-49-10	B	119	Yes	positive	yes	28-05-2020	18:49
20200528_Y010_3_2020-05-28T18-50-55	B	120	Yes	positive	yes	28-05-2020	18:50
20200528_Y011_1_2020-05-28T18-52-33	B	121	Yes	positive	yes	28-05-2020	18:52
20200528_Y011_2_2020-05-28T18-54-13	B	122	Yes	positive	yes	28-05-2020	18:54
20200528_Y011_3_2020-05-28T18-55-46	B	123	Yes	positive	yes	28-05-2020	18:55
20200528_Y013_1_2020-05-28T19-04-38	B	124	Yes	positive	yes	28-05-2020	19:04
20200528_Y013_2_2020-05-28T19-06-37	B	125	Yes	positive	yes	28-05-2020	19:06
20200528_Y013_3_2020-05-28T19-08-19	B	126	Yes	positive	yes	28-05-2020	19:08
20200528_Y014_1_2020-05-28T18-59-54	B	127	Yes	positive	yes	28-05-2020	18:59
20200528_Y014_2_2020-05-28T19-01-21	B	128	Yes	positive	yes	28-05-2020	19:01
20200528_Y014_3_2020-05-28T19-02-56	B	129	Yes	positive	yes	28-05-2020	19:02

a) Indirect ELISA

4.4 Analytical Method

In order to extract information and conclusions from the spectra obtained on sample readings a multivariate Data Analysis had been performed to the data set.

Initially, a Principal Component Analysis, that provides data compression and information extraction. It allows the finding of variables, or factors, that describe major trends in the data. Afterwards, aiming to predict the separation of the samples, the data were submitted to a Partial Least Squares Discriminant Analysis. As Cross-validation methods, which access the optimal complexity of the model and estimate the performance of the model when submitted to unknown data, two strategies were applied:

Venetian Blinds: In this approach, each test set is determined by selecting every nth object in the data set, starting at objects numbered 1 through n.

Contiguous Blocks: Each test set is determined by selecting contiguous blocks of three objects, in this case, in the data set, starting at object number 1. This cross-validation aims to eliminate the deleterious influence of replicates (3 readings of each sample).

An external validation of the model, testing it against a set of unknown samples, had been done to test its predictability.

Chapter 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 [12]. The description of each functional group spectral characteristics is detailed thereafter. Even though figure 5.1 refers to the spectra obtained from microbiology material [12], specifically from *Staphylococcus aureus*, 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.

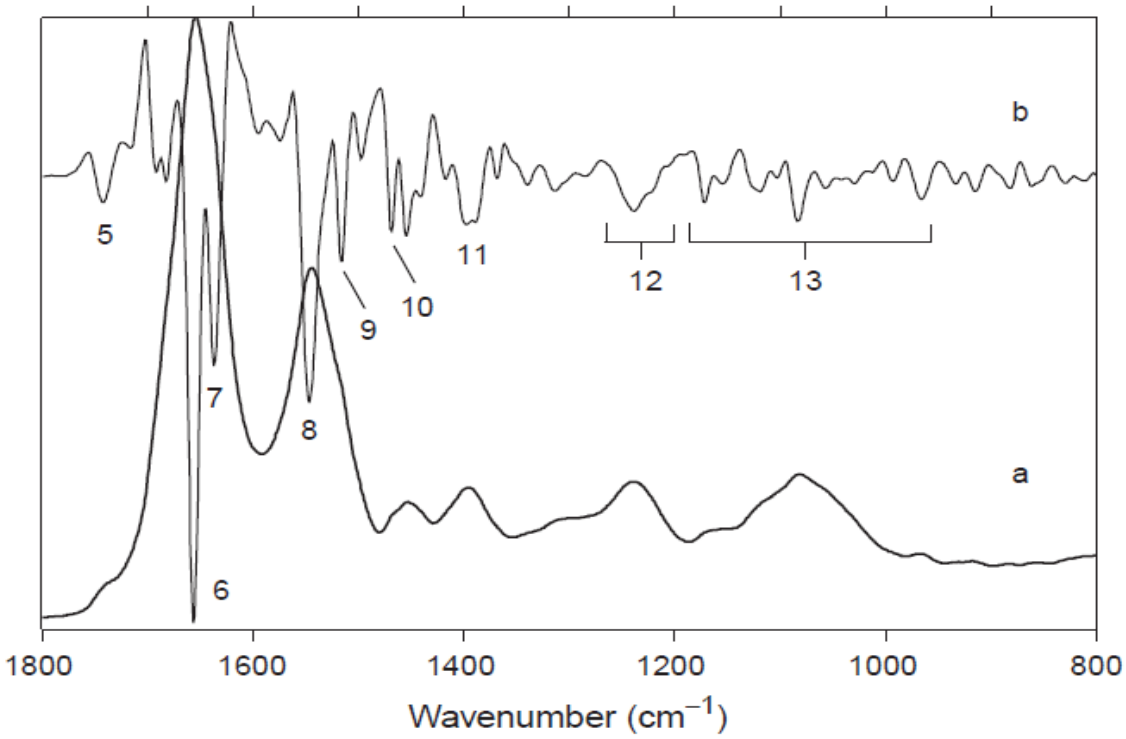


Figure 5.1 - Tentative band assignment of some bands frequently observed in bacterial MIR spectra (*Staphylococcus aureus* (strain SG 511)). (a) Original absorbance spectrum; (b) second derivative. (5) =C=O stretching; (6) amide I of α-helical structures; (7) amide I of β-sheet structures; (8) amide II; (9) tyrosine ring vibration band; (10) CH_2 bending; (11) -COO- symmetric stretching; (12) PO_2 asymmetric stretching; (13) spectral range dominated by complex ring vibrations of carbohydrates, P=O=P stretching, C-O=P stretching, PO_2 symmetric stretching. [12]

a) The region between 4000 and 3100 cm^{-1} is influenced by large spectral features resulting from -OH groups (~3400 cm^{-1}) and N-H (amide A~3300 cm^{-1} and amide B ~3030 cm^{-1}).

b) The region between 3100 and 2800 cm^{-1} reflects the C-H stretching vibrations of $-CH_3$ and $=CH_2$ functional groups. This region is dominated by the spectral characteristics of fatty acid chains and by some amino acid side-chain vibrations.
c) From the region between 1470 and 1350 cm^{-1}, various deformation influence of amino acids molecules are found. A weak band near 3015 cm^{-1} can be rarely observed, as a result the =C-H double bond of unsaturated fatty acid chains.
d) The region between 1800 and 1500 cm^{-1} is dominated by the bands that represent amide I and amide II – strongly representing the presence of proteins at the sample.
e) Bands near 1740 cm^{-1} result from =C=O stretching vibrations of the ester functional groups in lipids. Absorptions Also this spectral domain show absorption of nucleic acids due to =C=O, =C=N, =C=C= stretching of the DNA or RNA heterocyclic base structures.
f) Bands related to amino acid side-chain vibrations, occur weakly near 1498 cm^{-1} (phenylalanine), 1516 cm^{-1} (tyrosine) and between 1585 and 1570 cm^{-1} (aspartate and glutamate carboxylate stretching).
g) Between 1300 and 1500 cm^{-1} absorption profiles are observed predominantly from CH_2 and CH_3 modes of lipids and proteins. Also around 1400 cm^{-1}, a characteristic band may be attributed to the symmetric stretching vibrations of -COO- of amino acid side chains or free fatty acids.
h) Stretching vibrations around 1230 cm^{-1} superimposed bands typical of different =P=O double bond asymmetric stretching vibrations of phosphodiester, free phosphate and monoester phosphate functional groups are observed. The weak band that is visible near 1220 cm^{-1}, most probably due to the phosphodiester functional groups of DNA/RNA polysaccharide backbone structures.
i) The spectral region between 1200 and 900 cm^{-1} is generally dominated by the symmetric stretching vibration of PO_2^- groups in nucleic acids. It also exhibits a sequence of peaks mainly due to C-O-C and C-O-P stretching vibrations of various oligo and polysaccharides. [12]

5.2 General

From all 43 samples collected from all patients (as previously described) all infrared data were acquired under the same conditions:

- Scan region: 600 to 4000 cm^{-1}.
- Resolution: 2 cm^{-1}.
- Number of scans: 64.

Each sample was measured three times, obtaining three spectra of the same sample (Reading in triplicates). Also, background collection had been performed, through an empty ATR crystal, before each measurement.

The spectra obtained from all readings can be seen on Figure 5.2.

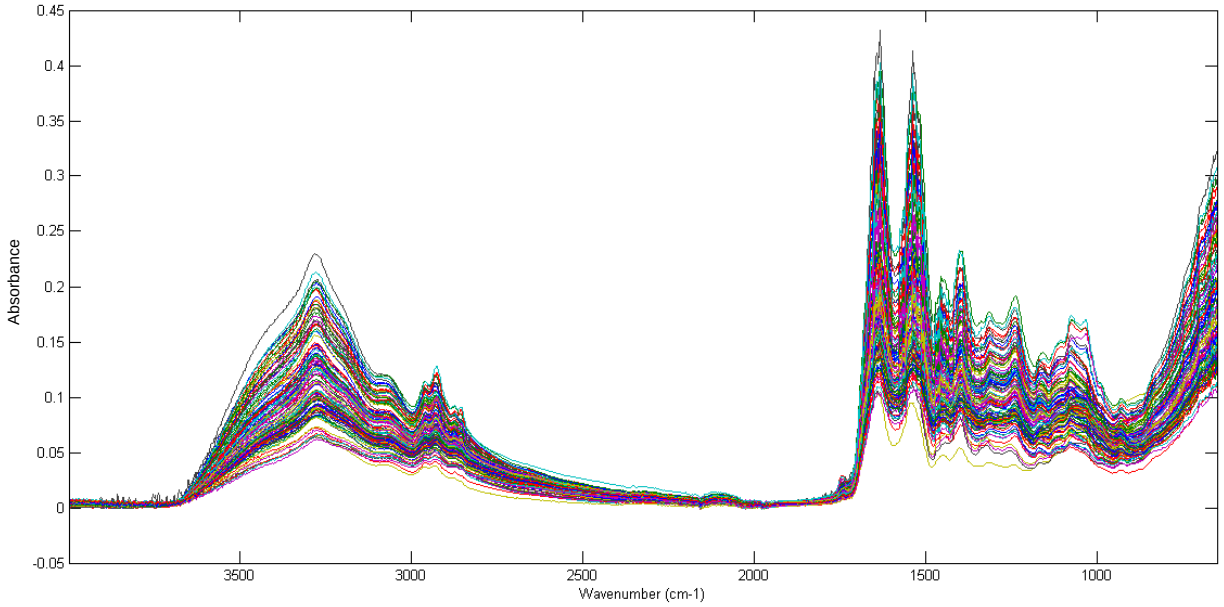


Figure 5.2 - Raw FTIR-ATR spectra for the entire set of collected samples. Wavenumbers from 4000cm^{-1} to 600 cm^{-1}

Aiming to check whether there is a significant difference between the raw spectra of a positive and a negative sample, the spectra of samples 20200527_PC47_1_2020-05-27T20-44-35 (negative) and 20200528_Y001_1_2020-05-28T17-46-33 (PCR positive) were plotted simultaneously. Results are shown in Figure 5.3, demonstrating that, at least in untreated spectra, there is no significant difference between a positive and a negative sample.

