

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE
FACULDADE DE MEDICINA

DÉBORA LARA ZUZA SCHEUCHER

**Technological platforms in the prognosis of COVID-19:
ATR-FTIR spectroscopy, metabolomic salivary analysis
and machine learning**

UBERLÂNDIA

2022

DÉBORA LARA ZUZA SCHEUCHER

**Technological platforms in the prognosis of COVID-19:
ATR-FTIR spectroscopy, metabolomic salivary analysis
and machine learning**

Tese apresentada ao Programa de Pós-graduação em Ciências da Saúde da Faculdade de Medicina da Universidade Federal de Uberlândia, como requisito parcial para a obtenção do Título de Doutor em Ciências da Saúde.

Área de concentração: Ciências da Saúde.

Orientador: Prof. Dr. Thúlio Marquez Cunha

Coorientadora: Profa. Dra. Yara Cristina de Paiva Maia

UBERLÂNDIA

2022

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

S328t
2022 Scheucher, Débora Lara Zuza, 1985-
Technological platforms in the prognosis of COVID-19 [recurso eletrônico] : ATR-FTIR spectroscopy, metabolomic salivary analysis and machine learning / Débora Lara Zuza Scheucher. - 2022.
121 f.

Orientador: Thúlio Marquez Cunha.

Coorientadora: Yara Cristina de Paiva Maia.

Tese (Doutorado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Ciências da Saúde.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.te.2023.7004>

Inclui bibliografia.

1. Ciências médicas. 2. COVID-19. 3. ATR-FTIR. 4. Metabolomic. 5. Machine learning. 6. Support vector machine. 7. Saliva. I. Cunha, Thúlio Marquez, 1979-, (Orient.). II. Maia, Yara Cristina de Paiva, 1975-, (Coorient.). III. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Ciências da Saúde. IV. Título.

CDU: 61

André Carlos Francisco
Bibliotecário – CRB-6/3408

FOLHA DE APROVAÇÃO

DÉBORA LARA ZUZA SCHEUCHER

**Technological platforms in the prognosis of COVID-19: ATR-FTIR
spectroscopy, metabolomic salivary analysis and machine learning**

Presidente da banca: Prof. Dr. Thúlio Marquez Cunha

Tese apresentada ao Programa de Pós-graduação em Ciências da Saúde da Faculdade de Medicina da Universidade Federal de Uberlândia, como requisito parcial para a obtenção do Título de Doutor em Ciências da Saúde.

Área de concentração: Ciências da Saúde.

Banca examinadora:

Titular: Flávia Bittar Britto Arantes

Instituição: Universidade Federal de Uberlândia

Titular: Guilherme Marques Andrade

Instituição: Universidade Federal de Uberlândia

Titular: Suzana Erico Tanni Minamoto

Instituição: Faculdade de Medicina de Botucatu - UNESP

Titular: Larissa Prado Maia

Instituição: Universidade do Estado de Minas Gerais

Suplente: Marcus Vinícius de Paula Netto

Instituição: Universidade Federal de Uberlândia

Suplente: Elaine Cristina Gonçalves

Instituição: Faculdade de Medicina de Ribeirão Preto - USP

Uberlândia, 16 de dezembro de 2022



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciências da Saúde				
Defesa de:	Tese de Doutorado Nº 24/PPCSA				
Data:	16.12.2022	Hora de início:	13:00h	Hora de encerramento:	18:30h
Matrícula do Discente:	11713CSD003				
Nome do Discente:	Débora Lara Zuza Scheucher				
Título do Trabalho:	Technological platforms in the prognosis of COVID-19: ATR-FTIR spectroscopy, metabolomic salivary analysis and machine learning				
Área de concentração:	Ciências da Saúde				
Linha de pesquisa:	2 - DIAGNÓSTICO, TRATAMENTO E PROGNÓSTICO DAS DOENÇAS E AGRAVOS À SAÚDE				
Projeto de Pesquisa de vinculação:	DESENVOLVIMENTO DE PLATAFORMAS NANOBIOTECNOLOGICAS PARA O DIAGNÓSTICO DE DOENÇAS PULMONARES				

Reuniu-se em web conferência pela plataforma Mconf-RNP, em conformidade com a PORTARIA Nº 36, DE 19 DE MARÇO DE 2020 da COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR - CAPES, pela Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciências da Saúde, assim composta: Profs. Drs. Suzana Erico Tanni Minamoto (UNESP), Larissa Prado Maia (UEMG), Guilherme Marquez Andrade (UFU), Flávia Bittar Britto Arantes (UFU) e Thúlio Marquez Cunha (UFU), orientador da candidata.

Iniciando os trabalhos, o presidente da mesa, Prof. Dr. Thúlio Marquez Cunha, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu a Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovada.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



16/12/2022, às 15:35, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Larissa Prado Maia, Usuário Externo**, em 16/12/2022, às 15:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Flávia Bittar Britto Arantes, Professor(a) do Magistério Superior**, em 16/12/2022, às 15:42, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Thúlio Marquez Cunha, Professor(a) do Magistério Superior**, em 16/12/2022, às 15:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Guilherme Marques Andrade, Professor(a) do Magistério Superior**, em 16/12/2022, às 15:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4134809** e o código CRC **A9339732**.

DEDICATÓRIA

“Se enxerguei mais longe, foi porque estava sobre os ombros de gigantes.”

Isaac Newton

Ao meu querido esposo **Públio**, parceiro nas lutas e desafios que surgiram ao longo do caminho. Dear, você esteve comigo em meio às tempestades, me lembrando sempre que existe sol atrás das nuvens cinzentas. Te amo! Ao nosso **Felipe**, herança preciosa que o Senhor nos concedeu, pelo sorriso e alegria que nos transportam em segundos para lugares celestiais. À nossa querida **Amanda**, que está chegando para iluminar ainda mais os nossos dias.

Aos meus pais maravilhosos, **Carlos e Elaine**, que me proporcionaram uma formação intelectual e espiritual que me permitiu concluir mais esta etapa com êxito e equilíbrio. Mãe, obrigada por dispor de tanto tempo! Você sabe bem como abriu mão de tantas coisas em meu auxílio. Muito obrigada pelo apoio constante, por suas orações e suporte físico e espiritual!

Aos pacientes **vítimas de COVID** que motivaram a busca incansável pelo conhecimento.

AGRADECIMENTOS ESPECIAIS

Ao **Prof. Dr. Thúlio Marquez Cunha** pela oportunidade, apoio, orientação, exigência e incentivo durante o desenvolvimento deste trabalho. Muito obrigada pela confiança e pelo aprendizado. Compartilho essa conquista com você!

À querida **Profª. Drª. Yara Cristina de Paiva Maia**, que me acolheu com tanto carinho. A sua sensibilidade, competência e amor ao que faz transformam tudo ao seu redor. Você me transformou! Obrigada pelo incentivo constante e por acreditar em mim mesmo quando eu mesma duvidava. Sou profundamente grata a você!

Ao **Prof. Dr. Robinson Sabino-Silva** por chegar em hora oportuna. Você transformou degraus em rampas me conduzindo aonde eu precisava chegar. Muito obrigada!

À querida **Drª. Luciana Oliveira e Silva**, que foi um anjo usado por Deus para trazer “resultados”, cura, força, equilíbrio e tantos ensinamentos. Lu, obrigada pela doçura, paciência e perseverança!

À **Natália Cristina Martins Nunes Mendes**, amiga fiel, que desde a proposta inicial acreditou! Naty, você encurtou o caminho e ajudou a transformar o impossível em possível. Obrigada!

À equipe do **Laboratório de Nanobiotecnologia**, em especial a **Alinne Tatiane Faria Silva, Mário Machado Martins e Hebréia Oliveira Almeida Souza**, pela parceria e aprendizado. Vocês foram pessoas chave nesse processo.

Ao **Prof. Dr. Murillo Guimarães Carneiro** pela parceria e apoio neste trabalho. Muito obrigada!

À equipe do grupo de pesquisa em **Biologia Molecular e Nutrição (BioNut)** por todas as discussões e aprendizado. Aprendi muito com todos vocês!

AGRADECIMENTOS

À querida amiga **Amanda Chan Yuk Ki Freitas**, sempre pronta a ajudar nas horas de maior dificuldade!

Ao meu irmão, **Adriano**, pelas trocas de experiência e pelo apoio sempre.

À minha querida **vovó Santinha** que me abençoa e me cobre de orações incessantes.

A todos os queridos **familiares e amigos** pelo apoio e carinho, nesses dias que não foram nada fáceis!

À **equipe UFU** pelo auxílio, apoio e compreensão em tantos momentos de cansaço e sobrecarga.

Aos **funcionários UTI COVID, do HC-UFU**. Vocês, com toda a simpatia e disponibilidade, tornaram esse período mais leve e alegre. Agradeço muito por toda a ajuda!

Ao **Programa de Pós-Graduação em Ciências da Saúde** e à **Universidade Federal de Uberlândia** por tornarem esse sonho possível. Em especial às secretárias **Gisele de Melo Rodrigues e Viviane Gonçalves**, pela

disponibilidade, atenção, carinho e apoio. Sempre tão dispostas a ajudar!

A **todos** aqueles que eu não citei e que deveria ter citado. Perdoem-me pela involuntária indelicadeza e aceitem meus sinceros agradecimentos. Meu coração tem plena certeza de que nada se faz sem a ajuda de muitos.

A DEUS

Só tu conheces Senhor as tantas vezes que os joelhos se dobraram e as lágrimas surgiram incessantes, mas a tua presença constante, e o teu amor abundante renovaram a graça e as tuas misericórdias a cada manhã.

A ti Senhor consagro este trabalho e a minha vida.

“Mas os que esperam no Senhor renovam as suas forças, sobem com asas como águias, correm e não se cansam, caminham e não se fatigam.”

Isaías 40:31

“Porque dele, e por ele, e para ele são todas as coisas; glória, pois, a ele eternamente.

Amém!”

Romanos 11:36

EPÍGRAFE

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.”

Martin Luther King

SCHEUCHER, D. L. Z. **Plataformas tecnológicas no prognóstico do COVID-19: espectroscopia ATR-FTIR, análise metabolômica salivar e aprendizado de máquina**. 2022. 121 p. Tese (Doutorado) – Universidade Federal de Uberlândia, 2022. Programa de Pós Graduação em Ciências da Saúde. Disponível em: <http://doi.org/10.14393/ufu.te.2023.7004>

RESUMO

A COVID-19 é uma infecção respiratória aguda grave causada pelo coronavírus SARS-CoV-2 que resultou em mais de 636 milhões de casos confirmados e 6,6 milhões de mortes no mundo, e o seu manejo continua desafiando as equipes de saúde. O desenvolvimento de novas plataformas tecnológicas para prever a gravidade e o prognóstico da doença são necessários para auxiliar os profissionais de saúde no processo de tomada de decisão, e melhorar o conhecimento sobre a fisiopatologia da doença e seu manejo. Neste contexto, a reflexão total atenuada associada à espectroscopia de infravermelho por transformada de Fourier (ATR-FTIR) consiste em uma espectroscopia vibratória que pode fornecer informações sobre a estrutura e composição química de materiais biológicos em nível molecular. A análise de metabolômica identifica compostos desconhecidos e determina a estrutura e as propriedades químicas das moléculas, refletindo o estado metabólico de órgãos e tecidos. Além disso, o aprendizado de máquina é outra plataforma promissora que usa inteligência artificial para tomar a decisão mais adequada com base no reconhecimento de padrões nos dados analisados. A maioria dos estudos utiliza essas ferramentas ligadas ao uso do sangue, porém são necessários métodos não invasivos, como por exemplo, o uso de saliva. Este estudo tem como objetivos: desenvolver modelos de algoritmos de aprendizado de máquina com valor preditivo e prognóstico baseados em assinaturas espectrais infravermelhas salivares de pacientes com COVID-19 para o prognóstico de (i) COVID-19 moderado versus grave, (ii) alta versus óbito e (iii) ausência de ventilação mecânica invasiva (VMI) versus VMI (Artigo 1); analisar os metabólitos da saliva de pacientes com COVID-19 coletados precocemente na internação para estratificação da doença por (i) gravidade – COVID-19 moderada vs grave, (ii) alta versus óbito e (iii) necessidade de VMI (Artigo 2). A saliva de 118 pacientes infectados com SARS-CoV-2 foi

coletada no início da internação (média de $3,0 \pm 3,2$ dias) e analisada por ATR-FTIR, aprendizado de máquina e cromatografia líquida de alta eficiência hifenizada ao espectrômetro de massa (HPLC-MS). A classificação dos espectros de infravermelho salivar por máquina de vetores de suporte (SVM) mostrou (i) valor preditivo de gravidade – moderado vs grave [sensibilidade de 68%, especificidade de 52% e acurácia de 60%], (ii) valor preditivo de sobrevivência – alta vs morte [sensibilidade de 60%, especificidade de 64% e acurácia de 64%], (iii) valor preditivo de necessidade de VMI (sem VMI x VMI) [sensibilidade de 46%, especificidade de 72% e acurácia de 66%] (Artigo 1). O perfil metabolômico dos pacientes com COVID-19 mostrou envolvimento de purinas, amida, lipídios, sulfóxidos e cetonas. A metabolômica salivar permitiu identificar os metabólitos relacionados à maior gravidade da doença nos desfechos avaliados, que podem ser usados na seleção de um potencial biomarcador salivar para a avaliação da gravidade no início da internação e abrir novas perspectivas terapêuticas contra a COVID-19 (Artigo 2). Este estudo revela o potencial da análise salivar por meio da espectroscopia de ATR-FTIR, metabolômica e aprendizado de máquina como plataformas tecnológicas prognósticas para determinar o desfecho de pacientes com COVID-19.

Palavras chave: COVID-19, ATR-FTIR, metabolômica, aprendizado de máquina, máquina de vetores de suporte, saliva.

SCHEUCHER, D. L. Z. **Technological platforms in the prognosis of COVID-19: ATR-FTIR spectroscopy, metabolomic salivary analysis and machine learning.** 2022. 121 p. PhD Thesis – Federal University of Uberlandia, 2022. Post Graduate Program of Health Sciences. Available at: <http://doi.org/10.14393/ufu.te.2023.7004>

ABSTRACT

The COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, resulted in more than 636 million confirmed cases and 6.6 million deaths worldwide, and its management remains a challenge for health teams. The development of novel tools for predicting COVID-19 severity and prognosis could assist healthcare professionals in the decision-making process, and improve knowledge about the pathophysiology of the disease. In this context, attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy consists of a vibratory spectroscopy that can provide information about the structure and chemical composition of biological materials at the molecular level. Metabolomic analysis identifies unknown compounds, and determine structure and chemical properties of molecules, reflecting the metabolic state of organs and tissues. Moreover, machine learning is another promising platform that uses artificial intelligence to make the most appropriate decision based on pattern recognition within the analyzed data. Most of studies use these tools linked to the use of blood, however non-invasive methods are needed, for example, using saliva. This study aims to: develop predictive and prognostic value models of machine learning algorithms based on salivary infrared spectral signatures to the prognosis of (i) moderate vs severe COVID, (ii) discharge vs death, and (iii) no need of invasive mechanical ventilation (IMV) vs IMV outcomes for COVID-19 patients (Article 1); analyze the saliva metabolites from COVID-19 patients for disease stratification by (i) the disease progression - moderate or severe COVID, (ii) discharge vs death, and (iii) the need of IMV, performing the metabolomics of saliva collected in early stages of hospitalization (Article 2). We analyzed the saliva of 118 patients infected with SARS-CoV-2 by ATR-FTIR, machine learning and high-performance liquid chromatography hyphenated to the mass spectrometer (HPLC-MS) early in hospitalization (mean 3.0 ± 3.2 days). The classification of salivary infrared spectra by support vector machine

(SVM) showed (i) a predictive value of severity – moderate vs severe [sensitivity of 68%, specificity of 52%, and accuracy of 60%], (ii) a predictive value of survival – discharge vs death [sensitivity of 60%, specificity of 64%, and accuracy of 64%], (iii) a predictive value of need of IMV (no IMV x IMV) [sensitivity of 46%, specificity of 72%, and accuracy of 66%] (Article 1). Metabolomic profile of COVID-19 patients showed an involvement of purines, amide, lipid, sulfoxide and ketones. Saliva metabolome allowed us to identify metabolites related to greater disease severity in the evaluated outcomes which might be used in selection of potential saliva biomarker for severity evaluation early in hospitalization and open new therapeutic perspectives against COVID-19 (Article 2). This study reveals the potential of salivary ATR-FTIR spectroscopy, metabolomic analysis and machine learning as prognostic technological platforms to determine outcome in COVID-19 patients.

Keywords: COVID-19, ATR-FTIR, metabolomic, machine learning, support vector machine, saliva.

ILLUSTRATION LIST

THEORETICAL FOUNDATION

Figure 1	Global overview of COVID-19	27
Figure 2	Structure and genomic organization of SARS-CoV-2	28
Figure 3	Coronavirus replication cycle	30
Figure 4	Types of vibrations on the functional groups	37
Figure 5	The electromagnetic spectrum	38
Figure 6	Major absorptions of bonds in organic molecules in the MIR region (4000 - 400 cm ⁻¹)	39
Figure 7	Molecular families and its location in the MIR region (4000 - 400 cm ⁻¹)	40
Figure 8	Basic components of an FTIR spectrometer	40
Figure 9	(a) Illustration of the evanescent wave; (b) principle of attenuated total reflection (ATR). dp = depth of the evanescent wave; n = refraction index	42
Figure 10	A simplified illustration of how the support vector machine works	48

ARTICLE 1

Figure 1	Study Flowchart	55
Figure 2	Representative average of ATR-FTIR spectra (800–1800 cm ⁻¹) in saliva of COVID-19 patients with different outcomes	64

ARTICLE 2

Figure 1	Study Flowchart	79
Figure 2	Volcano plot and heatmap of COVID-19 patients	88
Figure 3	Normalized intensity values in Log 2 of COVID-19 saliva metabolites	89
Figure S1	Individual heatmap of severe COVID-19 versus moderate COVID-19, discharge versus death and no IMV versus IMV	90

TABLE LIST

ARTICLE 1

Table 1	Participant characteristics and personal history at hospital admission	60
Table 2	Clinical Data in Saliva Collection	62
Table 3	Outcome of the clinical characteristics of patients with COVID-19 analyzed at the end of hospitalization	63
Table 4	Discrimination of SVM Algorithms Applied to Classify Outcomes in COVID-19 Inpatients	65

ARTICLE 2

Table 1	Demographics and Baseline Characteristics of COVID-19 Patients	83
Table S1	Clinical Data on timepoint saliva collection	85
Table 2	List of differentiative metabolites among COVID-19 saliva patients	87

LIST OF ACRONYMS AND ABBREVIATIONS

SUMMARY, INTRODUCTION, THEORETICAL FOUNDATIONS AND PERSPECTIVES

FTIR	Fourier-transform infrared Spectroscopy
2019-nCoV	2019 novel coronavirus
WHO	World Health Organization
COVID-19	Coronavirus disease 2019
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
ATR-FTIR	Attenuated total reflection Fourier-transform infrared Spectroscopy
FUU	Federal University of Uberlandia
MERS-CoV	Middle East respiratory syndrome coronavirus
EVD	Ebola virus disease
ICTV	International Committee on Taxonomy of Viruses
ORFs	Open reading frames
S	Spike surface glycoprotein
M	Membrane protein
E	Envelope protein
N	Nucleocapsid protein
RdRp	RNA polymerase
PLpro	Papain-like protease
3CLpro	Coronavirus main protease
ER	Endoplasmic reticulum
ERGIC	Endoplasmic reticulum-Golgi intermediate compartment
ACE 2	Angiotensin Converting Enzyme-2
RT-PCR	Reverse transcription-polymerase chain reaction
cDNA	Complementary DNA
VOC	Variant of concern
PRPs	Proline-rich proteins
IgA	Immunoglobulin A

IR	Infrared
MIR	Mid-infrared
NIR	Near-infrared
FIR	Far-infrared
ATR	Attenuated total reflection
MS	Mass spectrometry
m/z	Mass/charge ratios
GC	Gas chromatograph
LC	Liquid chromatograph
Q-TOF	Quadrupole-time-of-flight
ESI	Electrospray ionization source
TCA	Tricarboxylic acid
IMV	Invasive mechanical ventilation
AI	Artificial intelligence
ML	Machine learning
SVM	Support vector machine

ARTICLES 1 e 2

ATR-FTIR	Attenuated total reflectance Fourier-transform infrared spectroscopy
COVID-19	Coronavirus disease 2019
IMV	Invasive mechanical ventilation
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
RT-PCR	Real-Time Polymerase Chain Reaction
SVM	Support vector machine
TPR	True positive rate
TNR	True negative rate
ICU	Intensive care unit
S	Spike
RBD	Receptor binding domain
IR	Infrared
HC-UFU	Clinical Hospital of Uberlandia
UMC	Uberlandia Medical Center

BMI	Body mass index
CKD	Chronic kidney disease
O ₂	Oxygen
NRB	Non-rebreather mask
ATR	Attenuated total reflectance
GLzM	General Linear Model
CHF	Chronic heart failure
SOFA	Sequential-organ-failure-assessment-sofa-score
NIV	Non-invasive ventilation
CT	Computer tomography
ACE 2	Angiotensin converting enzyme II
HPLC-MS	High-performance liquid chromatography hyphenated to the mass spectrometer
TCA	Tricarboxylic acid cycle
HPLC	Spectroscopic grade methanol
l/min	Litter per minute
kV	Kilo volt
MFE	Molecular feature extraction
MPP	Mass Profiler Professional
FC	Fold change
PLpro	Papain-like proteases
3CLpro	3C-like protease
6MP	6-mercaptopurine
6TG	6-thioguanine
PGE2	Prostaglandin E2
AA	Arachidonic acid
IL-6	Interleukin 6
TNF α	Tumor necrosis factor alpha
ARDS	Acute respiratory distress syndrome
25-HC	Oxysterol 25-Hydroxy-cholesterol
ROS	Reactive oxygen species
UP	Undecylprodigiosin
RDS	Respiratory detox shot

SUMMARY

Ficha catalográfica	ii
Folha de Aprovação	iii
Ata de Defesa	iv
Dedicatória	vi
Agradecimentos Especiais	vii
Agradecimentos	viii
Epígrafe	xi
Resumo	xii
Abstract	xiv
Illustration list.....	xvi
Table list	xvii
List of acronyms and abbreviations.....	xviii

1	INTRODUCTION	23
1.1	Initial considerations	24
2	THEORETICAL FOUNDATION	26
2.1	General aspects of COVID-19	26
2.2	Saliva: a promising biological fluid	32
2.3	Fourier transform infrared (FTIR) spectroscopy: technical principles, advantages and applications	36
2.4	Metabolomic	43
2.5	Machine Learning	46
3	OBJECTIVES	49
3.1	General Objectives	49
3.2	Specific Objectives	49
4	ARTICLE 1	50
5	ARTICLE 2	74
6	CONCLUSION	102
7	PERSPECTIVES	103
8	PATIENT REPORT	104
9	REFERENCES	105
	ATTACHMENTS	116
	APPENDIX	119

1 INTRODUCTION

In December 2019, several cases of pneumonia of unknown etiology were identified in Wuhan City, Hubei Province, China (Huang et al., 2020) and sequencing analysis from lower respiratory tract samples indicated a novel coronavirus, which was named 2019 novel coronavirus (2019-nCoV) by the World Health Organization (WHO) (WHOa, 2020).

The high transmissibility and rapid spread through the respiratory tract by droplets or direct contact, resulted in a world dissemination and a pandemic situation was declared on March, 2020 (MAJUMDER, MINKO, 2021). Globally, the coronavirus disease 2019 (COVID-19), caused more than 6.6 million deaths until Nov 26, 2022 (WHO, 2022), overwhelmed many healthcare systems and had a catastrophic effect on the global economy (CASCELLA et al., 2022).

Symptoms can range from a simple cold to a severe disease with deleterious effects on many organ systems as pulmonary, neurological, renal, hepatic, gastrointestinal, thromboembolic, cardiac, endocrine, and dermatological system, eventually leading to multi-organ failure (GUPTA et al., 2020).

During this time of pandemic scenario there has been a significant advance in clinical research, enabling a better understanding of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), however the increasing number of mutations limit the containment of the disease and its management remains a challenge for health teams (HODGSON et al., 2021; KANTARCIOGLU et al., 2021).

Several countries have already faced the fifth wave of outbreaks that can be attributed mainly to the emergence of mutant variants of the virus. Advances in vaccination reduced the occurrence of severe cases, but the high number of mutations have been related to increased ability of evade infection-blocking antibodies, high transmissibility and higher risk of reinfection (QUARLERI et al., 2022; KUMAR et al., 2022).

In this context, saliva is a promising biofluid which contains circulating biomolecules produced locally or even from a systemic disease that can modify the composition of saliva, and can be used for early diagnosis and also explored for

prognosis and severity of the disease (PFAFFE et al., 2011; CAIXETA et al., 2020; FERNANDES et al., 2020; KAZMER et al., 2022).

When we talk about prognostic models, we think in tools that can assist in decision making regarding therapeutic goals, choosing the appropriate treatment, better allocation of medical equipment and human resources, as well as adequate triage of these patients. On this purpose, technological platforms allied to machine learning are being used to determine prognostic salivary biomarkers for COVID-19, as attenuated total reflection Fourier-transform infrared (ATR-FTIR) Spectroscopy and salivary metabolomic. The expectation with this longitudinal study is to establish possible molecular salivary biomarkers with prognostic capacity for COVID-19.

1.1 Initial considerations

The format of this thesis follows the alternative model proposed by the Post Graduate Program in Health Sciences, Faculty of Medicine - Federal University of Uberlandia (FUU), which determines that the study results may be presented in the scientific articles format.