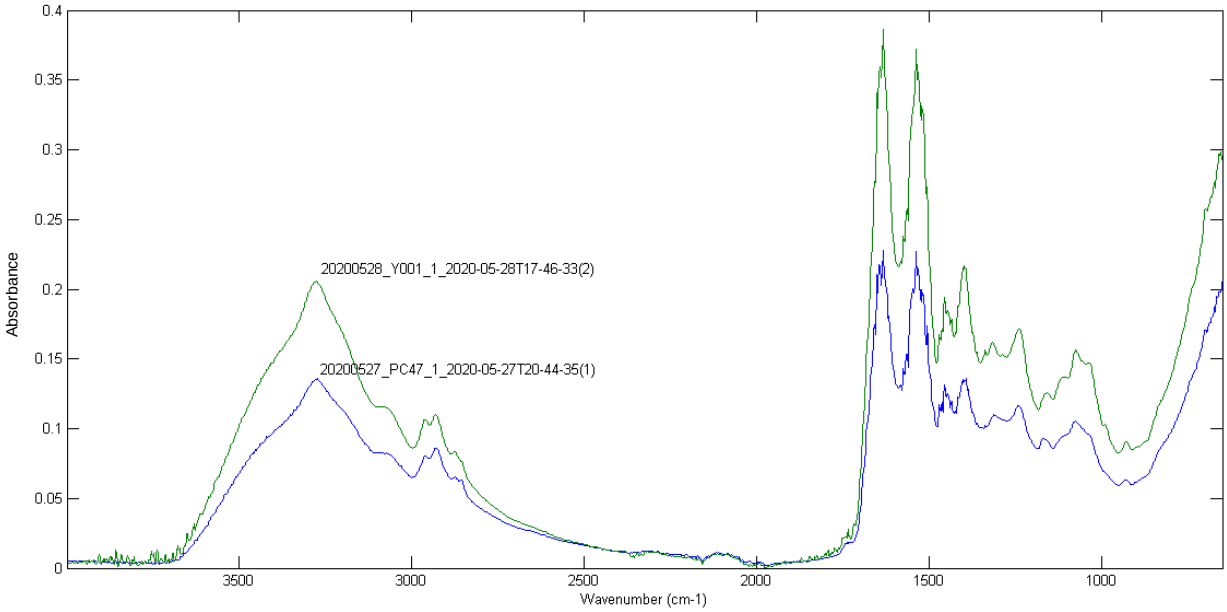


Figure 5.3 - Raw FTIR-ATR spectra of a positive (green curve) and a negative (blue curve) samples for COVID-19. Wavenumbers from 4000cm^{-1} to 600 cm^{-1}

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. This is the case of Ficoll®, a hydrophilic polysaccharide used in this study to separate the different components of blood.

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®.

These differences on the treatment of samples can cause a disturbance on the interpretation of the results, once, instead of capturing if the sample is positive/negative, the technique might capture the presence/absence of the reagent.

In order to avoid any misinterpretations and to spot the influence of the reagent on the spectra, a plot of one positive sample (20200528_Y001_1_2020-05-28T17-46-33), one negative (20200527_PC47_1_2020-05-27T20-44-35) and Ficoll® (ficoll_T30 - dried during 30 minutes) was performed, as demonstrated on figure 5.4.

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.

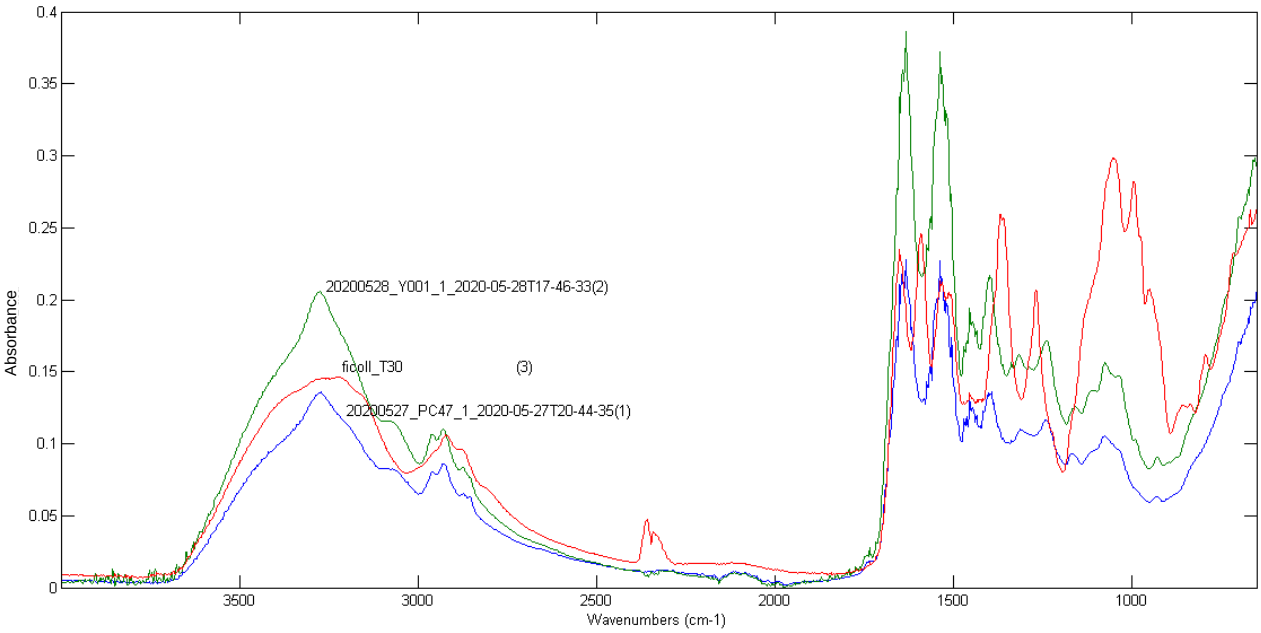


Figure 5.4 - Raw FTIR-ATR spectra of a positive (green curve), a negative (blue curve) samples for COVID-19 and Ficoll® (red curve). Wavenumbers from 4000cm^{-1} to 600 cm^{-1}

As described on Methods item, the samples had been dried during an optimal time of 30 minutes to reduce the influence of the free water on the readings. Regardless, it is possible to observe a very strong disturbance on the spectra, on wavenumbers from 4000 cm^{-1} to 2000 cm^{-1}. This results from a rather broad spectral feature resulting from -OH groups stretching over the entire wavenumber range, especially at approximately 3400 cm^{-1}, reflecting the presence of intramolecular residual water, bonded to proteins and other molecules present in plasma samples.

Wavenumbers from 4000 cm^{-1} to 2000 cm^{-1} were eliminated from the study, reducing the influence of intramolecular water on the spectra readings. Therefore, this procedure also minimizes the influence

of C-H, -CH_3 and =CH_2 functional groups (3100 and 2800 cm^{-1}), generally present on fatty acid chains that form the phospholipidic membrane of the cells – not elements of interest in the current study and that can be present in a residual form on plasma samples.

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. This region may contain weakly expressed bands arising from aromatic ring vibrations of phenylalanine, tyrosine, tryptophan and the various nucleotides [12].

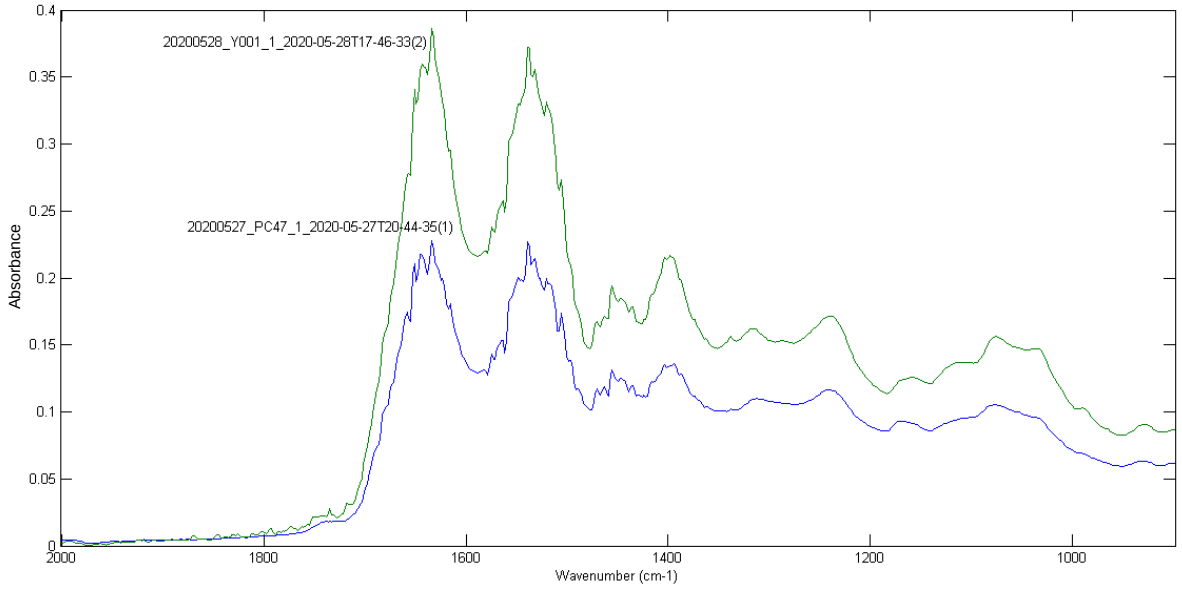


Figure 5.5 - Raw FTIR-ATR spectra of a positive (green curve), a negative (blue curve) samples for COVID-19 - Excluding wavenumbers from 4000 to 2000 cm^{-1} and 900 to 6000 cm^{-1}

5.3 Repeatability Analysis

A repeatability analysis 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)

In this case a PCA score plot considering three samples, each one belonging to one of the different groups (Table 5.1) was performed. Readings were executed in triplicate, each sample being read three times. Sample 20200522_111, is in this study, represents Group A; 20200528_Y009, group B and 20200527_PC35, group C. Samples were read on the 22th, 27th and 28th of May, 2020.

Table 5.1 - Samples for repeatability analysis for each group - A, B and C

Sample #	Sample ID	Date	Time
1	20200522_111_1_2020-05-22T15-22-46	22-05-2020	15:22
2	20200522_111_2_2020-05-22T15-24-40	22-05-2020	15:24
3	20200522_111_3_2020-05-22T15-26-21	22-05-2020	15:26
4	20200527_PC35_1_2020-05-27T18-56-54	27-05-2020	18:56
5	20200527_PC35_2_2020-05-27T18-58-40	27-05-2020	18:58
6	20200527_PC35_3_2020-05-27T10-00-20	27-05-2020	19:00
7	20200528_Y009_1_2020-05-28T18-41-58	28-05-2020	18:41
8	20200528_Y009_2_2020-05-28T18-43-31	28-05-2020	18:43
9	20200528_Y009_3_2020-05-28T18-45-26	28-05-2020	18:45

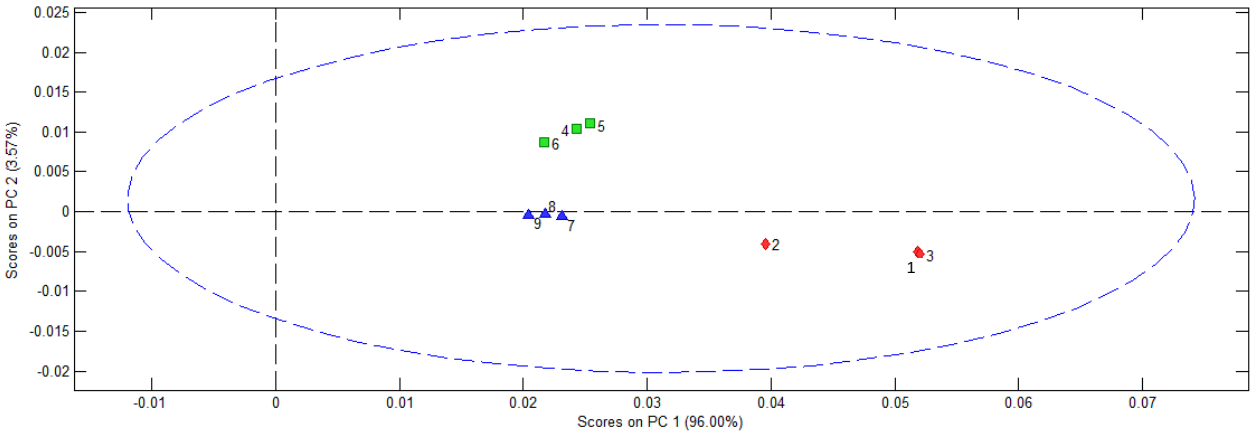


Figure 5.6 - PCA scores plot for Principal Components 1 and 2 showing three different samples read in triplicate: 20200522_111 (1, 2 and 3), 20200527_PC35 (4, 5 and 6) and 20200528_Y009 (7, 8 and 9). Wavelengths: 2000 cm^{-1} to 900 cm^{-1}.

Figure 5.6 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. For group A, it is possible to observe one of the readings (number 2) diverging from the other two readings. It might be caused by a variation on how the sensor is positioned on the microscope slide, once the blood sample is not completely uniform over the slide.

b) 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 (Table 5.2) was performed. This investigation was done to eliminate any possible inter operator variability that can influence the methodology repeatability. Moreover, it also aims to reduce the influence of environmental aspects on the readings, such as air temperature, humidity, movement, electrical supply and device conditions, that can change from day to day. 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. The difference that can be seen between points on Figure 5.7, even though very small, can be a result from a variation on how the sensor is positioned on the microscope slide, once the blood sample is not completely uniform over the slide. Samples were read on the 28th of May, 2020, from 18:24 to 18:45.