The thesis was organized into the following sections: **Theoretical Foundation**, presented as a form of literature review of the topics covered in the thesis; **Objectives**, exposes the purposes of the study; **Results**, includes two elaborated manuscripts; **Conclusion**, discusses the synthesis of the main results of the study; **Perspectives**, presents expectations for future studies; **Post-text**, includes references, attachments and appendices.

The first manuscript titled “**Salivary ATR-FTIR spectroscopy for COVID-19 prognosis**”, had as main objective the purpose to determine the predictive and prognostic value of algorithms developed for COVID-19 disease stratification by (i) the disease progression - moderate or severe COVID, (ii) discharge vs death, and (iii) the need of invasive mechanical ventilation, using saliva ATR-FTIR spectra and machine learning in early stages of hospitalization. It is intended to submit it to the Scientific Reports (Impact Factor = 4.996).

The second manuscript titled “**Saliva Metabolomics for COVID-19 Prognosis**”, had as main objective to analyze the saliva metabolites from COVID-19 patients for disease stratification by (i) the disease progression - moderate or severe COVID, (ii) discharge vs death, and (iii) the need of invasive mechanical ventilation,

performing the metabolomics of saliva collected in early stages of hospitalization. It is intended to submit it to the PLOS One (Impact Factor = 3.752).

2 THEORETICAL FOUNDATION

2.1 General aspects of COVID-19

During the last two decades, outbreaks of several viral diseases have been reported including the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002, H1N1 influenza in 2009, the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012, Ebola virus disease (EVD) in 2013, and Zika virus in 2015 (MAJUMDER, MINKO, 2021).

In December 2019, several cases of pneumonia of unknown etiology were identified in Wuhan City, Hubei Province, China, with an epidemiological link to the Huanan Seafood Wholesale Market where there was also sale of live animals (Huang et al., 2020). Sequencing analysis from lower respiratory tract samples indicated a novel coronavirus, which was named 2019 novel coronavirus (2019-nCoV) by the World Health Organization (WHO) on January 12, 2020 (WHOa, 2020).

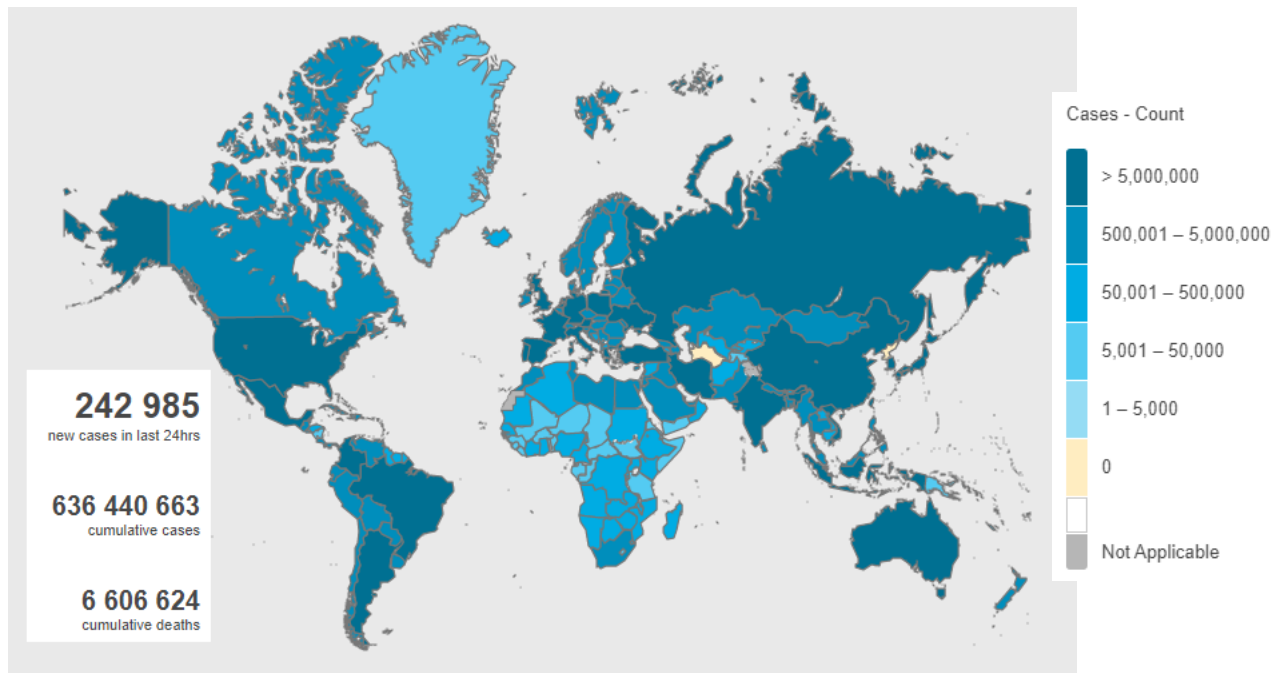
The 2019-nCoV is a betacoronavirus of group 2B with at least 70% similarity in genetic sequence to SARS-CoV (HUI et al., 2020; SU et al., 2016). It may originate from Chinese horseshoe bats, which are the natural reservoirs of SARS-CoV and the virus uses the human angiotensin converting enzyme II (ACE2) for cell entry (GE et al., 2013).

The International Committee on Taxonomy of Viruses (ICTV), which is responsible for developing the classification of viruses and taxon nomenclature of the family Coronaviridae, has assessed the placement of the human pathogen and designated it as SARS-CoV-2, recognized as a specie of severe acute respiratory syndrome-related coronavirus (GORBALENYA et al., 2020).

At the end of January 2020, the WHO declared this viral outbreak as a public health emergency of international concern and due to its rapid spread to several countries, a pandemic situation was declared on March 11, 2020 (EUROSURVEILLANCE, 2020; CUCINOTTA, VANELLI, 2020).

Globally, the coronavirus disease 2019 (COVID-19), resulted in 636.440.663 confirmed cases, including 6.606.624 deaths, reported to WHO until 26 November 2022 (WHO, 2022) (Figure 1).

Figure 1 Global overview of COVID-19

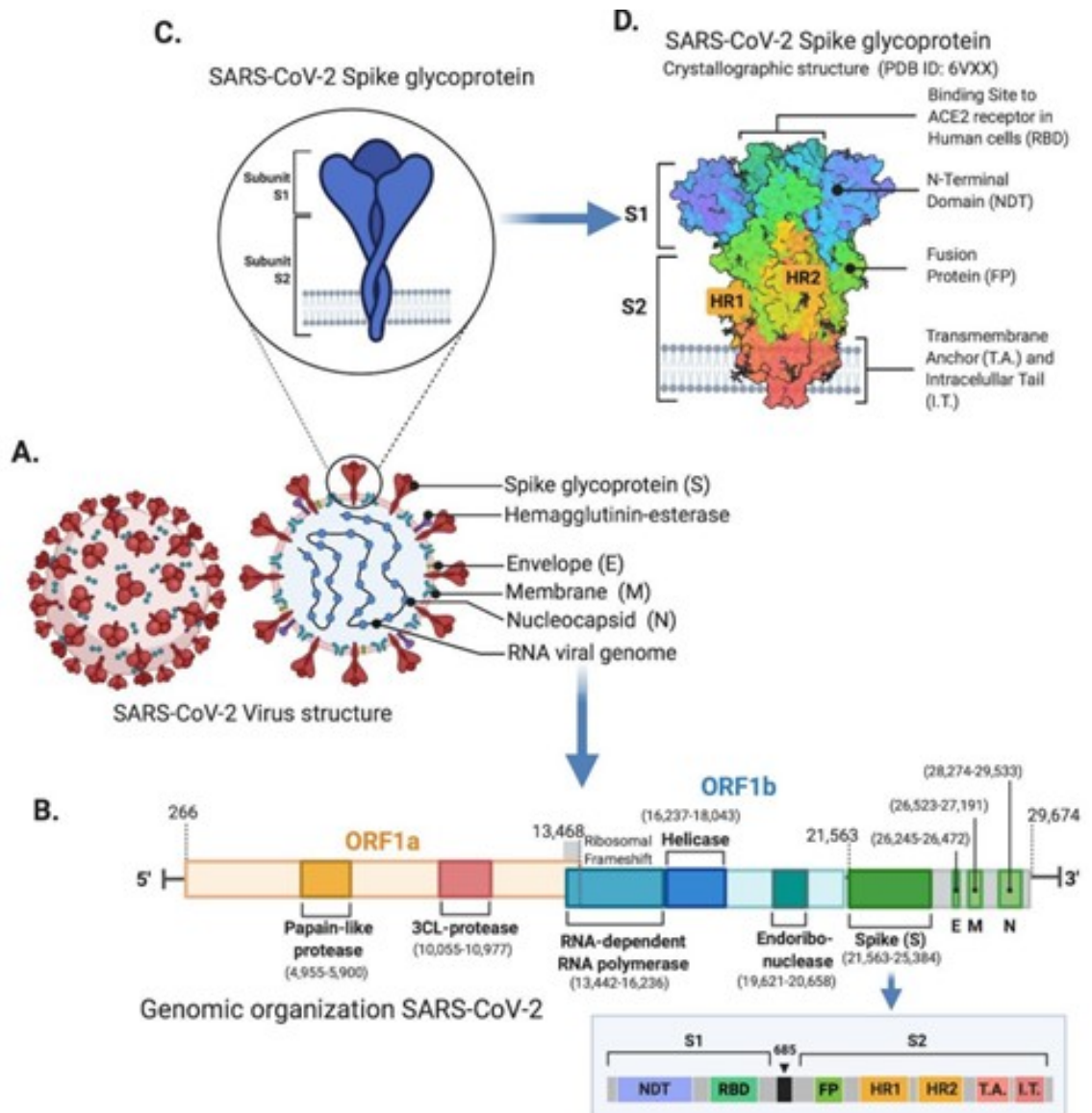


Reproduced from World Health Organization. <https://covid19.who.int/>. Accessed on 11/26/2022.

SARS-CoV-2 genome contains single-stranded positive-sense RNA encapsulated within a membrane envelop with an average diameter of 60-140 nm. The envelope is covered with glycoprotein spikes giving coronaviruses their crown-like appearance (corona is the Latin term for crown). The genome of SARS-CoV-2, composed of the open reading frames (ORFs), has a length of 27-32 Kilobases and encodes several structural and nonstructural proteins (Figure 2) (MITTAL et al., 2020; CHILAMAKURI, AGARWAL, 2021).

Four main structural proteins are encoded in SARS-CoV-2 viral genome, namely, spike surface glycoprotein (S), membrane protein (M), envelope protein (E), and nucleocapsid protein (N). These structural proteins are encoded by S, M, E, N genes at ORFs 10 and 11 on the one-third of the genome near the 3' -end. These mature structural proteins are responsible for viral maintenance and replication (MOUSAVIZADEH, GHASEMI, 2020). Several other nonstructural proteins are also encoded by the viral genome (RNA polymerase, RdRp; papain-like protease, PLpro; coronavirus main protease, 3CLpro) (MAJUMDER, MINKO, 2021).

Figure 2 Structure and genomic organization of SARS-CoV-2



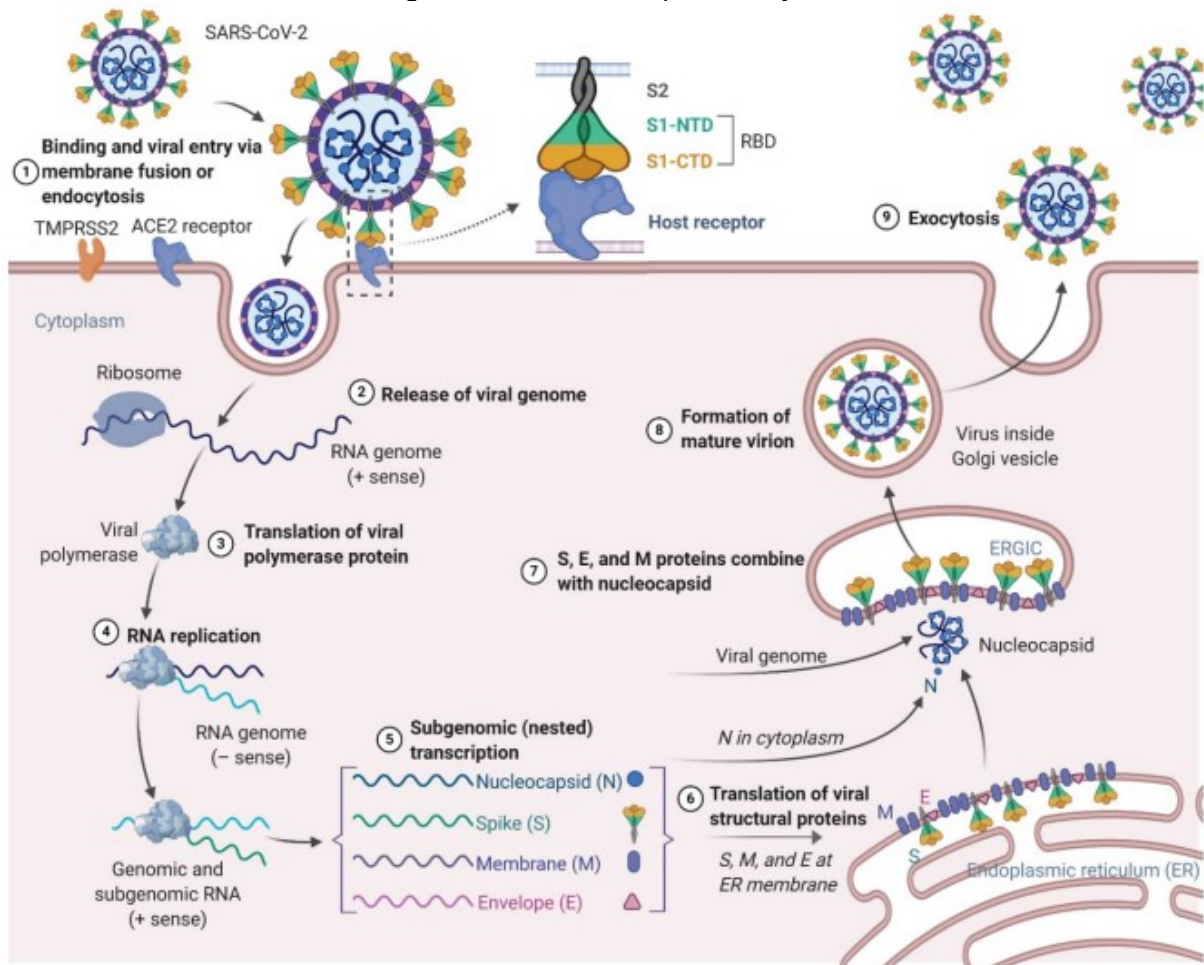
(A) Schematic representation of SARS-CoV-2 virus structure. (B) Genomic organization of SARS-CoV-2 representing ORF1a, ORF1b which encode for nonstructural proteins (papain-like protease, 3CL-protease, RNA-dependent RNA polymerase, helicase, and endoribonuclease). Genes coding for spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. Genomic positions are shown with dashed lines followed by nucleotide position number in RNA viral genome. The box highlights the genomic organization of spike (S) gene showing distinct S1 and S2 subunits coding segments. (C) Schematic magnified representation of SARS-CoV-2 spike glycoprotein showing S1 and S2 subunits. (D) Crystallographic structure of SARS-CoV-2 spike glycoprotein. Receptor binding domain (RBD) representing ACE2 receptor binding site in human cells, N-terminal domain (NTD), fusion protein (FP), transmembrane anchor (T.A.), and intracellular tail (I.T.) protein domains. Reproduced from CHILAMAKURI et al., 2021.

The structural proteins are responsible for host infection, membrane fusion, viral assembly, morphogenesis, and release of virus particles, among other functions,

and the nonstructural proteins facilitate viral replication and transcription (MITTAL et al., 2020).

The spike (S) protein of coronaviruses uses the angiotensin converting enzyme-2 (ACE 2) for cellular entry. The surface unit S1 binds to cellular proteases and allows fusion of viral and cellular membranes by the S2 subunit. Cell entry requires priming of the spike protein by the cellular serine protease TMPRSS2 which activity is essential for viral spread and pathogenesis (HOFFMANN et al., 2020). After endocytosis of the virus into the host cells and the release of viral genome, the host ribosome is used to translate viral polyproteins. Viral proteinases 3CLpro and PLpro cleave viral polyproteins into effector proteins. RNA-dependent RNA polymerase synthesizes a full-length negative strand RNA template which is used to replicate viral genomic. At that time, four essential structural viral proteins (N, S, M, and E) are produced by transcription then translation of the viral genome. N protein binds genomic RNA while S, M, and E proteins are integrated into the membrane of the endoplasmic reticulum (ER) forming ERGIC - endoplasmic reticulum-Golgi intermediate compartment. Assembled nucleocapsid with helical twisted RNA is encapsulated into the ER lumen, viral progeny is transported by the ERGIC toward the plasma membrane of the host cell, and finally, daughter virus is released by exocytosis (LIU et al., 2020; MAJUMDER, MINKO, 2021). A schematic representation of cellular uptake and replication of SARS-CoV-2 virus is shown in figure 3.

Figure 3 Coronavirus replication cycle



The figure describes membrane fusion, viral assembly, morphogenesis, and release of mature virion by exocytosis.

Reproduced from LIU et al., 2020.

The high transmissibility of SARS-CoV-2 mainly occurs via respiratory droplets and human-to-human skin contact, and the incubation period ranges from 2-14 days (mean of 6 days) (GAUTRET et al., 2020). Advanced age and the presence of comorbidities such as hypertension, diabetes, cardiovascular or pulmonary disease, obesity and cancer, were associated with disease severity and worse outcome (GAUTRET et al., 2020; STEFAN et al., 2020; RICHARDSON et al., 2020; GALLO et al., 2021).

The typical clinical symptoms of COVID-19 are fever, cough, myalgia or fatigue, dyspnea, and bilateral lung infiltrates on imaging with predominant distribution in peripheral and posterior regions (SONG et al., 2020; MAO et al., 2020). The clinical manifestation of symptoms is related to direct viral load and toxicity on

the affected organ or system, and body response to the viral attack, reflecting the symptoms reported by the patient (LI et al., 2020; GUPTA et al., 2020).

While all coronaviruses distress the respiratory tract, SARS-CoV-2 virus additionally is also associated to deleterious effects on many other organ systems as neurological, renal, hepatic, gastrointestinal, thromboembolism, cardiac, endocrine, and dermatological system, eventually leading to multi-organ failure (GUPTA et al., 2020).

Key mechanisms that may have a role in the pathophysiology of multi-organ injury secondary to ACE 2 mediated infection with SARS-CoV-2 include direct viral toxicity, endothelial cell damage and thromboinflammation, dysregulation of the immune response, and dysregulation of the renin–angiotensin–aldosterone system (GUPTA et al., 2020). In addition, the immune pathogenesis caused by the systemic cytokine storm and microcirculation dysfunctions may also occur secondary to viral sepsis (LI et al., 2020).

Among the various diagnostic tests for COVID-19 commercially available, reverse transcription-polymerase chain reaction (RT-PCR) stands out as the standard method (UDUGAMA et al., 2020). In general, test sample collected via nasopharyngeal or oropharyngeal swab undergo viral RNA extraction and its conversion into complementary DNA (cDNA) by the reverse transcription. Thereafter, the amplification of the DNA sample is performed in quantitative PCR, and the viral cDNA is detected by a fluorescent or electrical signal (VANGUILDER et al., 2008; MAJUMDER, MINKO, 2021).

The detectable viral load depends on the days after illness onset. In the first 14 days after onset, SARS-CoV-2 could most reliably be detected in sputum followed by nasal swabs, whereas throat swabs were unreliable 8 days after symptom onset. (UDUGAMA et al., 2020; YANG et al., 2020).

Although the RT-PCR test remains the reference gold standard in diagnosis of COVID-19 infection, this test has rigorous laboratory specifications and takes a long time to report results. Another concern is the false-negative findings that are related to improper sampling techniques, low viral load in the area sampled, or mutations in the viral genome (UDUGAMA et al., 2020; AI et al., 2020; WINICHAKOON et al., 2020).

As November 26, 2021 and until this publication, there were five SARS-CoV-2 lineages designated as the “variant of concern” (VOC), alpha, beta, gamma, delta, and omicron variants (KUMAR et al., 2022). The term VOC refers to the high mutational load in the virus and has been related to high transmissibility, increased viral loads and virulence, higher risk of reinfection, ability to escape natural immunity and decrease in effectiveness of public health measures (QUARLERI et al., 2022; KUMAR et al., 2022; CHOI e SMITH, 2021).

Vaccines against COVID-19 began to be administered by China in July 2020, followed by other countries such as the United States of America (USA), Russia and Europe in December 2020 and Brazil and India in January 2021. Currently, vaccines have resulted in a drop in mortality (WHO, 2022), but offer limited protection as viral mutations continue to occur (ZENG et al., 2022) and their effectiveness declines over time (ANDREWS et al., 2022).

Due to the high rate of infection and mortality caused by COVID-19, and its high mutational load, there is an urgent need to develop strategies for fast screening, diagnosis and prognosis, for quick isolation and clinical intervention (CAIXETA et al., 2021). Studies have shown the successfully use of saliva as an early diagnostic tool and also explored for prognosis and severity of many diseases, including COVID-19 (PFAFFE et al., 2011; CAIXETA et al., 2020; FERNANDES et al., 2020; KAZMER et al., 2022).

2.2 Saliva: a promising biological fluid

Saliva is a complex biological fluid that has been widely studied in relation to its structure, composition and function. The different branches of study encompass salivary biochemistry, genomics, epigenomics, transcriptomics, proteomics, metabolomics and microbiomics in an attempt to find diagnostic and monitoring biomarkers (DAWES and WONG, 2019; BEL'SKAYA et al., 2018).

Saliva is a clear, mucoserous, slightly acidic, from exocrine secretion, composed of 99% water into a viscoelastic solution, such as electrolytes (Na, K, Ca, Mg, hydrogen carbonates, and phosphates), mucus (mucins, proline-rich proteins (PRPs), cystatins, mucopolysaccharides and glycoproteins), antiseptic substances [hydrogen peroxide, Immunoglobulin A (IgA), lactoferrin, statherin, histatins], and

several enzymes (α -amylase, carbonic anhydrases, lysozymes, lingual lipase). The complexity of saliva can be seen by the number of components that compose it, and presents more than 3,000 proteins and glycoproteins in total saliva (HUMPHREY et al., 2001; KHAUSTOVA et al., 2010; PFAFFE et al., 2011; CARPENTER, 2013; DAWES e WONG, 2019).

Salivary fluid is mostly formed by the major and minor salivary glands, from gingival crevicular fluid and a little amount comes from some molecules that passes into saliva from blood, via passive diffusion (steroid hormones), active transport (polymeric IgA) and ultrafiltration (water, ions, catecholamines and steroids). The major salivary glands, which exist as bilateral pairs (parotid, submandibular and sublingual glands) are located in and around the mouth and throat, while the minor glands (around 600) are found in the lower lip, tongue, palate, cheeks and pharynx (HUMPHREY et al., 2001; PFAFFE et al., 2011). During unstimulated salivary flow, 20% comes from parotid, 65% from submandibular, 7 to 8% from sublingual, and less than 10% from the minor glands, however during stimulated high flow, parotid contribute with 60% of the flow (HUMPHREY et al., 2001; CARPENTER, 2013).

Parotid secretes a serous secretion, similar to water, rich in amylase and PRPs, while submandibular and sublingual glands secrete a mixture of mucoserous saliva and are less responsive in diet changes. Minor salivary glands are composed of mucins and a small amount of amylase, maintaining a mucin-rich layer adjacent to the mucosa and also shows no increase in salivary flow in response to food stimuli (CARPENTER, 2013).

Salivary glands are composed of specialized cells, the acini, which make the saliva, and the ductal cells, that modify and convey the saliva to the mouth, mediated through multiple ion transporters and channels (BAUM et al., 2011; PFAFFE et al., 2011; CARPENTER, 2013). Briefly, acinar cells secrete sodium and chloride ions by active transport into the ductal lumen of the gland after a neural signal from the brain. Acinar cells absorb water from the blood through aquaporin 3 and 5 channels and via the tight junctions. In the parotid and submandibular glands, the salt is mostly recovered by the striated ducts, which are impermeable to water and the final saliva is markedly hypotonic (BAUM et al., 2011; CARPENTER, 2013). The autonomic nervous system densely innervates the salivary glands. While parasympathetic nerve

impulses produce high-flow, low-protein saliva, sympathetic impulses produce low-flow, protein-rich saliva (CARPENTER, 2013).

Salivary functions can be listed as follows: lubrication and protection, buffering action and clearance, tooth integrity, antimicrobial activity with immunomodulation and anti-inflammatory properties, and taste and digestion (HUMPHREY et al., 2001).

It is well established that saliva reflects the physiological state of the body, including endocrine, nutritional, metabolic, and emotional condition. Circulating biomolecules that originate from a local or even systemic disease can be transported to the salivary glands, which consequently will modify the composition of saliva that can be used for early diagnosis and also explored for prognosis and severity of the disease. Among the various diseases studied through saliva, we highlight: caries, periodontal diseases, oral cancer, diabetes, cardiovascular, autoimmune and renal diseases, pancreatic, breast, lung and prostate cancers, and also COVID-19 (PFAFFE et al., 2011; MALAMUD, 2011; ABRAO et al., 2016; BUNACIU et al., 2015; ZHANG et al., 2016; FERREIRA et al., 2020; CAIXETA et al., 2020; FERNANDES et al., 2020; KAZMER et al., 2022).