Table 5.2 - Samples for repeatability analysis three samples on group B**.**

Sample #	Sample ID	Date	Time
1	20200528_Y007_1_2020-05-28T18-24-07	28-05-2020	18:24
2	20200528_Y007_2_2020-05-28T18-25-36	28-05-2020	18:25
3	20200528_Y007_3_2020-05-28T18-27-15	28-05-2020	18:27
4	20200528_Y008_1_2020-05-28T18-28-52	28-05-2020	18:28
5	20200528_Y008_2_2020-05-28T18-31-37	28-05-2020	18:31
6	20200528_Y008_3_2020-05-28T18-33-03	28-05-2020	18:33
7	20200528_Y009_1_2020-05-28T18-41-58	28-05-2020	18:41
8	20200528_Y009_2_2020-05-28T18-43-31	28-05-2020	18:43
9	20200528_Y009_3_2020-05-28T18-45-26	28-05-2020	18:45

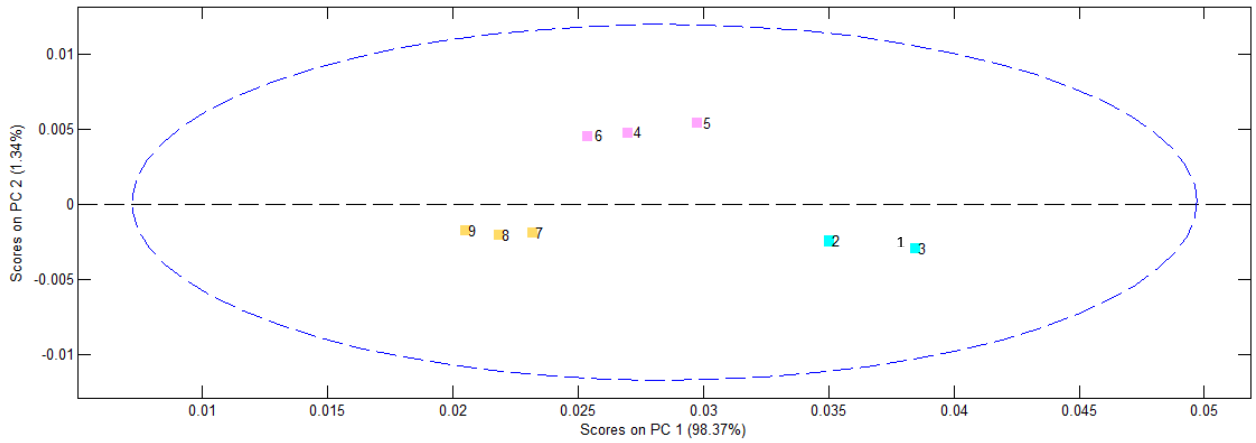


Figure 5.7 - PCA scores plot for Principal Components 1 and 2 showing three samples from group B read in triplicate: 20200528_Y007 (1, 2 and 3), 20200528_Y008 (4, 5 and 6) and 20200528_Y009 (7, 8 and 9). Wavelengths: 2000 cm^{-1} to 900 cm^{-1}.

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

This analysis intends to evaluate if the proposed methodology allows to predict whether a person is contaminated with SARS-CoV-2, considering groups aforementioned A, B and C. In this case, IgG positive had been excluded for group C samples, resulting in a fully negative control group.

Additionally, analysis was also carried out to determine if the methodology is capable of differentiating between IgG positive and negative plasma samples, applying only group C samples. This is further detailed on item 5.5.

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.

The model's ability to separate positives and negative samples on the entire dataset was also confirmed through a PLS-DA.

First of all, a pretreatment of the original spectrum was carried out in order to remove the spectral drift caused by the differences between sample physical properties, improve the resolution and sensitivity, distinguish the overlapping peaks and shoulder peaks, reduce the signal-to-noise ratio, and ensure a good correlation between the spectral data and component information. Chosen pretreatment was 1st derivative, avoiding the effects of baseline drift [28].

Then, a reasonable selection in the principal components for the Principal Component Analysis was performed. This aims to make full use of the effective information of the spectrum signal and avoid overfitting. In this study, the 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, as figure 5.8. Therefore, five Principal Components are the best for this model, as per table 5.3.

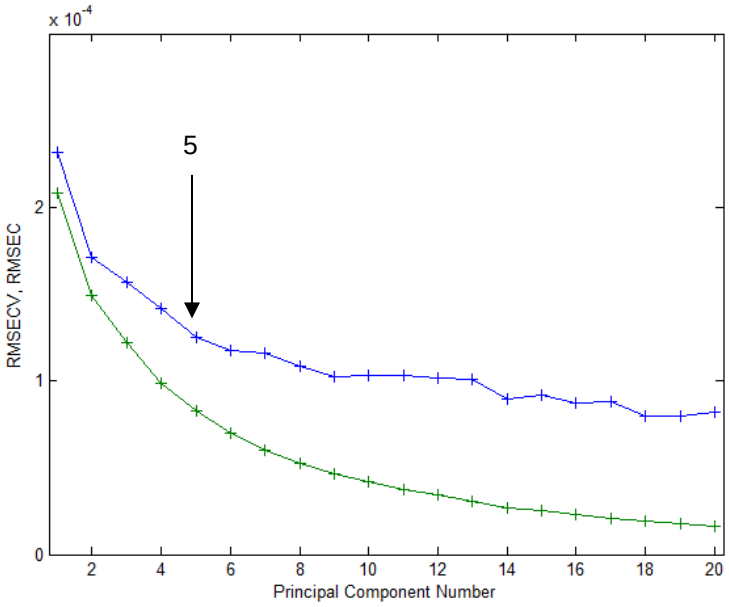


Figure 5.8 - RMSEC and RMSECV for groups A, B and C

Table 5.3 - Principal Components and variance captured for the entire dataset (groups A, B and C)

Principal Component	% Captured Variance	% Total Captured Variance
1	98.96	98.96
2	0.51	99.47
3	0.17	99.64
4	0.12	99.77
5	0.07	99.84

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, having, the second principal component, the biggest influence on the differentiation of the samples.

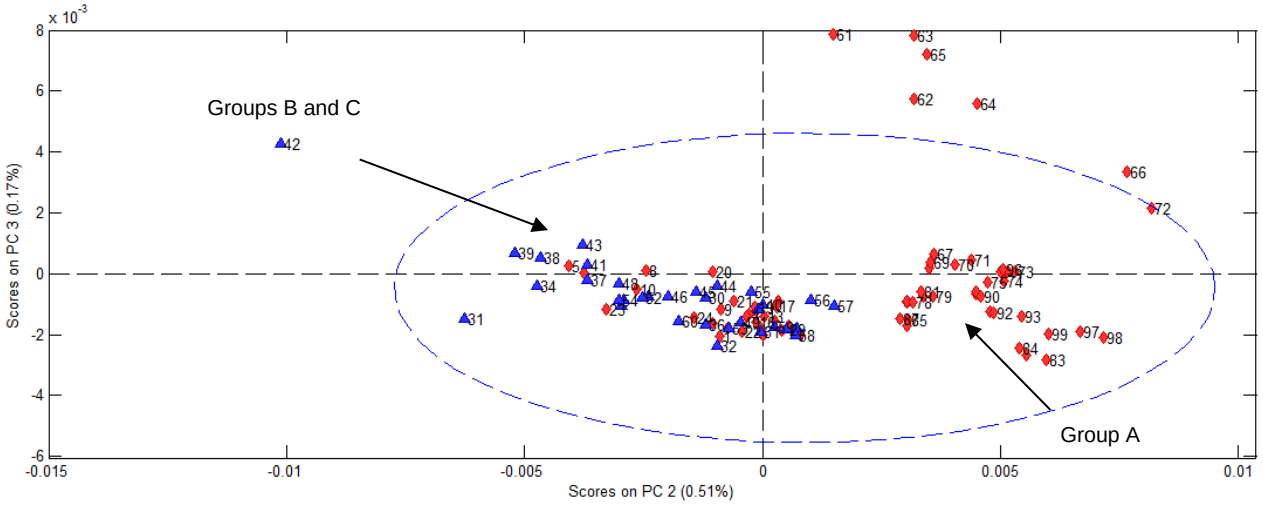


Figure 5.9 - Score plots of group A and B (PCR Positive samples in red) and group C (Control in blue). Principal Components 2 and 3

It is possible to evidence the formation of two clusters on the figure 5.9: 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). These are red and blue markers, by the left side of the score plot.

Once the control group was not treated with the reagent and positive samples, on the contrary, were treated with Ficoll®, this score plot does not evidence if the method is able to differentiate presence or absence of Ficoll®. This inference will be confirmed later.

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 5.10.

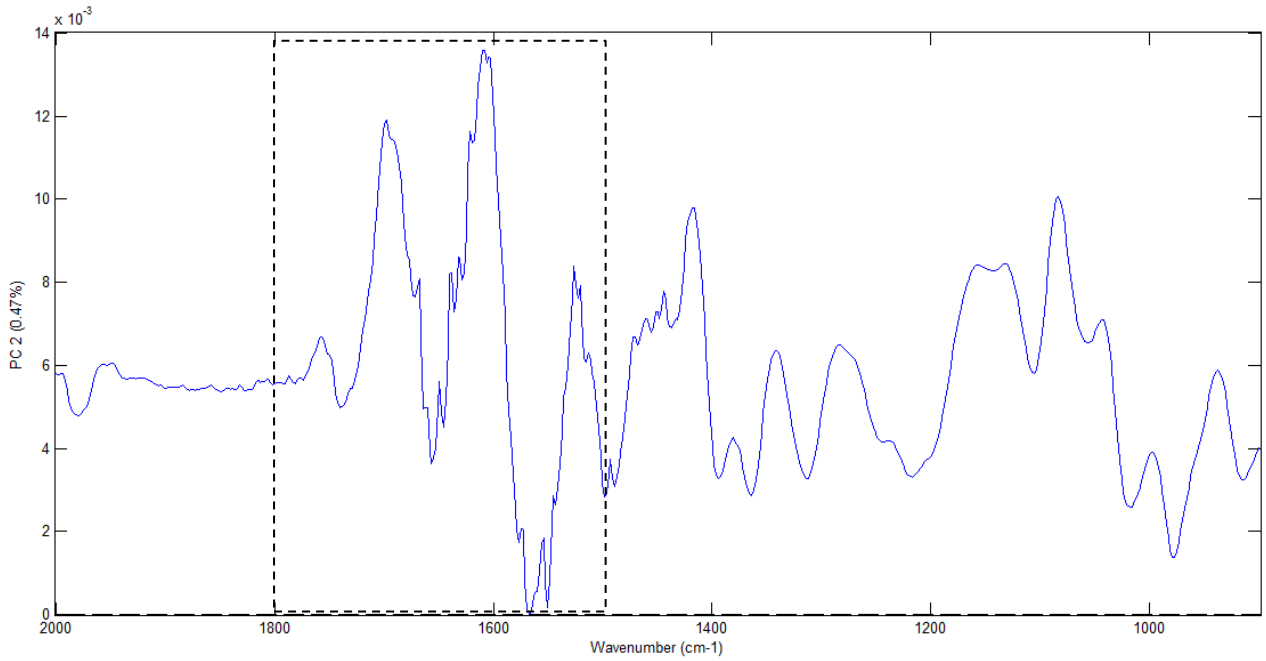


Figure 5.10 - Loadings for Principal Component 2 – Groups A, B and C. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1}

It is possible to notice stretching vibrations on the region between 1800 and 1500 cm^{-1}, which is dominated by amide I and amide II bands, 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.

There is also a high influence of bands near 1740 cm^{-1}, resulting from =C=O stretching vibrations from lipids and also absorption of nucleic acids due to =C=O, =C=N, =C=C= stretching of the DNA or RNA heterocyclic base structures. Deformation influence by amino acid side-chain vibrations, near 1498 cm^{-1} (phenylalanine), 1516 cm^{-1} (tyrosine) and between 1585 and 1570 cm^{-1} (aspartate and glutamate carboxylate stretching) can be also highlighted.

Besides, it is possible to evidence an important deformation influence on wavenumbers from approximately 1200 cm^{-1} to 1000cm^{-1}, which evidences a spectral range dominated by complex ring vibrations of carbohydrates and P=O=P stretching. This can be related to the presence of Ficoll® on the sample, being it a polysaccharide. Although not definitive and inferior to the groups previously mentioned, according to the loadings, the influence of the presence/absence of the reagent in the samples must be considered.