It is known that some viral strains may be detected in saliva as long as 29 days after infection (SABINO-SILVA et al., 2020). In a similar way, SARS-CoV-2 RNA is highly detectable in the oropharyngeal cavity, and therefore saliva can be used as an excellent diagnostic fluid to control the spread of COVID-19 (FERNANDES et al., 2020). The presence of viral RNA was detected in more than 50% of the asymptomatic patients as well as in symptomatic patients before the onset of symptoms (NAGURA-IKEDA et al., 2020; FERNANDES et al., 2020). There are three possible pathways that explain the viral presence in saliva. Firstly, the virus can be found in the upper and lower airways, as the virus enters the oral cavity in the form of droplets present in gas exchange. Secondly, the SARS-CoV-2 present in the blood can access the mouth via crevicular fluid, a specific exudate produced by the sulcular epithelium of the oral mucosa, that contains local proteins derived from the extracellular matrix and serum-derived proteins. Finally, infection of the major and minor salivary glands by COVID-19 can occur, with the release of viral particles in the saliva via salivary ducts (SABINO-SILVA et al., 2020).

Another factor that makes saliva the ideal fluid for viral detection is the presence of ACE 2 receptors in the salivary glands, which are fundamental in the process of viral coupling and replication in host cells, making these glands the perfect target for SARS-CoV-2 (FERNANDES et al., 2020). Nagura Ikeda et al., showed that the detection of viral RNA in saliva was higher in samples collected within 9 days of symptom onset, since the viral load of SARS-CoV-2 declines over time (NAGURA-IKEDA et al., 2020; FERNANDES et al., 2020).

Saliva is a biofluid that can be obtained non-invasively, with minimal discomfort, through rapid collection and adequate safety in the context of the COVID-19 pandemic (FERNANDES et al., 2020). The collection method can even be self-collected, reducing the risks of exposure of health professionals (NAGURA-IKEDA et al., 2020; CAIXETA et al., 2021).

SARS-CoV-2 has an important global impact due to high mortality rates, overload of health systems, greater need for medical and human inputs, in addition to its high transmission potential. This context demonstrates the real need for rapid and accurate platform diagnoses, as well as the early identification of the prognosis of the disease. Prognostic models are tools that can assist in decision-making regarding therapeutic goals, choosing the appropriate treatment, accurate allocation of hospital medical equipment and triage of these patients.

In this context, attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy is being used in an attempt to detect a “fingerprint” of biofluids samples that could predict severity in patients with COVID-19. ATR-FTIR spectroscopy look for biomarkers that allows the detection and quantitative analysis of biochemical characteristics of saliva in mid-infrared (MIR) region which identifies components with highly specific bands in a particular set of wavenumbers depending on the molecular composition and structure (KHAUSTOVA et al., 2009).

ATR-FTIR as a non-invasive, non-destructive and inexpensive method, combined with the many biochemical benefits of saliva as a biological fluid for disease diagnosis and prognosis applied to the COVID-19 study shows great potential and importance. ATR-FTIR spectroscopy proved to be valuable in the detection of SARS-CoV-2 in the saliva of infected patients, and can be better explored regarding its prognostic potential and determination of disease severity.

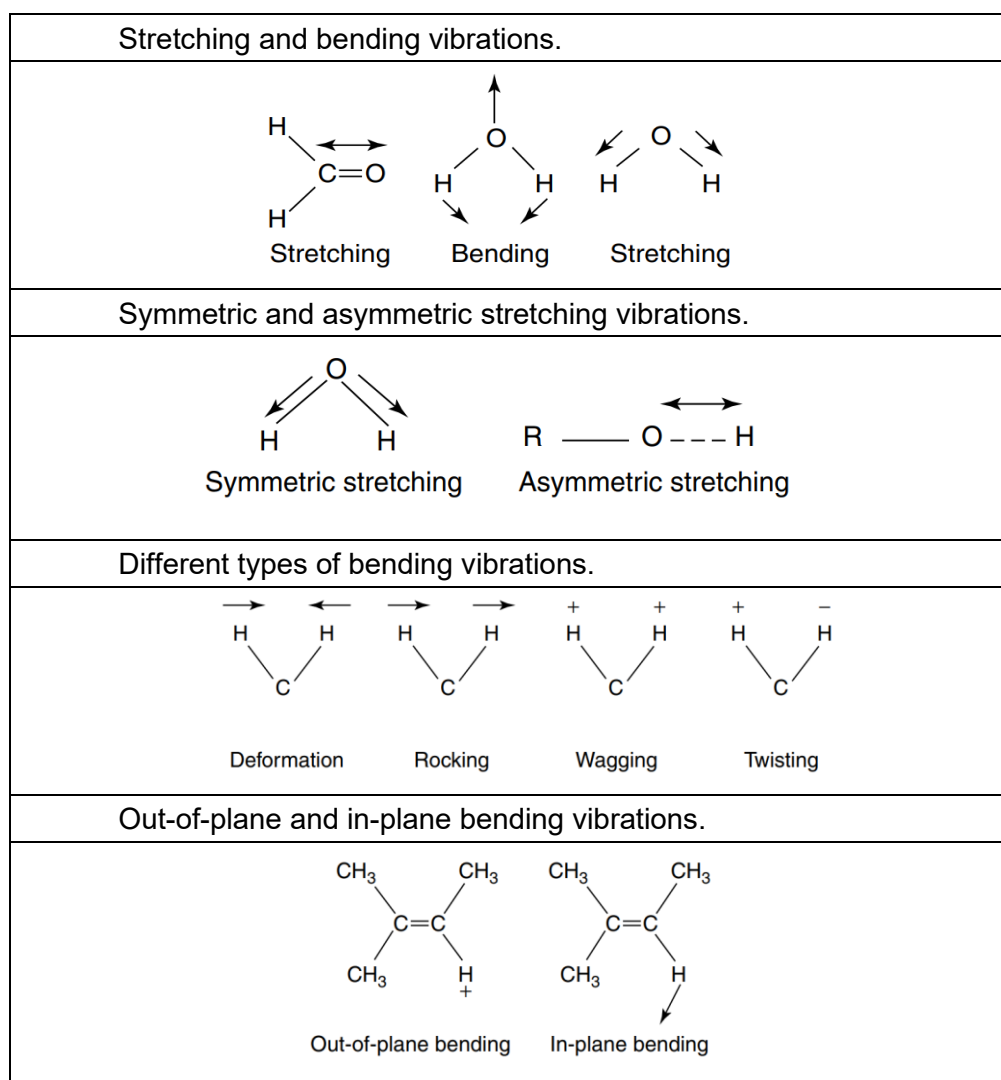
2.3 Fourier transform infrared (FTIR) spectroscopy: technical principles, advantages and applications

Fourier transform infrared (FTIR) spectroscopy is a method of vibrational spectroscopy used for qualitative and structural analysis that identifies structural molecules based on their infrared (IR) absorption (CAIXETA et al., 2020; BELSKAYA et al., 2018).

The IR spectroscopy is related to the fact that molecules absorb light in the IR region of the electromagnetic spectrum, converting it into molecular vibration. The wavelength of the incident IR radiation absorbed depends on the atoms involved and the strength of any intermolecular interactions. Changes in the electric dipole moment of a molecule during vibration results in an infrared-activated molecule, whereas IR spectroscopy is quantified and directly correlated to the biochemical composition. The greater the change in electric dipole moment, the more intense will be the IR absorption band. Chemical compounds absorb IR radiation and the frequencies and intensities of the IR bands characterize the material, and its IR spectrum can be used to identify and quantify the particular substance in an unknown sample. The final IR absorption spectrum of a specific molecule is called the 'fingerprint' of the sample. (PUTZIG et al., 1994; STUART, 2004; ELLIS, GOODACRE, 2006; MOVASAGHI et al., 2008; CLEMENS et al., 2014).

The absorbed IR radiation causes vibration on the functional groups in different ways, like stretching, bending, deformation or combined vibrations, and involves changes mainly in bond length (stretching) or bond angle (bending). If a molecule has similar terminal atoms, bonds can stretch in-phase (symmetrical stretching) or out-of-phase (asymmetric stretching). Bending vibrations includes scissoring (or deformation), rocking, wagging and twisting (Figure 4). In case of different terminal atoms, then the vibrations will have varying proportions of stretching motion according to the strength of bonds between atoms and changes in the electric dipole moment during vibration (STUART, 2004; ELLIS e GOODACRE, 2006).

Figure 4 Types of vibrations on the functional groups



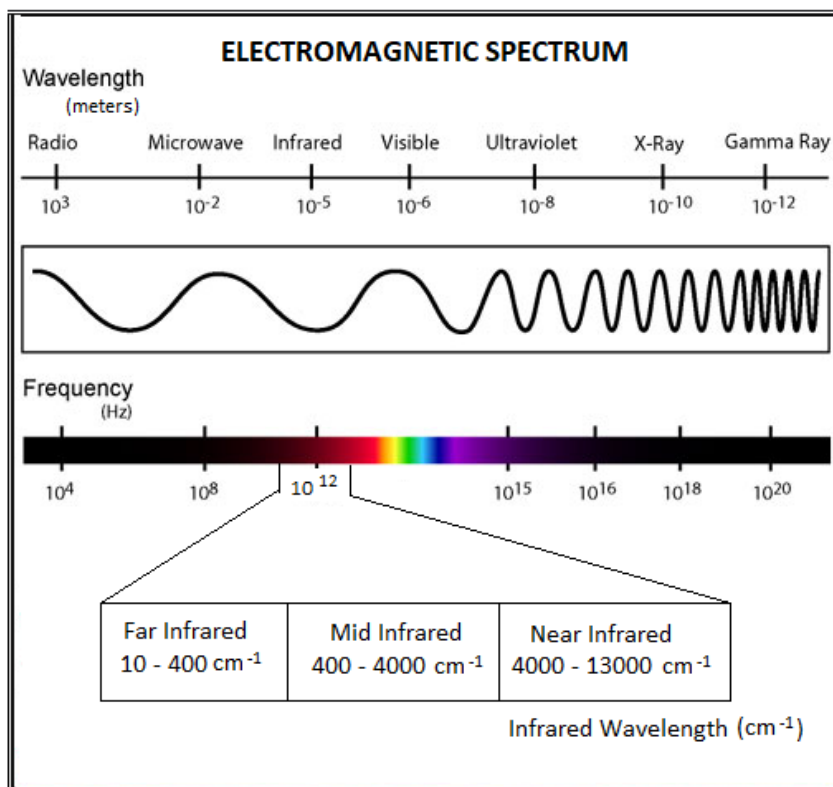
Reproduced with modifications from STUART, 2004.

The region of the IR radiation of electromagnetic spectrum is located between the visible (about 13000 cm^{-1}) and the microwave wavelengths (about 10 cm^{-1}) (Figure 5) and is conventionally divided in near-infrared (NIR) - 13000 to 4000 cm^{-1} , mid-infrared (MIR) - 4000 to 400 cm^{-1} and far-infrared (FIR) - 400 to 10 cm^{-1} (BARTH, 2007; HUCK, 2016).

NIR spectroscopy basically is associated to a specific fingerprint or spectrum as a result of the absorption of parts of the incoming IR radiation by solid, liquid or gaseous samples. The functional groups (C-H, C-O, C=O, N-H and O-H) when excited perform stretching, bending, deformation and scissor-vibrations giving qualitative and quantitative parameters such as physico-chemical parameters. However, the overtones or combinations of the fundamental stretching bands results

in weak intensity and overlapped bands. In contrast, the MIR region contains only fundamental vibrations (“signatures”) of bio samples with a clearer spectrum, sharp peaks and very rich information which allows the identification of highly sensitive detectors or biomarkers. The FIR regions are less used but also provide relevant information about heavy atoms, molecular skeleton vibrations, molecular torsions and crystal lattice vibrations (BARTH, 2007; HUCK, 2016; ELLIS e GOODACRE, 2006; STUART, 2004).

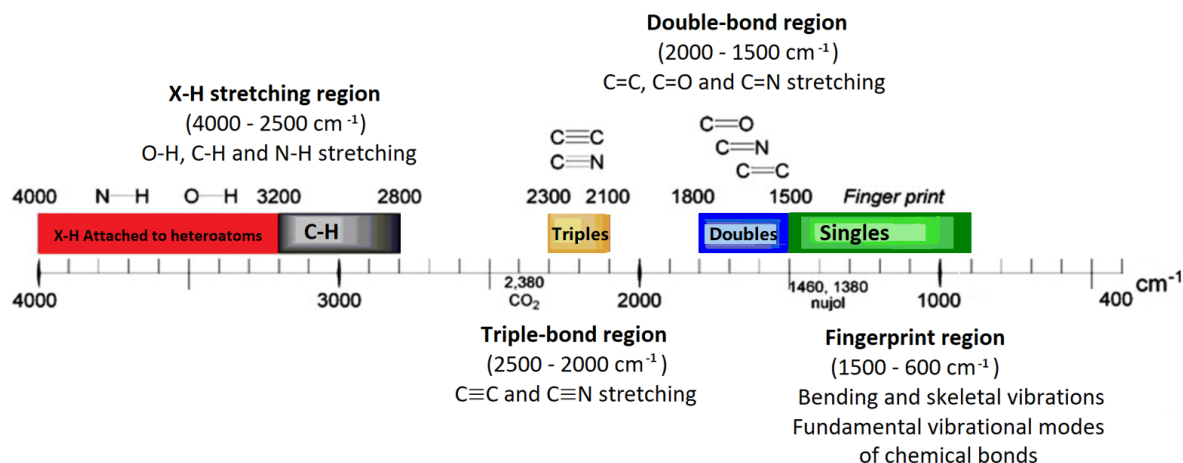
Figure 5 The electromagnetic spectrum



Reproduced with modifications from NOREEN et al., 2012.