PLS-DA was also applied, as a supervised pattern recognition technique used to classify the unknown samples to positive/negative predefined classes, to confirm or refute the separation previously demonstrated

Two internal cross validation methods had been tested: Venetian Blinds and Contiguous Blocks. To submit the model to external Validation, all dataset had been splited in proportions of 70% of samples for calibration of the model and 30% for external validation, being this group completely independent to the calibration set, resulting in:

- 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. The difference between Cross-Validation and Calibration errors becomes progressively substantial with the inclusion of new latent variables. Therefore, five Latent Variables had been chosen to Venetian Blinds cross-validation, shown in figure 5.11 and three Latent Variables for Contiguous Blocks (figure 5.12).

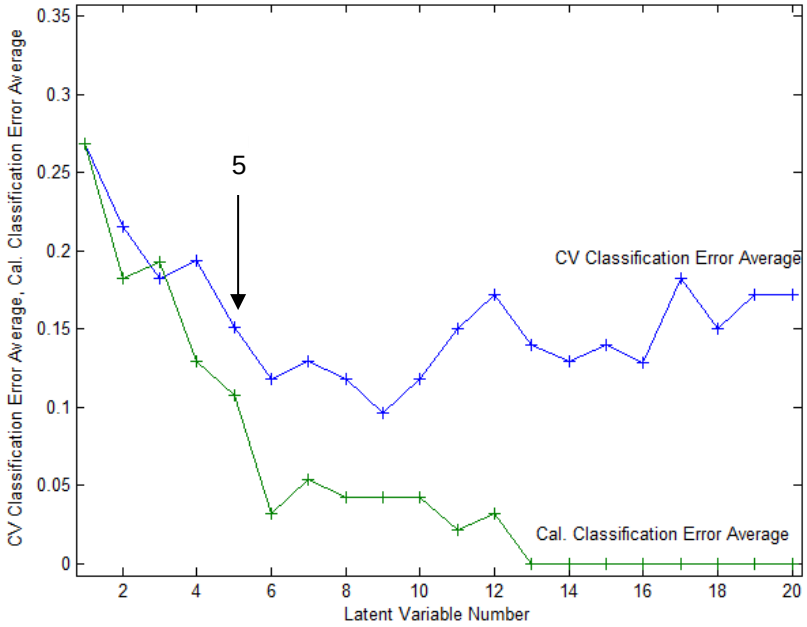


Figure 5.11 - Average Cross-Validation Error for Classification and Average Calibration Classification Error - Groups A, B and C. Cross Validation method: Venetian Blinds

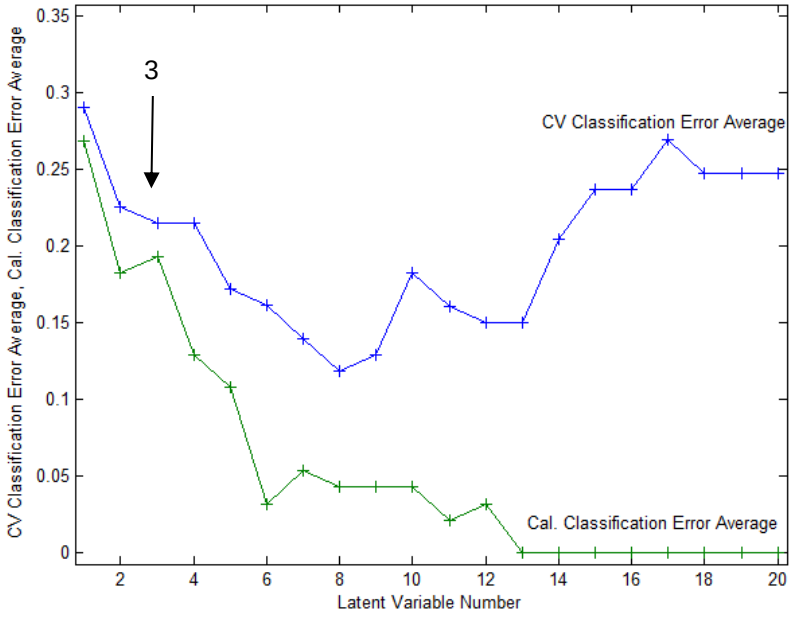


Figure 5.12 - Average Cross-Validation Error for Classification and Average Calibration Classification Error - Groups A, B and C. Cross Validation method: Contiguous Blocks

Considering the Venetian Blinds cross-validation method, Figure 5.13 demonstrates that the model can distinguish positive samples, provenient from groups A and B – in red and negative samples from group C – in blue. Nonetheless, some degree of overlapping on the comparison positives x negatives is present.

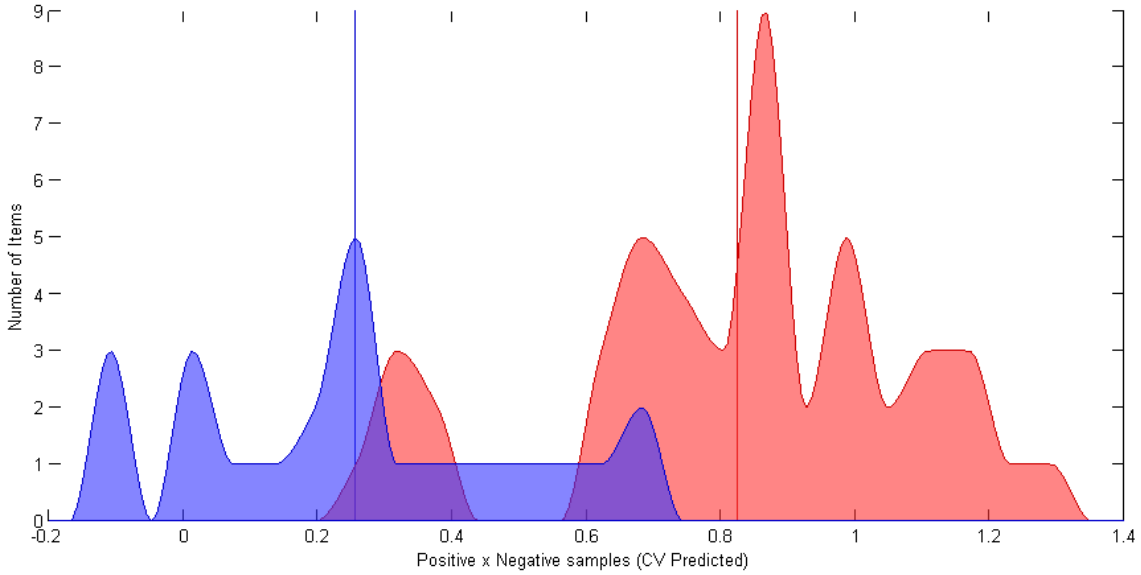


Figure 5.13 - Histogram for PLS-DA Cross-Validation predictions of Positive (groups A and B – red) x Negative (group C - blue) samples - Cross Validation method: Venetian Blinds

The same procedure was applied adopting Contiguous Blocks as the cross-validation method. In this occasion, the data set of 70 samples had been divided in 23 blocks, with 3 samples each. Once individual samples are read in triplicate at the laboratory, the similarity of the analysis can cause model overfitting. This division was made to avoid this effect.

Similar results were obtained, when comparing to Venetian Blinds, as detailed in figure 5.14. A degree of separation and some overlapping of the results can be observed.

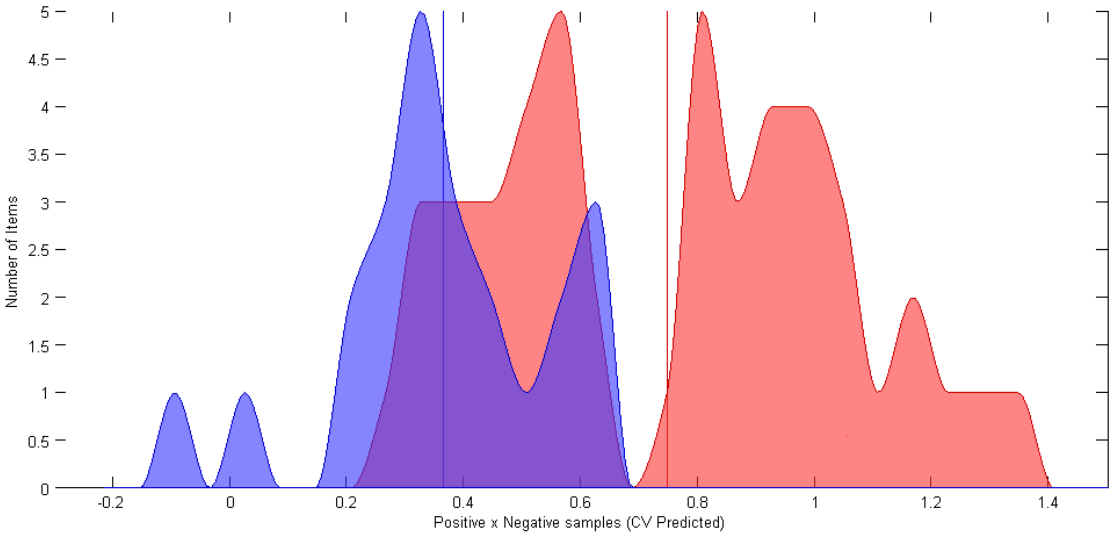


Figure 5.14 - Histogram for PLS-DA Cross-Validation predictions of Positive (groups A and B – red) x Negative (group C - blue) samples - Cross Validation method: Contiguous Blocks.

Afterwards, both models had been challenged with external data originated from the splitting Calibration x Validation dataset. As per results shown on table 5.4, both cross-validation methods have considerable calibration and cross-validation errors for an approximate threshold of 0.50 (RMSEC of 0.28 and 0.34 and RMSECV of 0.33 and 0.39 for the Venetian Blinds and Contiguous Block methodologies, respectively). The models shows a certain capacity of the model to differentiate PCR positive samples from groups A and B from negative ones from Control group C. Nonetheless, the errors are considerable to a threshold of 0.5, which brings uncertainty to the model.

Considering the predictive capacity of the model when it is submitted to external samples, both models demonstrate capacity on differentiate positive and negatives, even though the errors are still substantial. Venetian Blinds cross-validation method presents a RMSEP of 0.32 and Contiguous Block, RMSEP of 0.35. Contiguous Block, though, possess a better predictability capacity, whereas Venetian Block tends to overfitting.

Table 5.4 - RMSEC, RMSECV and RMSEP values for PLS-DA models, applying two different cross-validation methods: Venetian Blinds and Contiguous Block

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

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. Regarding the SARS-CoV-2 virus, however, this relationship is not yet fully defined. Moreover, diagnosis of the disease obtained through the detection of antibodies must be carefully analyzed, as there is a risk of false-positive results from previous contact with other types of coronaviruses and even influenza viruses, which cause common flu.

In any case, information about the presence/absence of IgG antibodies can be certainly useful, especially in population serological surveys or in situations where the determination of previous contact with the virus is of interest.

Considering such importance, 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. All samples in this group were also tested for the presence or absence of IgG antibodies against SARS-CoV-2.

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.

Initially, an exploratory analysis of data from group C (Control) was performed using PCA. The optimum number of factors was selected, as previously described, considering a correlation on the values of RMSECV and RMSEC. When the values start to digress, the Principal Components are no

longer considered, as figure 5.15. Therefore, five Principal Components were selected for this model, shown in table 5.5.

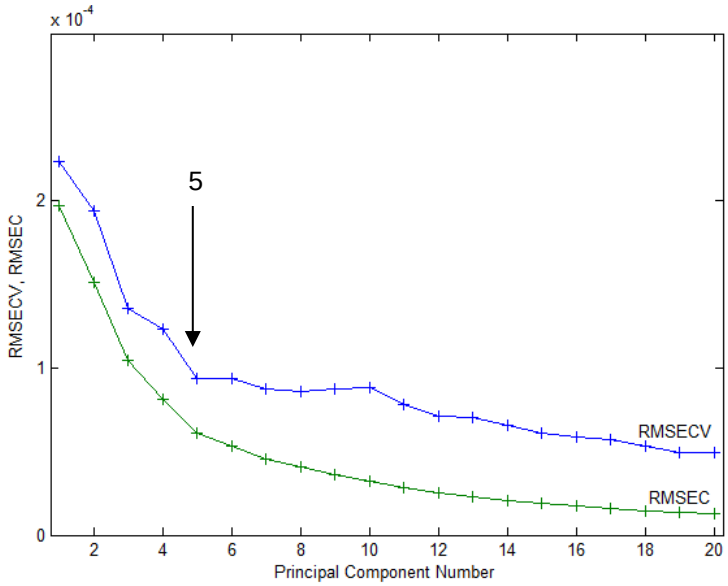


Figure 5.15- RMSEC and RMSECV for group C (IgG Positive and IgG Negative samples)

Table 5.5 - Principal Components and variance captured for group C (IgG Positive and IgG Negative samples)

Principal Component	% Variance Captured	% Total Variance Captured
1	99.19	99.19
2	0.33	99.53
3	0.25	99.77
4	0.09	99.86
5	0.06	99.92

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 5.16.