Didactically, the MIR spectrum ($400 - 4000 \text{ cm}^{-1}$) can be organized into four main regions (Figure 6): the fingerprint region ($600 - 1500 \text{ cm}^{-1}$), the double-bond region ($1500 - 2000 \text{ cm}^{-1}$), the triple-bond region ($2000 - 2500 \text{ cm}^{-1}$), and the X–H stretching region ($2500 - 4000 \text{ cm}^{-1}$) (STUART, 2004).

Figure 6 Major absorptions of bonds in organic molecules in the MIR region (4000 - 400 cm^{-1})



Modified from NOREEN et al., 2012.

The spectrum of a molecule can be formed by complex vibrational movements and result in a hundred or more absorption bands, without the need to identify the vast majority of these bands. The final spectrum can be considered a "fingerprint" of the molecule and therefore this region is called the fingerprint region (650 - 1500 cm^{-1}), important for biological samples, and contains the key chemical bonds as bending and skeletal vibrations. The double-bond region (1500 - 2000 cm^{-1}) are due to C=C, C=O and C=N stretching, and contains important regions such as amide I /amide II region from 1500 – 1750 cm^{-1} . The triple-bond stretching absorptions (2000 – 2500 cm^{-1}) results in a constant strong force of the bonds. C $\equiv$ C bonds absorb IR radiation between 2050 and 2300 cm^{-1} , while C $\equiv$ N occurs at the 2200 – 2300 cm^{-1} . In addition, the fundamental vibrations in the 2500 – 4000 cm^{-1} region are generally due to O-H, C-H and N-H stretching. The functional groups of a sample can be identified due to their location in the spectrum (STUART, 2004; ELLIS e GOODACRE, 2006; MOVASAGHI et al., 2008; BAKER et al., 2014).

The most relevant spectral information can be assigned to molecular families, represented in Figure 7. Lipids and fatty acyl chains are located in 2800 - 3050 cm^{-1} , proteins are found in 900 - 1750 cm^{-1} and saccharides are in regions 900 - 1200 cm^{-1} (NOREEN et al., 2012).

[illegible]

FTIR spectroscopy is a vibrational method that studies the absorption of electromagnetic waves in MIR region ($4000 - 400 \text{ cm}^{-1}$) widely used to analyze biological samples (CHAN e KAZARIAN, 2016). A source emits electromagnetic radiation that passes through an interferometer. The Michelson interferometer used in FTIR spectrometry consists of two perpendicular plane mirrors, one of which can move perpendicularly to the plane. The radiation passes through the interferometer to the sample and arrives at a detector. The signal is then amplified, and high frequency interference is eliminated by a filter. The data is converted to the digital version by an analog-to-digital converter, being transferred to the computer for Fourier transformation (Figure 8) (STUART, 2004).

```

graph LR
    Source[Source] --> Interferometer[Interferometer]
    Interferometer --> Sample[Sample]
    Sample --> Detector[Detector]
    Detector --> Amplifier[Amplifier]
    Amplifier --> ADC[Analog-to-digital converter]
    ADC --> Computer[Computer]

```

40

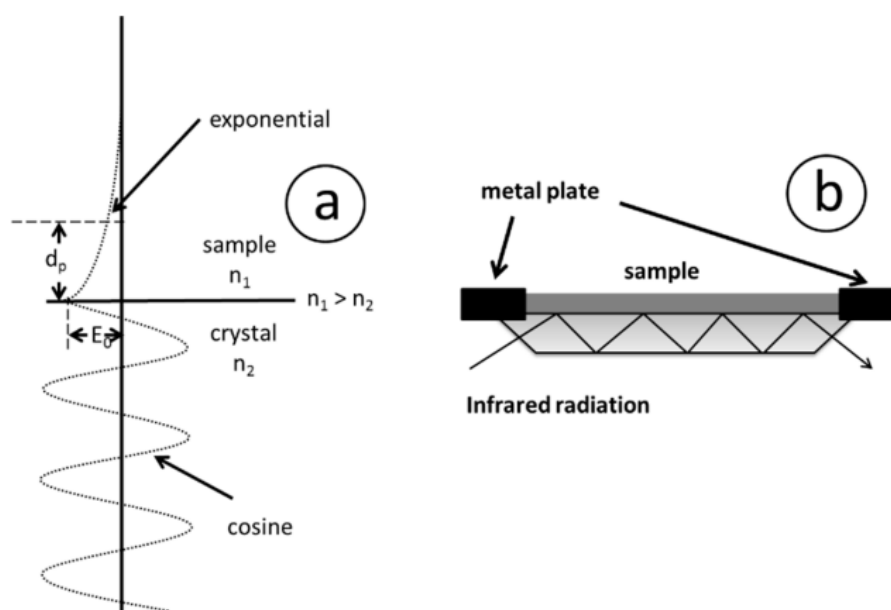
The light intensity measured by the Michelson interferometer is relative to the position of the moving mirror and it results in the interferogram, which is the first Fourier transform of the spectrum. A second Fourier transform, made by the computer, convert data back to spectrum (BARTH, 2007). In general, the Fourier transform extracts by mathematical methods the individual frequencies of the interferogram for final representation in an IR spectrum (BEEKES et al., 2007).

The configuration of the device for the acquisition of FTIR spectra can be done in different ways and the most used are: transmission, transflection or attenuated total reflection (ATR). The spectral analysis by reflection is generally more used in relation to the transmission ones, because the spectra obtained by transmission must be computationally corrected due to the occurrence of physical interferences since the IR radiation passes through the sample before reaching the detector (KAZARIAN e CHAN, 2013). In transflection mode, detection of absorbed IR radiation occurs after it has been transmitted through the sample, reflected and retransmitted back to the material. As a result, there is the formation of standing wave perpendicular to the surface of the sample, which cause spectral changes unrelated to the biochemistry of the material (BUNACIU et al., 2015; LIMA et al., 2015).

The ATR technique operates on the phenomenon of total internal reflection. The sample is placed in direct contact with a crystal (diamond, germanium, zinc selenite or thallium-iodide) with a higher refractive index than the sample, inducing multiple internal reflections of the incident radiation, which is attenuated and penetrates the sample as an evanescent wave (Figure 9a). The evanescent wave passes through the detector and provides a single spectrum, which represents the average signal from the sample area that light passed through (Figure 9b). The penetration depth of radiation into the sample can be controlled by changing the incidence angle of the radiation and allows measurement from liquid, solid or gas state. (BARTH, 2007; CLEMENS et al., 2014; BUNACIU et al., 2015; LIMA et al., 2015; HUCK, 2016).

The spectrum obtained by the ATR-FTIR analysis can be differentiated, through the first and second derivatives, or smoothed in order to reduce interferences and increase the accuracy of the analysis (STUART, 2004).

Figure 9 (a) Illustration of the evanescent wave; (b) principle of attenuated total reflection (ATR). d_p = depth of the evanescent wave; n = refraction index.



Reproduced from HUCK, 2016.

ATR-FTIR does not require special substrate or sample preparation, is a simple, fast, low cost, reagent-free, allows rapid and non-invasive detection of biochemical changes at the molecular level, and can be used in a variety of biological samples as tissue, cells, microorganisms, fingernails, hair, blood, and human secretions (saliva, urine, feces, tears, earwax, nasal mucus, vaginal mucus, breast milk and semen) (ELKINS, 2011; ORPHANOU et al., 2015; DEPCIUCH et al., 2016; BEL'SKAYA et al., 2018; DAS et al., 2021).

Considering that a molecule is determined by its unique structure, each biomolecule will exhibit its own ATR-FTIR spectrum ("fingerprint"), which represents the vibrational mode of this structural bond (STUART, 2004; ELLIS, GOODACRE, 2006; MOVASAGHI et al., 2008). The method's sensitivity to small changes in the biochemical composition of the sample makes ATR-FTIR the ideal technology as a diagnostic tool and promising as well as a prognostic method (BEL'SKAYA et al., 2021; KAZMER et al., 2022).

In the medical field, ATR-FTIR spectroscopy has been widely used for monitoring, diagnostic and prognosis purposes in different pathological conditions such as cancer (BUNACIU et al., 2015; BUTLER et al., 2019; FERREIRA et al.,

2020), diabetes (CAIXETA et al., 2020), chronic kidney disease (RODRIGUES et al., 2019), arthritis (ELLIS e GOODACRE, 2006), depressive disorders (DEPCIUCH et al., 2016), hepatitis, human immunodeficiency virus infection (LUKOSE et al., 2021), among many others.

The clinical applicability of ATR-FTIR in COVID-19 was reported as a diagnostic tool using saliva samples (MARTINEZ-CUAZITL et al., 2021; WOOD et al., 2021; NASCIMENTO et al., 2022; KAZMER et al., 2022), nasopharyngeal swab (BARAUNA et al., 2021; NOGUEIRA et al., 2021), plasma (ZHANG et al., 2021), RNA extracts (KITANE et al., 2021) and successfully distinguished healthy controls from positive COVID-19 cases with high specificity and sensitivity. Banerjee et al., investigated the potential of ATR-FTIR as a rapid blood test for classification of COVID-19 disease severity triage tool at the point of admission, correlating data to clinical parameters (age, sex, diabetes mellitus, and hypertension) (BANERJEE et al., 2021), but none prognostic studies were found using saliva sample and this method.

Kazmer et al., reinforce that further evaluation may establish the utility of ATR-FTIR for COVID-19 prognosis (KAZMER et al., 2022). The proposal to use saliva by ATR-FTIR spectroscopy as a determinant of patient outcome and prognosis in COVID-19 has not yet been elucidated in the literature.

2.4 Metabolomic

Metabolomics is the global search for metabolites, which are small molecules resulting from the metabolism of various substances by the body. The metabolic profile directly reflects the metabolic state of organs and tissues due to continuous exchange of metabolites between fluids and tissues. The metabolomic approach has been widely used to identify metabolites with a diagnostic and prognostic profile and to guide personalized treatment strategies (BLASCO et al., 2020). In this context, mass spectrometry (MS) is used to detect, quantify and characterize a large number of compounds simultaneously, by measuring its mass and identifying its chemical structure. The technique consists of ionizing simple molecules or complex structures, and separate efficiently ions based on different mass/charge ratios (m/z). The determination of the exact molecular weight of the sample components is used to

identify unknown compounds, and determine structure and chemical properties of molecules (SILVERSTEIN et al., 2005).

In summary, the main components of a mass spectrometer are: 1) Sample insertion: direct (syringe attached to an infusion pump) or hyphenated [Gas (GC) or Liquid (LC) Chromatograph]; 2) Ionization source: transform uncharged sample into charged species (ions). There are several ionization techniques, depending on the phase (solid, liquid, gas) of the sample. 3) Mass analyzers: contains electric and magnetic fields, which exert forces on ions traveling through these fields, increasing or decreasing the speed of a charged particle and changing the direction of the ion's trajectory. The magnitude of the deflection of the moving ion's trajectory depends on its m/z ratio; 4) Detectors: records the abundance of each ion type to determine the chemical element composition, and transform the chemical signal into an electrical signal; 5) Controllers: software that manage and process data acquisition by transforming the electric signal into graphs such as mass spectra (GATES, 2007).

In this study, analyzes were carried out using a high-performance liquid chromatography hyphenated to a high-resolution quadrupole-time-of-flight (Q-TOF) mass spectrometer with electrospray ionization source (ESI). Liquid chromatography enables full automation for large-scale analysis and provides greater sensitivity, selectivity, and greater structural determination capability. The Q-TOF mass spectrometer offers high ion-transmission efficiency, faster scan speed and high sensibility (SCHNEIDER et al., 2016; LEHMANN, 2017). The ESI allows sample analysis dissolved in solvents in a continuous flow, where the analyte solution is sprayed from a nozzle needle and put at high voltage. The ions are formed when the spray droplets cool down by rapid evaporation of solvent into a high vacuum region of the mass spectrometer. It is a method that allows ionization of labile and thermally unstable molecules, and works well for both small and large biomolecules (GATES, 2007; LEHMANN, 2017).

Studies involving multiomic approaches, such as metabolomic, search for diagnostic and prognostic biomarkers for COVID-19, and try to fill in the gaps of the metabolic disturbances caused by the viral infection (BATTAGLINI et al., 2022; BLASCO et al., 2020; POZZI et al., 2022). The clinical applicability of metabolomic in COVID-19 was reported in several studies with diagnostic purposes, disease evolution, need of hospitalization and comparison between groups with mild,

moderate, or severe disease via plasma, urine, serum, saliva and sebum (SU et al., 2020; SONG et al., 2020; SHEN et al., 2020; BI et al., 2022; SPICK et al., 2022; POZZI et al., 2022).