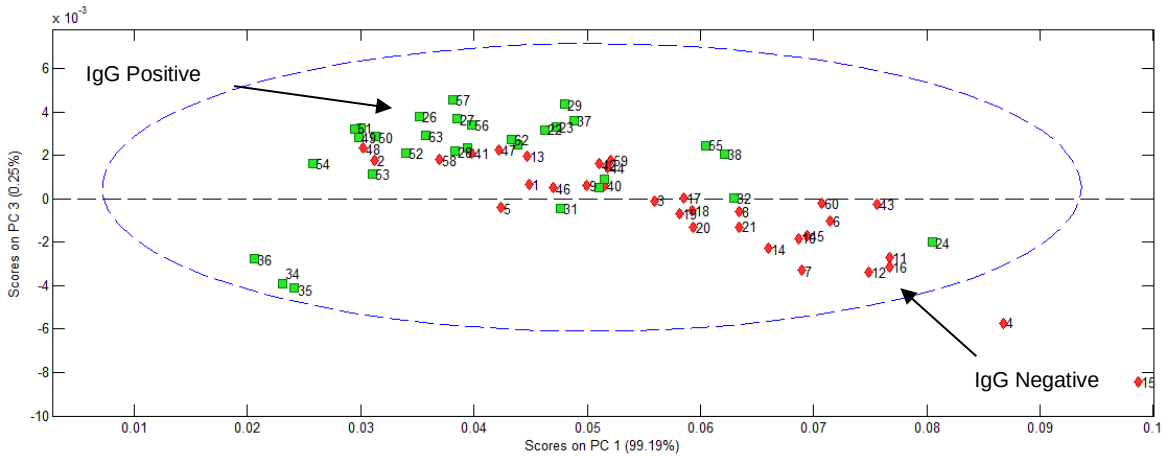


Figure 5.16 - Score plots of group C. IgG Positive samples in green and IgG Negative samples in red. Principal Components 1 and 3

Through loadings analysis (Figures 5.17 and 5.18) 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. Also, there is a big influence from =C= stretching vibrations of the ester functional groups in lipids, probably native from residues of cell membranes, near 1740 cm^{-1}.

Regions between 1470 and 1350 cm^{-1} exert a smaller influence in this PC, still important nonetheless, due to the presence of amino acids.

In PC3, which establishes a greater separation between positive and negative samples, there are also considerable peaks between wavenumbers 1800 and 1500 cm^{-1} for the amide I and amide II. It's possible to report bands to amino acid side-chain vibrations, as in 1516 cm^{-1} (tyrosine) and between 1585 and 1570 cm^{-1} (aspartate and glutamate carboxylate stretching).

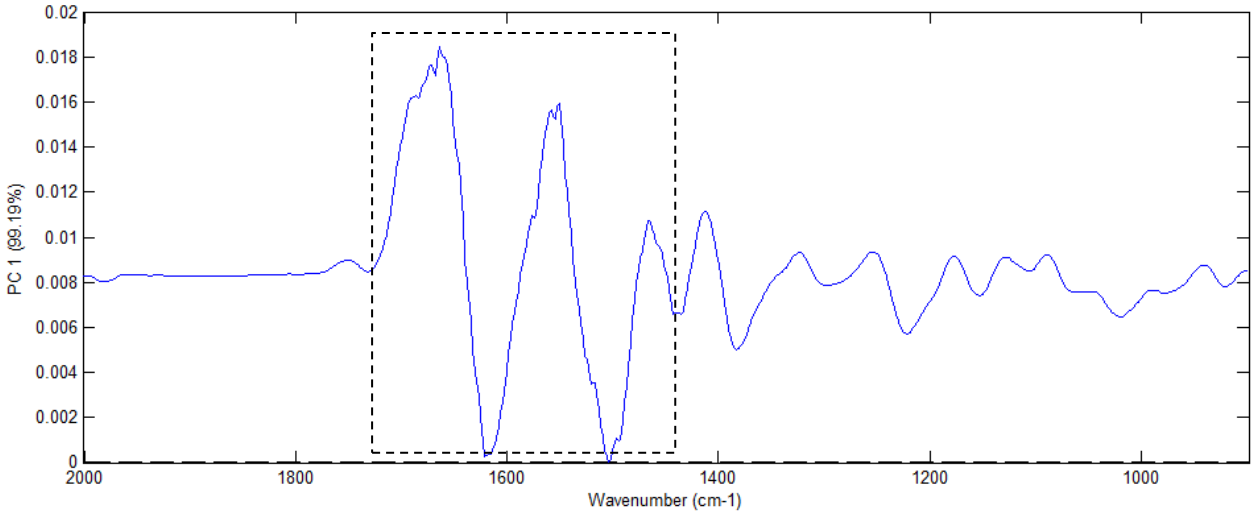


Figure 5.17 - Loadings for Principal Component 1 – Group C. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1}

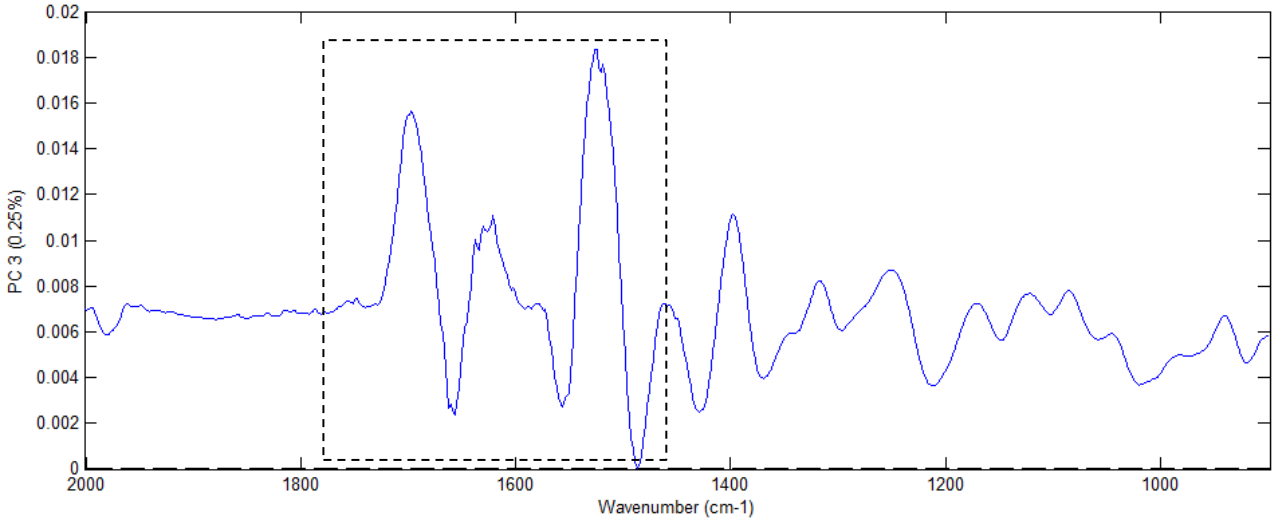


Figure 5.18 - Loadings for Principal Component 3 – Group C. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1}

Being one of the goals of this work to verify if the model is able to differentiate IgG positive samples from IgG negatives, a PLS-DA was performed with control samples (group C) only. Two Cross Validation methods were also tested: Venetian blinds and Contiguous Blocks.

Each model was submitted to external Validation: All dataset had been divided in proportions of 70% of samples for calibration of the model and 30% for external validation, to infer about the prediction ability of the model, resulting in:

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

A graph comparing the average Cross-Validation Error for Classification with average Calibration Classification Error was determined, aiming to define the number of Latent Variables needed to build each model. The difference between Cross-Validation and Calibration errors becomes progressively substantial with the inclusion of new latent variables. Therefore, four Latent Variables had been chosen to Venetian Blinds cross-validation, as figure 5.19, and two Latent Variables for Contiguous Blocks, showed on figure 5.20.

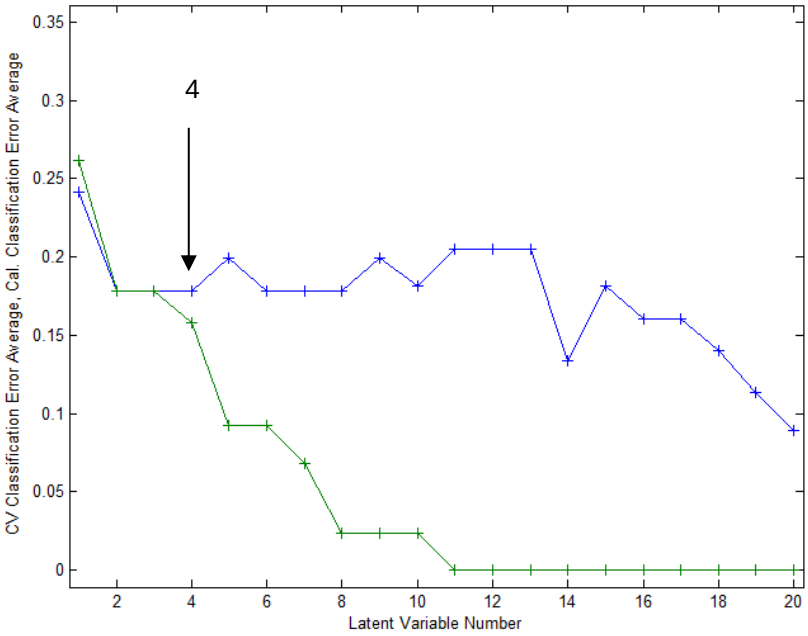


Figure 5.19 - Average Cross-Validation Error for Classification and Average Calibration Classification Error – Group C. Cross Validation method: Venetian Blinds

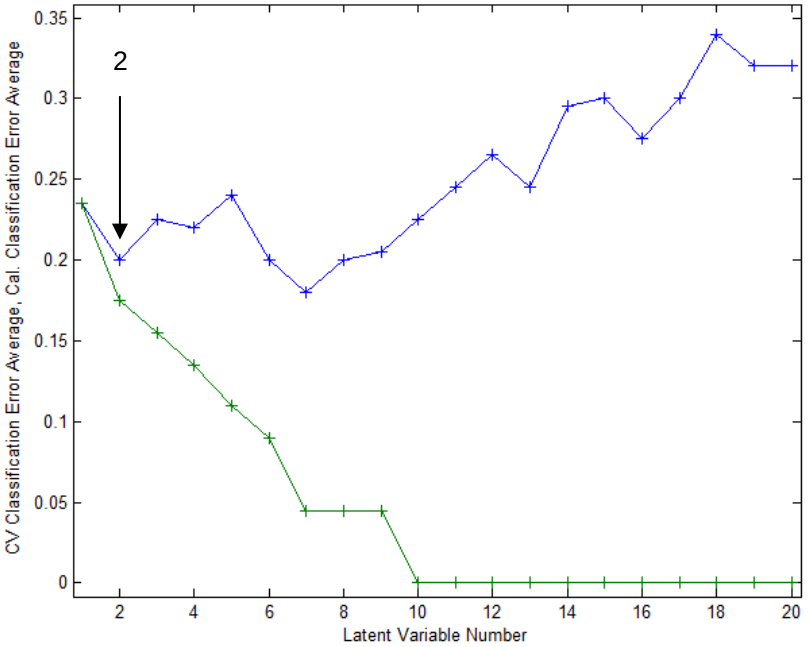


Figure 5.20 - Average Cross-Validation Error for Classification and Average Calibration Classification Error – Group C. Cross Validation method: Contiguous Blocks

As a result from the PLS-DA, Figure 5.21 demonstrates that the model is able to distinguish IgG positive samples, in green, from IgG negative samples in red. Nonetheless, there is some degree of overlapping on the comparison positive x negative samples.

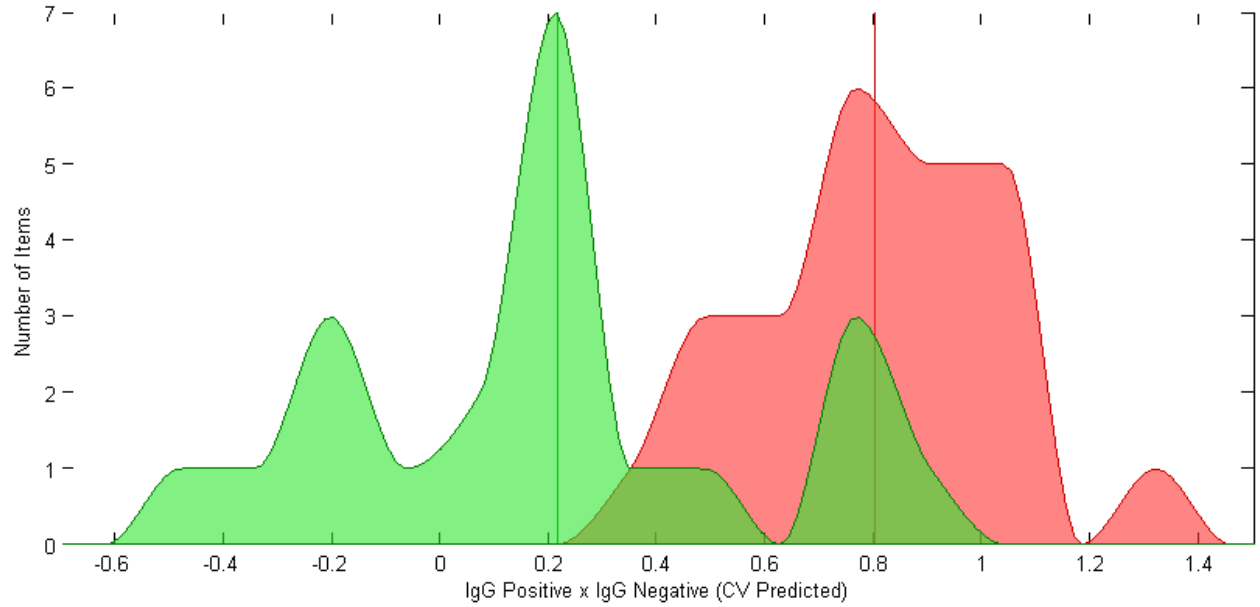


Figure 5.21 - Histogram for PLS-DA Cross-Validation predictions of IgG Positive (green) x IgG Negative (red) samples - Cross Validation method: Venetian Blinds

The same procedure was adopted for Contiguous Blocks as a cross-validation method. Therefore, 45 collected spectra was divided into 15 blocks, with 3 readings each. As each sample is read in triplicate in the laboratory, the similarity of the analysis can cause the model to overfit. This division was made to avoid this effect, as previously described on this work. Still, a fair degree of separation and some overlapping of results can be observed on figure 5.22.

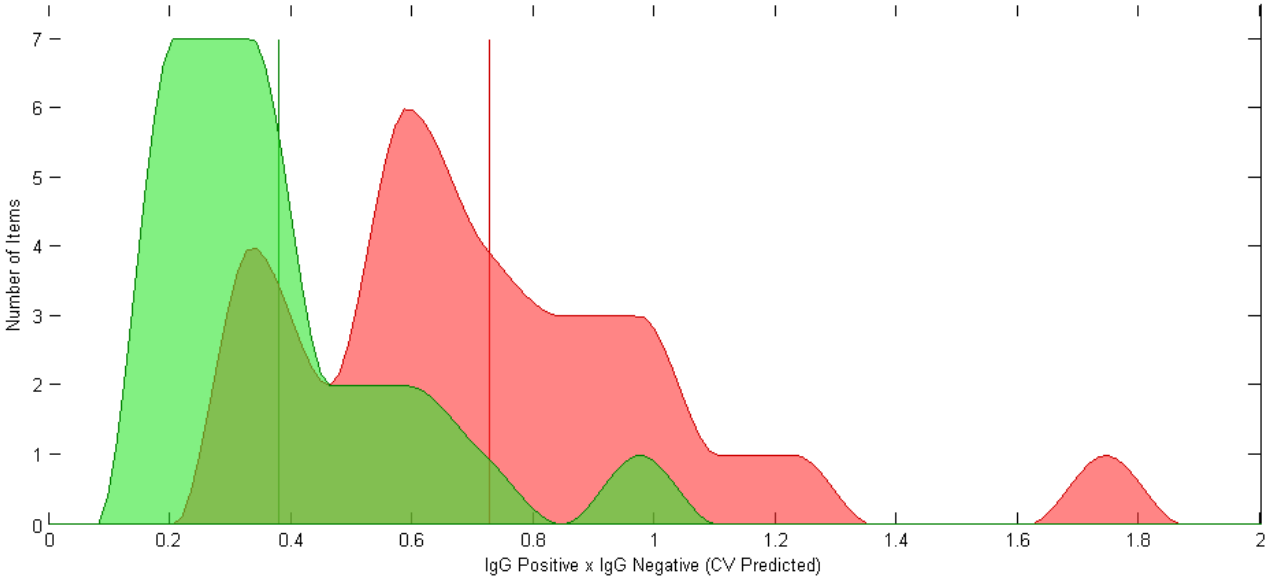


Figure 5.22 - Histogram for PLS-DA Cross-Validation predictions of IgG Positive (green) x IgG Negative (red) samples - Cross Validation method: Contiguous Blocks.

As previously described, both models had been challenged with external data originated from the splitting Calibration x Validation dataset. Both cross-validation methods have big calibration and cross-

validation errors (RMSEC of 0.32 and 0.38 and RMSECV of 0.36 and 0.42 for the Venetian Blinds and Contiguous Block methodologies, respectively), as shown in table 5.6. The model has, to a certain extent, a good capacity of the model to differentiate IgG positive and IgG negative samples. The errors are still big, nonetheless, to a threshold of approximately 0.5.

Regarding the model's predictive capacity when challenged by a new set of samples, it is possible to observe the following: Venetian Blinds cross-validation method presents a RMSEP of 0.39, a prediction error bigger than both RMSEC and RMSECV. For Contiguous Block RMSEP is 0.43, also slightly bigger than RMSCV for this model. 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.

Table 5.6 - RMSEC, RMSECV and RMSEP values for PLS-DA models, applying two different cross-validation methods: Venetian Blinds and Contiguous Block

	Venetian Blinds	**Contiguous Block**
RMSEC	0.32	0.38
RMSECV	0.36	0.42
RMSEP	0.39	0.43

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

During the exploratory analysis of the different datasets, it was possible to evidence a clear separation between the samples of groups A and B. These samples, despite all being PCR positive and processed in the same way in the laboratory of the University of Lisbon, were collected in different hospitals: Beatriz Ângelo and Curry Cabral.

It is possible to observe an important discrimination between the two datasets, originating two clusters on the PCA (Figure 5.23).

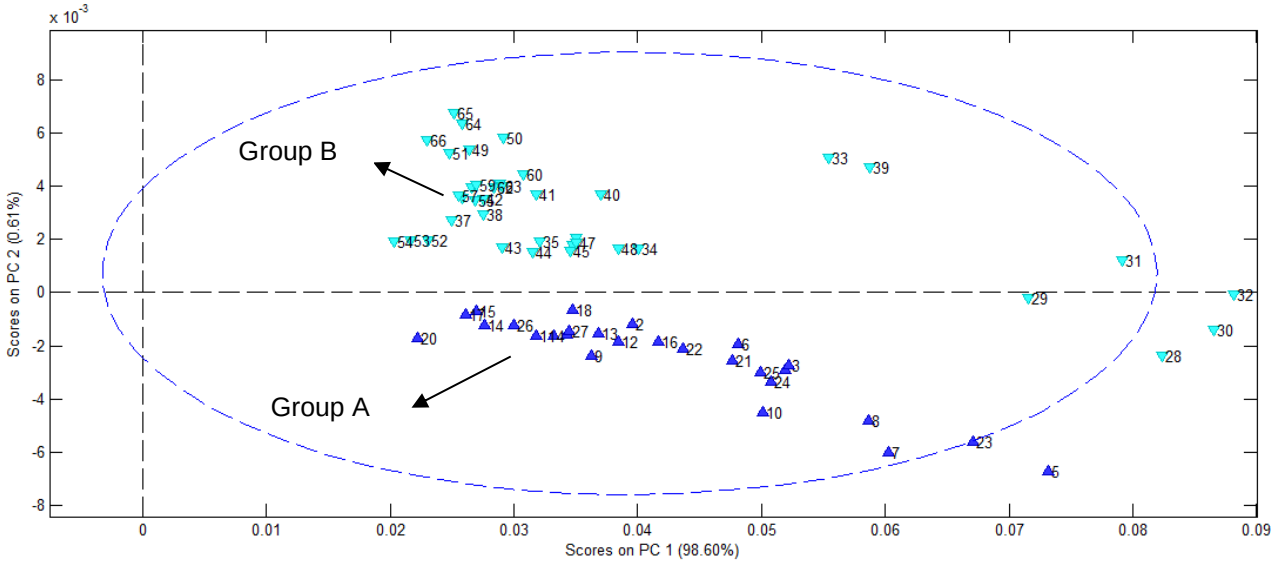


Figure 5.23 - Score plots of group A and B (PCR Positive samples). In light blue, group B – samples collected at Curry Cabral hospital. In blue, group A – samples collected at Beatriz Ângelo Hospital. Principal Components 1 and 2

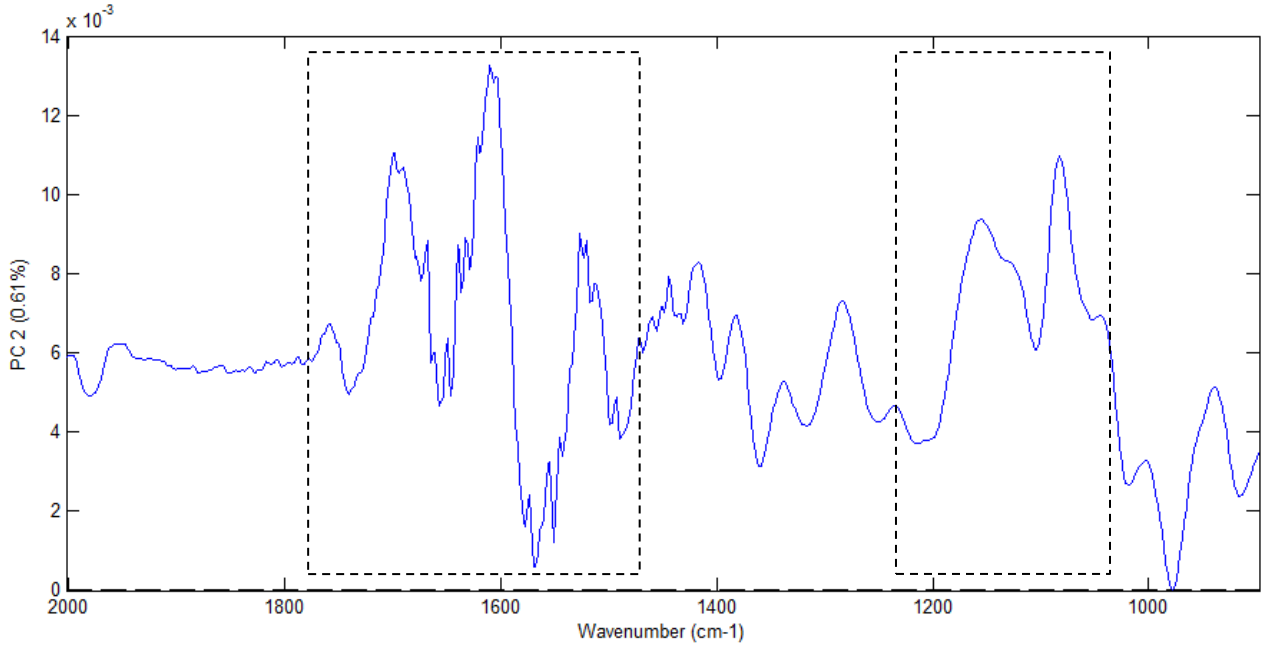


Figure 5.24 - Loadings for Principal Component 2 – Groups A and B. Spectra region highlighted: Considerable influence of loadings between 1800 and 1500 cm^{-1}.

It is possible to notice, at figure 5.24, stretching vibrations on the region between 1800 and 1500 cm^{-1}, as for other datasets, which is dominated by amide I and amide II bands, originated by the presence of proteins (antibodies, cytokines) in the samples. Besides, it is possible to evidence an important deformation influence on wavenumbers from approximately 1200 cm^{-1} to 1000cm^{-1}, which evidences a spectral range dominated by complex ring vibrations of carbohydrates and P=O=P stretching – as previously described, a probable influence of Ficoll®, even though both data sets are treated with the reagent.

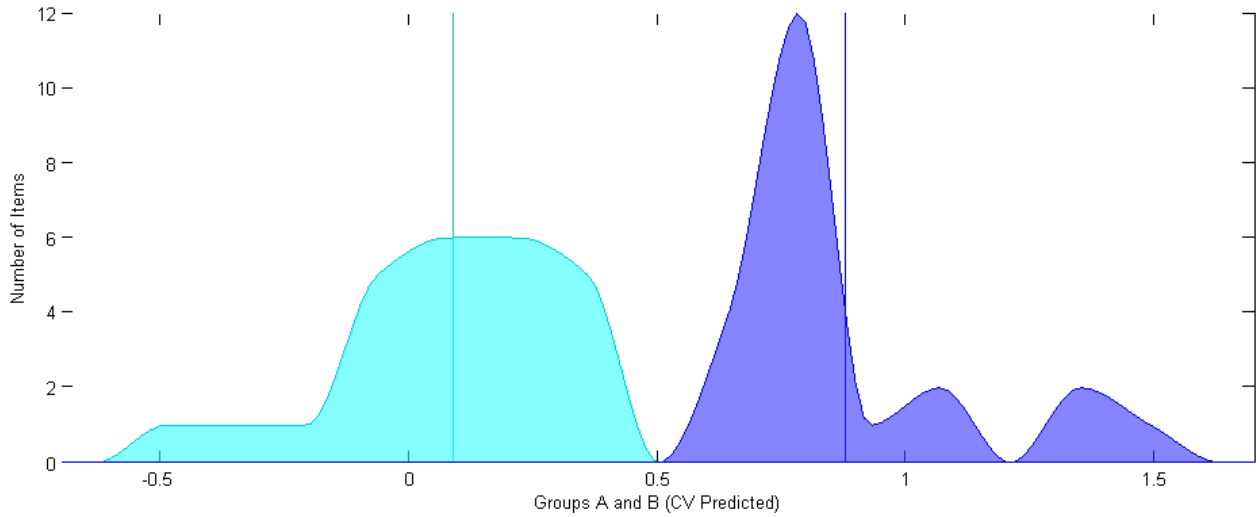


Figure 5.25 - Histogram for PLS-DA Cross-Validation predictions of samples collected in Curry Cabral (light blue) and Beatriz Ângelo (deep blue) hospitals - Cross Validation method: Contiguous Blocks.

The separation of clusters is confirmed by the discriminant PLS in figure 5.25 (Cross-validation: Contiguous Blocks), obtaining very low values of RMSEC (0.22) and RMSECV (0.24). The model is highly capable of separating which samples were collected in different hospitals.

Chapter 6.

Conclusion

To date, there have been more than five million deaths due to COVID-19 worldwide. Social distancing measures and massive and widespread vaccination campaigns have strongly reduced the virus propagation curve, mostly in developed countries. Therefore, these places had started to relax the confinement measures in the last months.

Nonetheless, even after massive vaccination and regardless of the high efficacy of the vaccines currently available, SARS-CoV-2 virus will remain among our society and causing hospitalization and, in last instances, causalities. This is a special problem considering the current situation in low income countries, with less access to vaccines. Besides, new strains of the virus can even put health public systems under strong pressure again. Also, the emergence of new pandemics is basically certain in the future.

Therefore, new and improved tools are still required to reduce the burden of COVID-19 (or, for that matter, new illnesses that might arise).

The repeatability 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, the method of analysis FTIT ATR possess capacity of differentiating between positive and negative samples for SARS-CoV-2, even though the current methodology still shows considerable values of RMSEC and RMSECV (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. This occurs because Contiguous Blocks tends to reduce the effects of reading samples in triplicate, methodology adopted in the current study. 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 considerable RMSEC and RMSECV values (0.32 / 0.36 for Venetian Blinds and 0.38/0.42 for Contiguous Block). The prediction errors obtained with the two cross-validation methodologies (0.39 and 0.43), being greater than the calibration and validation errors, 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. This can be justified 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 (Ex.: In a hospital, samples can be mixed with some reagent /anti-coagulant, while the second hospital follows a different procedure).

6.1 Future Perspectives

Being Infrared techniques extremely sensitive on the detection of different functional groups, the analytical methodology applied on the collection, preparation and reading of both positive and negative blood samples should be optimized and strictly identical. 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. Positive and negative samples should be submitted to the same sample processing methodology, reducing the influence of different reagents on the readings.

Also, acquiring a bigger amount 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, as observed on IgG testing. This can definitely increase the capability of the technique on diagnosing COVID-19 in patients worldwide.

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 (e.g., in a hospital, samples can be mixed with some reagent /anti-coagulant, while the second hospital follows a different procedure). This assumption must be confirmed in further studies, circumstance when the SOPs for blood collection must be evaluated and compared.

References

[01] Rota PA. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. Science. 2003 May 1;300(5624):1394–9.

[02] Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. International Journal of Infectious Diseases [Internet]. 2020 Mar 12 [cited 2021 May 20];0(0). Available from: https://www.ijidonline.com/article/S1201-9712(20)30123-5/fulltext

[03] 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

[04] 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

[05] 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

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

[07] 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

[08] 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

[09] Whiteside TL. Immune Responses to Cancer: Are They Potential Biomarkers of Prognosis? Frontiers in Oncology. 2013 May 17;3(107).

[10] Shokrani M. Cytokines: utility and laboratory measurement [Internet]. www.mlo-online.com. 2011 [cited 2021 May 20]. Available from: https://www.mlo-online.com/home/article/13004250/cytokines-utility-and-laboratory-measurement

[11] Beć KB, Huck CW. Breakthrough Potential in Near-Infrared Spectroscopy: Spectra Simulation. a Review of Recent Developments. Frontiers in Chemistry [Internet]. 2019 Feb 22 [cited 2021 Jun 30];7(48). Available from: https://www.frontiersin.org/articles/10.3389/fchem.2019.00048/full

[12] 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

[13] 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

[14] Detecção precisa de SARS-CoV 2 [Internet]. PerkinElmer, Inc. 2021 [cited 2021 Oct 1]. Available from: https://www.perkinelmer.com.br/

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

[16] Trimble S. Understanding Chemometrics for NIR Spectroscopy | Tools for Applied Food Science | felixinstruments.com [Internet]. Felix Instruments – Applied Food Science. 2021 [cited 2021 Oct 2]. Available from: https://felixinstruments.com/blog/chemometrics-for-spectroscopic-measurements-of-food/

[17] Biancolillo A, Marini F. Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. Frontiers in Chemistry [Internet]. 2018 Nov 21 [cited 2021 Aug 6];6. Available from: https://pubmed.ncbi.nlm.nih.gov/30519559/

[18] M. Wise B, B. Gallagher N, Bro R, M. Shaver J, Windig W, Scott Koch R. Chemometrics Tutorial for PLS_Toolbox and Solo. USA: Eigenvector Research, Inc.; 2006.

[19] What Is Principal Component Analysis (PCA) and How It Is Used? [Internet]. Sartorius AG. 2020 [cited 2021 Sep 15]. Available from: https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186

[20] Holland S. Data Analysis in the Geosciences [Internet]. Uga.edu. 2019 [cited 2021 Sep 10]. Available from: http://strata.uga.edu/8370/lecturenotes/principalComponents.html

[21] Lee LC, Liong C-Y, Jemain AA. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. Analyst [Internet]. 2018 Jul 23 [cited 2021 Sep 16];143(15):3526–39. Available from: https://pubs.rsc.org/lv/content/articlehtml/2018/an/c8an00599k

[22] Russell SM, Alba-Patiño A, Barón E, Borges M, Gonzalez-Freire M, de la Rica R. Biosensors for Managing the COVID-19 Cytokine Storm: Challenges Ahead. ACS Sensors [Internet]. 2020 Jun 2 [cited 2021 Jul 15];5(6):1506–13. Available from: https://pubs.acs.org/doi/10.1021/acssensors.0c00979

[23] Samson R, Navale GR, Dharne MS. Biosensors: frontiers in rapid detection of COVID-19. ACS Sens [Internet]. 2020 Aug 11 [cited 2021 Sep 20];10(9). Available from: https://pubs.acs.org/doi/10.1021/acssensors.0c00979

[24] Maddali H, Miles CE, Kohn J, O'Carroll DM. Optical Biosensors for Virus Detection: Prospects for SARS-CoV-2/COVID-19. ChemBioChem [Internet]. 2020 Dec 9 [cited 2021 Sep 20];22(7). Available from: https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.202000744

[25] Chen Z, Zhang Z, Zhai X, Li Y, Lin L, Zhao H, et al. Rapid and Sensitive Detection of anti-SARS-CoV-2 IgG, Using Lanthanide-Doped Nanoparticles-Based Lateral Flow Immunoassay. Analytical Chemistry [Internet]. 2020 Apr 23 [cited 2021 Sep 25];92(10):7226–31. Available from: https://pubs.acs.org/doi/10.1021/acs.analchem.0c00784

[26] Centrifugation_Brochure_Centrifuge-58XX-family [Internet]. www.eppendorf.com. Eppendorf AG; [cited 2021 Sep 22]. Available from: https://www.eppendorf.com/worldwide/

[27] 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

[28] About PLS Toolbox and Solo - Eigenvector Research Documentation Wiki [Internet]. wiki.eigenvector.com. Eigenvector Research Inc; 2012 [cited 2021 Sep 25]. Available from: http://wiki.eigenvector.com/index.php?title=About_PLS_Toolbox_and_Solo

Annexes

Table 01: Complete profile of samples

Samples			Methods data	Diagnosis Data		Date and Time (Readings)	
Sample ID	Group	Sequential #	Treatment with Ficoll	S1 IgG antibodies[a] (positive/negative borderline)	COVID-19 confirmed by PCR	Date	Time
20200522_111_1_2020-05-22T15-22-46	A	1	yes	NA	yes	22-05-2020	15:22
20200522_111_2_2020-05-22T15-24-40	A	2	yes	NA	yes	22-05-2020	15:24
20200522_111_3_2020-05-22T15-26-21	A	3	yes	NA	yes	22-05-2020	15:26
20200522_131_1_2020-05-22T13-33-58	A	4	yes	NA	yes	22-05-2020	13:33
20200522_131_2_2020-05-22T13-35-56	A	5	yes	NA	yes	22-05-2020	13:35
20200522_131_3_2020-05-22T13-38-29	A	6	yes	NA	yes	22-05-2020	13:38
20200522_132_1_2020-05-22T13-42-05	A	7	yes	NA	yes	22-05-2020	13:42
20200522_132_2_2020-05-22T13-43-58	A	8	yes	NA	yes	22-05-2020	13:43
20200522_132_2_2020-05-22T13-47-00	A	9	yes	NA	yes	22-05-2020	13:47
20200522_133_1_2020-05-22T13-50-22	A	10	yes	NA	yes	22-05-2020	13:50
20200522_133_2_2020-05-22T13-52-47	A	11	yes	NA	yes	22-05-2020	13:52
20200522_133_3_2020-05-22T13-54-51	A	12	yes	NA	yes	22-05-2020	13:54
20200522_136_1_2020-05-22T14-05-29	A	13	yes	NA	yes	22-05-2020	14:05
20200522_136_1_2020-05-22T14-14-02	A	14	yes	NA	yes	22-05-2020	14:14
20200522_136_2_2020-05-22T14-08-25	A	15	yes	NA	yes	22-05-2020	14:08
20200522_136_2_2020-05-22T14-15-53	A	16	yes	NA	yes	22-05-2020	14:15
20200522_136_3_2020-05-22T14-10-11	A	17	yes	NA	yes	22-05-2020	14:10
20200522_136_3_2020-05-22T14-17-43	A	18	yes	NA	yes	22-05-2020	14:17
20200522_137_1_2020-05-22T14-12-14	A	19	yes	NA	yes	22-05-2020	14:12
20200522_137_2_2020-05-22T14-19-48	A	20	yes	NA	yes	22-05-2020	14:19
20200522_137_3_2020-05-22T14-21-36	A	21	yes	NA	yes	22-05-2020	14:21
20200522_141_1_2020-05-22T14-43-10	A	22	yes	NA	yes	22-05-2020	14:43
20200522_141_2_2020-05-22T14-44-45	A	23	yes	NA	yes	22-05-2020	14:44
20200522_141_2_2020-05-22T14-46-15	A	24	yes	NA	yes	22-05-2020	14:46
20200522_143_1_2020-05-22T14-48-00	A	25	yes	NA	yes	22-05-2020	14:48
20200522_143_2_2020-05-22T14-49-48	A	26	yes	NA	yes	22-05-2020	14:49
20200522_143_3_2020-05-22T14-51-41	A	27	yes	NA	yes	22-05-2020	14:51
20200522_PC114_1_2020-05-22T15-33-24	C	28	No	negative	no	22-05-2020	15:33
20200522_PC114_2_2020-05-22T15-35-07	C	29	No	negative	no	22-05-2020	15:35
20200522_PC114_3_2020-05-22T15-36-55	C	30	No	negative	no	22-05-2020	15:36
20200522_PC115_1_2020-05-22T15-52-07	C	31	No	negative	no	22-05-2020	15:52
20200522_PC115_2_2020-05-22T15-53-58	C	32	No	negative	no	22-05-2020	15:53
20200522_PC115_3_2020-05-22T15-55-43	C	33	No	negative	no	22-05-2020	15:55
20200522_PC116_1_2020-05-22T15-57-30	C	34	No	negative	no	22-05-2020	15:57

20200522_PC116_2_2020-05-22T15-59-13	C	35	No	negative	no	22-05-2020	15:59
20200522_PC116_3_2020-05-22T16-01-33	C	36	No	negative	no	22-05-2020	16:01
20200522_PC117_1_2020-05-22T16-03-18	C	37	No	negative	no	22-05-2020	16:03
20200522_PC117_2_2020-05-22T16-04-51	C	38	No	negative	no	22-05-2020	16:04
20200522_PC117_3_2020-05-22T16-06-30	C	39	No	negative	no	22-05-2020	16:06
20200522_PC124_1_2020-05-22T16-26-45	C	40	No	negative	no	22-05-2020	16:26
20200522_PC124_2_2020-05-22T16-28-34	C	41	No	negative	no	22-05-2020	16:28
20200522_PC124_3_2020-05-22T16-30-28	C	42	No	negative	no	22-05-2020	16:30
20200522_PC125_1_2020-05-22T16-32-21	C	43	No	negative	no	22-05-2020	16:32
20200522_PC125_1_2020-05-22T16-33-58	C	44	No	negative	no	22-05-2020	16:33
20200522_PC125_2_2020-05-22T16-35-35	C	45	No	negative	no	22-05-2020	16:35
20200522_PC127_1_2020-05-22T16-48-22	C	46	No	negative	no	22-05-2020	16:48
20200522_PC127_2_2020-05-22T16-50-30	C	47	No	negative	no	22-05-2020	16:50
20200522_PC127_3_2020-05-22T16-52-22	C	48	No	negative	no	22-05-2020	16:52
20200527_PC101_1_2020-05-27T20-33-12	C	49	No	positive	no	27-05-2020	20:33
20200527_PC101_2_2020-05-27T20-34-35	C	50	No	positive	no	27-05-2020	20:34
20200527_PC101_3_2020-05-27T20-36-01	C	51	No	positive	no	27-05-2020	20:36
20200527_PC120_1_2020-05-27T19-07-19	C	52	No	positive	no	27-05-2020	19:07
20200527_PC120_2_2020-05-27T19-09-24	C	53	No	positive	no	27-05-2020	19:09
20200527_PC120_3_2020-05-27T19-11-21	C	54	No	positive	no	27-05-2020	19:11
20200527_PC122_1_2020-05-27T19-18-16	C	55	No	positive	no	27-05-2020	19:20
20200527_PC122_2_2020-05-27T19-20-01	C	56	No	positive	no	27-05-2020	19:21
20200527_PC122_3_2020-05-27T19-21-53	C	57	No	positive	no	27-05-2020	19:18
20200527_PC31_1_2020-05-27T20-28-37	C	58	No	positive	no	27-05-2020	20:28
20200527_PC31_2_2020-05-27T20-30-05	C	59	No	positive	no	27-05-2020	20:30
20200527_PC31_3_2020-05-27T20-31-42	C	60	No	positive	no	27-05-2020	20:31
20200527_PC35_1_2020-05-27T18-56-54	C	61	No	positive	no	27-05-2020	18:56
20200527_PC35_2_2020-05-27T18-58-40	C	62	No	positive	no	27-05-2020	18:58
20200527_PC35_3_2020-05-27T19-00-20	C	63	No	positive	no	27-05-2020	19:00
20200527_PC38_1_2020-05-27T19-13-05	C	64	No	positive	no	27-05-2020	19:13
20200527_PC38_2_2020-05-27T19-14-41	C	65	No	positive	no	27-05-2020	19:14
20200527_PC38_3_2020-05-27T19-16-14	C	66	No	positive	no	27-05-2020	19:16
20200527_PC43_1_2020-05-27T19-49-16	C	67	No	negative	no	27-05-2020	19:49
20200527_PC43_2_2020-05-27T19-50-45	C	68	No	negative	no	27-05-2020	19:50
20200527_PC43_3_2020-05-27T19-52-20	C	69	No	negative	no	27-05-2020	19:52
20200527_PC44_1_2020-05-27T20-19-03	C	70	No	negative	no	27-05-2020	20:19
20200527_PC44_2_2020-05-27T20-20-31	C	71	No	negative	no	27-05-2020	20:20
20200527_PC44_3_2020-05-27T20-21-59	C	72	No	negative	no	27-05-2020	20:21
20200527_PC47_1_2020-05-27T20-44-35	C	73	No	negative	no	27-05-2020	20:44
20200527_PC47_2_2020-05-27T20-46-19	C	74	No	negative	no	27-05-2020	20:46
20200527_PC47_3_2020-05-27T20-47-54	C	75	No	negative	no	27-05-2020	20:47

20200527_PC72_1_2020-05-27T19-02-13	C	76	No	positive	no	27-05-2020	19:02
20200527_PC72_2_2020-05-27T19-03-49	C	77	No	positive	no	27-05-2020	19:03
20200527_PC72_3_2020-05-27T19-05-31	C	78	No	positive	no	27-05-2020	19:05
20200527_PC73_1_2020-05-27T19-38-49	C	79	No	positive	no	27-05-2020	19:38
20200527_PC73_2_2020-05-27T19-40-37	C	80	No	positive	no	27-05-2020	19:40
20200527_PC73_3_2020-05-27T19-42-14	C	81	No	positive	no	27-05-2020	19:42
20200527_PC81_1_2020-05-27T19-54-07	C	82	No	positive	no	27-05-2020	19:54
20200527_PC81_2_2020-05-27T19-55-54	C	83	No	positive	no	27-05-2020	19:55
20200527_PC81_3_2020-05-27T19-57-35	C	84	No	positive	no	27-05-2020	19:57
20200527_PC84_1_2020-05-27T19-59-40	C	85	No	negative	no	27-05-2020	19:59
20200527_PC84_2_2020-05-27T20-01-15	C	86	No	negative	no	27-05-2020	20:01
20200527_PC84_3_2020-05-27T20-02-43	C	87	No	negative	no	27-05-2020	20:02
20200527_PC96_1_2020-05-27T20-39-21	C	88	No	positive	no	27-05-2020	20:39
20200527_PC96_2_2020-05-27T20-40-55	C	89	No	positive	no	27-05-2020	20:40
20200527_PC96_3_2020-05-27T20-42-24	C	90	No	positive	no	27-05-2020	20:42
20200528_Y001_1_2020-05-28T17-46-33	B	91	Yes	positive	yes	28-05-2020	17:46
20200528_Y001_2_2020-05-28T17-48-15	B	92	Yes	positive	yes	28-05-2020	17:48
20200528_Y001_3_2020-05-28T17-49-50	B	93	Yes	positive	yes	28-05-2020	17:49
20200528_Y002_1_2020-05-28T17-51-34	B	94	Yes	positive	yes	28-05-2020	17:51
20200528_Y002_2_2020-05-28T17-53-18	B	95	Yes	positive	yes	28-05-2020	17:53
20200528_Y002_3_2020-05-28T17-54-53	B	96	Yes	positive	yes	28-05-2020	17:54
20200528_Y003_1_2020-05-28T18-01-04	B	97	Yes	positive	yes	28-05-2020	18:01
20200528_Y003_2_2020-05-28T18-02-37	B	98	Yes	positive	yes	28-05-2020	18:02
20200528_Y003_3_2020-05-28T18-04-07	B	99	Yes	positive	yes	28-05-2020	18:04
20200528_Y004_1_2020-05-28T17-56-32	B	100	Yes	positive	yes	28-05-2020	17:56
20200528_Y004_2_2020-05-28T17-58-04	B	101	Yes	positive	yes	28-05-2020	17:58
20200528_Y004_3_2020-05-28T17-59-32	B	102	Yes	positive	yes	28-05-2020	17:59
20200528_Y005_1_2020-05-28T18-07-27	B	103	Yes	Negative	yes	28-05-2020	18:07
20200528_Y005_2_2020-05-28T18-09-00	B	104	Yes	Negative	yes	28-05-2020	18:09
20200528_Y005_3_2020-05-28T18-10-28	B	105	Yes	Negative	yes	28-05-2020	18:10
20200528_Y006_1_2020-05-28T18-18-29	B	106	Yes	positive	yes	28-05-2020	18:18
20200528_Y006_2_2020-05-28T18-20-25	B	107	Yes	positive	yes	28-05-2020	18:20
20200528_Y006_3_2020-05-28T18-22-07	B	108	Yes	positive	yes	28-05-2020	18:22
20200528_Y007_1_2020-05-28T18-24-07	B	109	Yes	Negative	yes	28-05-2020	18:24
20200528_Y007_2_2020-05-28T18-25-36	B	110	Yes	Negative	yes	28-05-2020	18:25
20200528_Y007_3_2020-05-28T18-27-15	B	111	Yes	Negative	yes	28-05-2020	18:27
20200528_Y008_1_2020-05-28T18-28-52	B	112	Yes	positive	yes	28-05-2020	18:28
20200528_Y008_2_2020-05-28T18-31-37	B	113	Yes	positive	yes	28-05-2020	18:31
20200528_Y008_3_2020-05-28T18-33-03	B	114	Yes	positive	yes	28-05-2020	18:33
20200528_Y009_1_2020-05-28T18-41-58	B	115	Yes	positive	yes	28-05-2020	18:41
20200528_Y009_2_2020-05-28T18-43-31	B	116	Yes	positive	yes	28-05-2020	18:43

20200528_Y009_3_2020-05-28T18-45-26	B	117	Yes	positive	yes	28-05-2020	18:45
20200528_Y010_1_2020-05-28T18-47-33	B	118	Yes	positive	yes	28-05-2020	18:47
20200528_Y010_2_2020-05-28T18-49-10	B	119	Yes	positive	yes	28-05-2020	18:49
20200528_Y010_3_2020-05-28T18-50-55	B	120	Yes	positive	yes	28-05-2020	18:50
20200528_Y011_1_2020-05-28T18-52-33	B	121	Yes	positive	yes	28-05-2020	18:52
20200528_Y011_2_2020-05-28T18-54-13	B	122	Yes	positive	yes	28-05-2020	18:54
20200528_Y011_3_2020-05-28T18-55-46	B	123	Yes	positive	yes	28-05-2020	18:55
20200528_Y013_1_2020-05-28T19-04-38	B	124	Yes	positive	yes	28-05-2020	19:04
20200528_Y013_2_2020-05-28T19-06-37	B	125	Yes	positive	yes	28-05-2020	19:06
20200528_Y013_3_2020-05-28T19-08-19	B	126	Yes	positive	yes	28-05-2020	19:08
20200528_Y014_1_2020-05-28T18-59-54	B	127	Yes	positive	yes	28-05-2020	18:59
20200528_Y014_2_2020-05-28T19-01-21	B	128	Yes	positive	yes	28-05-2020	19:01
20200528_Y014_3_2020-05-28T19-02-56	B	129	Yes	positive	yes	28-05-2020	19:02