Plasma metabolomic analysis showed dysregulation in biofluids after SARS-CoV-2 infection, with a disproportionate reduction in nutrient circulation (SU et al., 2020). Abnormally high levels of ketone bodies in serum analysis (acetoacetic acid, 3-hydroxybutyric acid and acetone), was related to liver injury and dyslipidemia, with impairment in hepatic glutathione synthesis and oxidative stress (BRUZZONE et al., 2020). Furthermore, groups of amino acids, carbohydrates, fatty acids, and glycerophospholipids showed differences according to the severity disease, with macrophage dysregulation, tryptophan-kynurenine metabolism activation, platelet degranulation, massive metabolic suppression and complement system pathways disturbances in COVID-19 patient sera and plasma samples (SHEN et al., 2020; BLASCO et al., 2020; THOMAS et al., 2020; SONG et al., 2022).

The tricarboxylic acid (TCA) cycle, also called Krebs cycle and citric acid cycle, showed metabolite depletion and was related to disease severity, confirmed by multiple independent studies (SU et al., 2020; WU et al., 2020; SHEN et al., 2020; WANG et al., 2021). Triglycerides and free fatty acids (arachidonic and oleic acid), activation of gluconeogenesis and the metabolism of porphyrins were also related to disease severity, while non-critical COVID-19 patients showed a lipid disorder and a dysregulation in phenylalanine, tyrosine and tryptophan biosynthesis (BARBERIS et al., 2020).

Saliva metabolomics are still under explored in the study of viral infections (COSTA JUNIOR et al., 2020; SAPKOTA et al., 2020). In a comparative analysis between serum, saliva and sebum, saliva had the worst diagnostic performance, with a sensitivity of 0.78 and a specificity of 0.83 (SPICK et al., 2022). A recent study used salivary metabolites as diagnostic chemical markers for COVID-19 using colorimetric electronic tongue (BORDBAR et al., 2022). Studies that evaluated the saliva metabolomic in patients with COVID-19 allowed the identification of metabolites that differentiate patients by the need of hospitalization, and differences in oral microbiota with SARS-CoV-2 infection (POZZI et al., 2022; IEBBA et al., 2021; HAO et al., 2022).

In general, a better understanding of the pathophysiological fingerprint and pathways of COVID-19 make metabolomics-related approaches an important tool for monitoring disease progression and for therapeutic treatments (COSTA JUNIOR et al., 2020). However, SARS-CoV-2 salivary metabolomic fingerprint are still under explored, and no studies were found investigating which metabolites are likely presented in saliva of patients who progressed to death or need of invasive mechanical ventilation (IMV).

2.5 Machine Learning

Artificial intelligence (AI) consists of computational mechanisms that rely on human behavior to solve problems. AI systems use publicly available data or private platforms that serve as training for AI algorithms. Machine learning (ML) is a branch of AI where computer systems are able to automatically learn and progressively improve the performance of a task from patterns and observations in data, or make the most appropriate decision based on pattern recognition within the analyzed data, without being explicitly programmed (SAMUEL, 1959; SCHLEDER et al., 2021). ML algorithms have shown potential to contribute to the integration of data, extraction of information and identification of complex and non-linear relationships, improving predictive performance models developed to assist decision-making in health (TOPOL, 2019).

In general, ML can be divided into three categories: supervised, unsupervised, and reinforced. Supervised learning uses a labeled dataset to train the algorithm by establishing the relationship between samples and their labels. This model allows algorithms to learn from historical or training data and apply them to unknown inputs to derive the correct output. Unsupervised learning is used to identify new patterns and is performed on unlabeled data. The model independently finds features and patterns, using clusters by the similarity in their characteristics, or other organization without previous classification. Reinforcement learning is a behavior-based learning, which an autonomous agent learns about its interactive environment through a programmed rewards system (WANG et al., 2022).

Supervised learning can be divided into classification and regression. Classification encompasses output data defined as a categorical or nominal finite set,

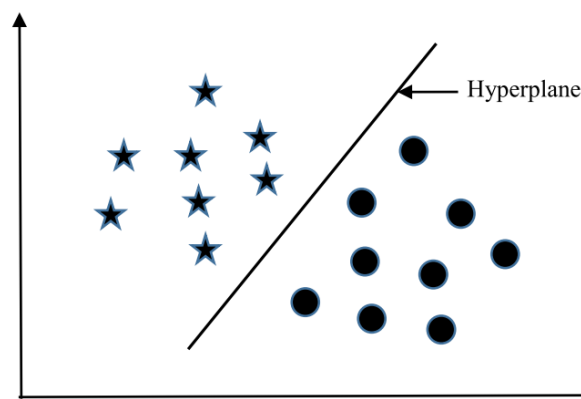
while in regression, models predict a continuous numerical output based on the predictors (SCHLEDER et al., 2021).

The supervised learning process goes through the training stage and subsequent validation of the predictive model. In the training process a learning algorithm is applied to a part of data set to create a function that maps the characteristics of predictor variables in relation to desired outcome. The validation phase consists of applying the function to verify the algorithm performance in the portion of data set that was not used in training. This process is repeated several times to obtain the average behavior of the model, and the function is considered adequate if it has a percentage of success or acceptable statistical metrics in the validation stage (MITCHELL, 1997).

The comparison of algorithms with different characteristics allows evaluating different performances in the search for the best predictive model. In addition, the choice of a classifier must consider computational cost, model interpretability and clinical plausibility. Several supervised machine learning algorithms can be used for disease prediction, such as logistic regression, decision tree, random forest, support vector machine (SVM), among others (UDDIN et al., 2019).

SVM model has been shown to be superior to other regression techniques as it finds both linear and nonlinear global solutions and high dimensional input support vectors, which are representative of the problem, in addition to high accuracy prediction and less spectral noise (THISSEN et al., 2004). In a first step, the algorithm maps each data item into a n-dimensional feature space, according to the number of features. After that, it identifies the line, named as hyperplane, used to separate the data items into two classes with maximal distance for both classes to minimize errors classification. Each data point is plotted as a point in the n-dimension space (where n is the number of features) and the value of each feature is a specific coordinate. Figure 10 presents a simple illustration of an SVM classifier (UDDIN et al., 2019).

Figure 10 A simplified illustration of how the support vector machine works.



The SVM has identified a hyperplane (actually a line) which maximizes the separation between the 'star' and 'circle' classes.

Reproduced from UDDIN et al., 2019.

The use of AI creates a high-performance medicine scenario, where the impact can be seen for healthcare teams through accurate interpretation of images, for health systems, by improving workflow and the potential for reducing medical errors, and for patients to use their own data to promote health (TOPOL, 2019; BARBERIS et al., 2022).

ML has been a valuable tool in the study of COVID-19 and has been widely used in the search for diagnostic and prognostic biomarkers. A recent study using plasma samples determined key proteins and metabolites and built an algorithm with a 92% accuracy for survival prediction based on the day of hospitalization of COVID-19 patients (RICHARD et al., 2022). Several studies investigated ML, COVID-19 and its relation to sex (CELAYA-PADILLA et al., 2021), and multiomic approaches in hospitalized patients (CASTAÑÉ et al., 2022; LIPMAN et al., 2022; D'ALESSANDRO et al., 2021; DELAFIORI et al., 2021; GANGLOFF et al., 2021). Only one study was found associating salivaomics data and ML to predict severity in COVID-19 patients (WANG et al., 2022).

We believe that the integration of technological platforms such as ATR-FTIR and metabolomic approaches with machine learning can have the potential to greatly enhance prognosis in COVID-19, and be used to identify more effective therapy targets, and accurate disease outcome prediction.

3 OBJECTIVES

3.1 General Objectives

The aim of the study was to use ATR-FTIR and metabolomic platforms in an attempt to find prognostic salivary biomarkers for COVID-19, which could be identified at early stages and used as prognostic predictors of disease severity and outcome determination.

3.2 Specific Objectives

- Search for clinical conditions that can predict severity based on personal history, comorbidities and clinical data on hospital stay (Article 1 and 2);
- Develop predictive algorithms for COVID-19 disease stratification by the disease progression – moderate or severe COVID-19, death or discharge, and need of invasive mechanical ventilation using saliva ATR-FTIR and machine learning (Article 1);
- Analyze the saliva metabolites from COVID-19 patients by the disease progression – moderate or severe COVID-19, death or discharge, and need of invasive mechanical ventilation using saliva metabolomics (Article 2).

Salivary ATR-FTIR spectroscopy for COVID-19 prognosis

Debora Lara Zuza Scheucher¹, Luciana Oliveira e Silva¹, Alinne Tatiane Faria Silva¹, Natalia Cristina Martins Nunes Mendes¹, Mario Machado Martins², Murillo Guimarães Carneiro³, Robinson Sabino-Silva⁴, Yara Cristina de Paiva Maia⁵, Thulio Marquez Cunha^{5*}

¹ Health Sciences Program, Faculty of Medicine, Federal University of Uberlandia, Uberlandia, 38405-320, Brazil.

² Laboratory of Nanobiotechnology, Institute of Genetics and Biochemistry, Federal University of Uberlandia, Uberlandia, 38405-320, Brazil.

³ Artificial Intelligence in Networks Group, Laboratory of Bioinspired Computing, Faculty of Computing, Federal University of Uberlandia, Uberlandia, 38408-100, Brazil.

⁴ Innovation Center in Salivary Diagnostic and Nanobiotechnology, Department of Physiology, Institute of Biomedical Sciences, Federal University of Uberlandia, Uberlandia, 38400-902, Brazil.

⁵ Faculty of Medicine, Federal University of Uberlandia, Uberlandia, 38400-902, Brazil.

* Corresponding author

Thulio Marquez Cunha

Faculty of Medicine, Federal University of Uberlandia, Uberlandia, 38400-902, Brazil

E-mail: thcunha@yahoo.com.br

ABSTRACT

Several studies revealed the importance of attenuated total reflectance Fourier-transform infrared (ATR-FTIR) spectroscopy on salivary diagnostic approaches for COVID-19. However, predictive prognostic salivary biomarkers of disease severity and outcomes remain unknown. The objective of this study was to develop predictive and prognostic value models of machine learning algorithms based on salivary infrared spectral signatures to the prognosis of (i) moderate vs severe COVID-19, (ii) discharge vs death, and (iii) no need of invasive mechanical ventilation (IMV) vs IMV outcomes for COVID-19 patients. Saliva was collected at the beginning of hospitalization from 118 patients with confirmed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) detection using RT-PCR at two Brazilian Hospitals. Patients were followed during all hospitalization period, and clinical information, laboratory test results and outcomes were collected. The classification of salivary infrared spectra by support vector machine (SVM) showed (i) a predictive value of severity – moderate vs severe [sensitivity of 68%, specificity of 52%, and accuracy of 60%], (ii) a predictive value of survival – discharge vs death [sensitivity of 60%, specificity of 64%, and accuracy of 64%], (iii) a predictive value of need of IMV (no IMV x IMV) [sensitivity of 46%, specificity of 72%, and accuracy of 66%]. This study reveals the potential of salivary ATR-FTIR spectroscopy as a prognostic tool to determine outcome in COVID-19 patients using an SVM algorithm.

Introduction

The COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, resulted in more than 636 million confirmed cases and 6.6

million deaths worldwide, overwhelmed health systems and caused an international health crisis with a drastic socio-economic repercussion^{1, 2}. Over the past 2 years of disease, there has been a significant advance in clinical research enabling a better understanding of SARS-CoV-2 and its management, however the increasing number of mutations limit the containment of the disease and result in even greater dissemination of the virus^{3, 4}.

COVID-19 manifests with a broad range of signs and symptoms ranging from fully asymptomatic to severe diseases requiring admission to the intensive care unit (ICU) and death. Prognostic models can predict expected outcome from each patient based on clinical data and dynamic biomarker profile⁵. In this context, the development of novel tools for predicting COVID-19 severity and prognosis could assist healthcare professionals in the decision-making process, and improve knowledge about the pathophysiology of the disease. Several prognostic model analysis for COVID-19 based on different parameters, such as plasmatic markers⁶, lung auscultation⁷, age, vital signs and laboratory results⁸ using intermediate and large cohorts showed distinct limitations.

The viral load of SARS-CoV-2 in saliva performed by Real-Time Polymerase Chain Reaction (RT-qPCR) was directly associated with inflammatory cytokines (IL-6, IL-10, IL-18 and CXCL10) with anti-spike (S) and anti-receptor binding domain (RBD) IgG levels, and it contributes to detection of prognostic factors for severity and mortality in COVID-19⁹. Furthermore, novel technological diagnostic platforms for COVID-19 using salivary biomarkers can offer non-invasive and a self-collected sample method¹⁰.

The attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectroscopy is a highly sensitive and reproducible analytical technique to detect

molecular composition of sample based on infrared (IR) absorption^{11, 12}. The applications of biofluids in ATR-FTIR platforms permits rapid, nondestructive, cost-effective, and reagent-free detection capable of discriminate normal and pathological conditions¹³. The clinical applicability of ATR-FTIR in COVID-19 was reported as a diagnostic tool using saliva samples^{11, 14-16}, nasopharyngeal swab^{12, 17}, plasma¹⁸ and RNA extracts¹⁹. The application of plasma samples on ATR-FTIR associated with clinical parameters was suggested as a severity triage tool at the point of admission²⁰.

COVID-19 is a complex multi-organ involvement disease, and so far, no clinical parameter or isolated biomarker has been able to predict disease outcome, and individual prognostic factors are unclear²¹. The discriminant analysis of the infrared spectra can be used by support vector machine (SVM) algorithms, which performs the group separation by hyperplanes (linear, quadratic, cubic)²². The proposal to use saliva in ATR-FTIR platform with a predictive and prognostic value in COVID-19 has not yet been described in the literature.

In the present study, we tested the hypothesis that unique spectral biomarkers can be identified in saliva of COVID-19 patients by ATR-FTIR, and applied to discriminate between those who progressed to severe and non-severe disease. Thus, the objective of this study was to determine the predictive and prognostic value of algorithms developed for COVID-19 disease stratification by (i) the disease progression - moderate or severe COVID-19, (ii) discharge or death, and (iii) the need of invasive mechanical ventilation, using saliva ATR-FTIR spectra and machine learning in early stages of hospitalization.

Methods

Ethical aspects and biosafety statement

In this study, all methods were carried out following relevant ethical guidelines and in accordance with the Declaration of Helsinki. All experimental procedures were approved by the Institutional Review Board of the UNA Faculty of Uberlandia (protocol number 4.602.081). Written informed consent was obtained from all the participants of the study. Samples were identified by numbers and time of collection avoiding any relation to the name of the patient or other information that could lead to personal identification. Clinical data were restricted to the smallest possible number of scientific team. All samples used for ATR-FTIR analyses were managed according to laboratory security protocols for COVID-19. All materials were cleaned before and after each experiment to prevent any source of contamination.

Study design, sample collection and data source

This observational prospective study was conducted in wards and intensive care units (ICU) for COVID-19 patients at the Clinical Hospital of Uberlandia (HC-UFU, Uberlandia, Minas Gerais, Brazil), a public hospital, and at the Uberlandia Medical Center (UMC, Uberlandia, Minas Gerais, Brazil), a private hospital. Recruitment followed a convenience sampling method between June 17, 2020 and March 17, 2021. Patients were prospectively enrolled and included a total of 118 adult inpatients, older than 18 years-old, with a positive SARS-CoV-2 RT-PCR test collected via nasopharyngeal swab. Exclusion criteria were orotracheal intubation prior or at the time of saliva collection, patients classified as mild COVID-19, and vaccinated patients.

Participants were followed during the hospitalization period, and timepoints were defined as admission, saliva collection and at the end of hospitalization. The disease severity classification was performed at the end of hospitalization, and the positive COVID-19 patients were categorized into: (i) moderate versus severe COVID-19; (ii) discharge versus death; and (iii) invasive mechanical ventilation (IMV) versus no need of IMV (Figure 1).

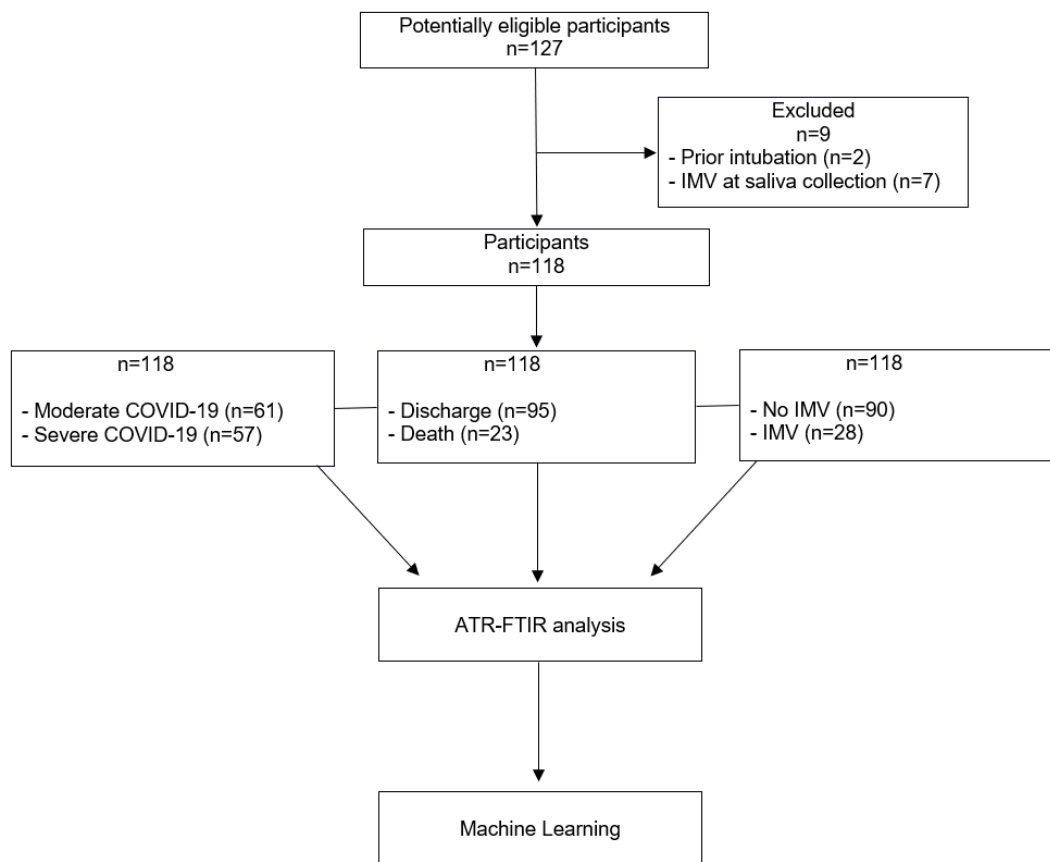


Figure 1. Study Flowchart. Note: IMV, invasive mechanical ventilation; ATR-FTIR, attenuated total reflection-Fourier transform infrared spectroscopy.

A saliva sample was obtained from each patient within a few days after hospital admission according to security protocols, which was performed by an unstimulated saliva collection followed by a spitting method²³. Participants were asked to abstain from eating, drinking or performing oral hygiene for at least 1 h before sample collection²⁴. During collection, participants were instructed to maintain

saliva in the mouth, and when appropriated, spit saliva into a 15 ml Falcon™ sterile tube until complete a minimum volume of 1 ml. The sample was immediately stored in an ice-chamber for around 1-h and after stored at -80C° until sample processing.

Routine laboratory measurements and blood counts were obtained as part of standard medical care by electronic database. Additional data as vital signs, O2 therapy, imaging reports for chest imaging, diagnostic outcomes, medications used, aggravating factors (need for tracheostomy, IMV, non-invasive ventilation (NIV), acute hemodialysis (HD), acute stroke) and outcomes were also recorded.

Variables

Our data included four different categories: 1) demographic data at admission (age, sex, smoking status, alcohol intake, COVID-19 symptoms on hospital admission, date of symptoms onset, data of SARS-CoV-2 PCR-positive result), 2) clinical data at the time of sample collection (heart rate, temperature, respiratory rate, O2 saturation, systolic blood pressure, diastolic blood pressure, disease status, imaging exams, blood count, clinical biochemistry profile), 3) general clinical information (body mass index (BMI) on admission, medications in use – prior and during hospitalization, O2 therapy start date, date of orotracheal intubation, outcome - discharge or death), 4) pre-existing medical conditions (comorbidities) (chronic kidney disease (CKD), dyslipidemia, hypothyroidism, pulmonary disease, hypertension, obesity, diabetes, cardiac disease, cancer). All categories were obtained from observational and prospective data collection and from the patient's electronic medical records.

COVID-19 disease severity classification

A severity scale for COVID-19 was based on a previous study²⁵ and modified as follows: 1) mild COVID-19, based on mild hypoxemia ($>94\%$) O_2 saturation, without need for O_2 therapy; 2) moderate COVID-19, based on hypoxemia ($\leq 94\%$) O_2 saturation, and need for O_2 therapy by nasal cannula (up to 5 l/min), or pneumonia on imaging; 3) severe COVID-19, based on the use of non-rebreather mask (NRB) (>6 l/min), or invasive mechanical ventilation (IMV). Intubation status and O_2 therapy modality were obtained by clinical notes and electronic database of each patient during hospitalization. SOFA score was calculated for all patients (<https://www.mdcalc.com/sequential-organ-failure-assessment-sofa-score>). Mild COVID-19 patients were not included in this study, as they did not require hospitalization.

ATR-FTIR spectroscopy

Infrared spectroscopy was used with the attenuated total reflection accessory Cary 630 (Agilent Technologies, United States) connected to a MicroLab (Agilent) software. This platform uses a micro component attenuated total reflectance (ATR) device assembled with a platinum diamond as an internal reflection element. Considering the low cost and a unique slight vibrational mode at 950 cm^{-1} in the biofingerprint infrared region²⁶, small aluminum pellets (Biogenetics, Minas Gerais, Brazil) were used to perform a large-scale system of analysis. In this system 10 μ l of saliva were inserted in a specific aluminum pellet and heated to 80°C in a dry bath for three minutes. The aluminum pellets on the side covered by a thin layer of dried salivary pellicle were analyzed in the wavenumber region from 4000 cm^{-1} to 650 cm^{-1} ,

32 scans were performed per analysis with a resolution of 4 cm^{-1} , with a room temperature around 22–23°C.

Spectral data pre-processing and algorithms development

The infrared spectral data analysis of dried salivary pellicle was divided into two steps: pre-processing and classification. Pre-processing is designed by aggregation, attribute selection, and data transformation. The aggregation gathered the spectral reading of each patient. The attribute selection involved the selection of spectrum regions that would be analyzed. The spectral data were truncated with the lipidic region (3050-2800 cm^{-1}) associated with the biofingerprint region (1800-900 cm^{-1}). Then, in data transformation, the Savitzky-Golay smoothing filter was applied to each spectrum followed by a first order derivative and pre-processed by vector normalization.

The classification was tested with Support Vector Machine (SVM) algorithm based on better results during model training. To analyze the predictive performance of the machine learning algorithms, ten times stratified cross-validation was used. All spectra were divided into ten subsets, with each interaction; nine of them were used to train the algorithm and one exclusively to test it, so that each subset was part of the test once²⁷.

In addition, the procedure was repeated ten times with changes in the samples configurations in these subsets to achieve a closer estimate of the real performance of the model. To measure the results obtained, three performance measures consolidated in the literature were used: sensitivity, specificity, and accuracy. Sensitivity is the proportion of positives (worst outcome) that were correctly classified, and the specificity is the proportion of negatives (better outcome) that were correctly

classified. The accuracy is defined as the total number of samples correctly classified considering true and false negatives^{28, 29}.

Statistical analysis

Patient characteristics were summarized using standard descriptive statistics: mean value/standard deviation for continuous variables and count/percent for categorical variables. Differences in proportions were assessed with Chi Square test. The normality of the data was assessed using Kolmogorov-Smirnov test and graphically through histograms. The non-normal distribution data were transformed with Z score. General Linear Model (GLZM) was used to compare clinical and metabolic profile.

Reported p-values were based on two-sided tests and compared at a significance level of 5%. Statistical Package for Social Sciences software (version 24.0; Chicago) was used for all statistical analyses.

Results

Clinical and demographic characteristics on admission

Laboratory, health and demographic data were obtained from 118 non-vaccinated patients with confirmed SARS-CoV-2 infection hospitalized at Clinical Hospital of Federal University of Uberlandia and at Uberlandia Medical Center between June 17, 2020 and March 17, 2021, under review board approval. Participants were followed during hospitalization and agreed to provide saliva samples used to assess its relation with severity and survival through the analysis of (i) moderate versus severe COVID-19; (ii) discharge versus death; and (iii) invasive

mechanical ventilation (IMV) versus no need of IMV. The COVID-19 disease severity classification was analyzed at the end of hospitalization.

Participant characteristics are listed in Table 1 with no difference between moderate and severe COVID-19. Regarding race/ethnicity, 78.8% were White, 19.5% Hispanic and 1.7% African American, with no differences between categories. The most prevalent comorbidities were hypertension (53%), followed by diabetes (27%), and obesity (23%), and among diabetic patients, 5.1% were insulin dependent on admission. 5.1% of chronic kidney disease (CKD) patients were undergoing hemodialysis prior to hospitalization. No differences were found on admission between categories, except for age (older patients had higher mortality), gender (73% of deaths were among males and 75% of patients who required IMV were men), also related to smoking history.

Table 1. Participant characteristics and personal history at hospital admission.

Timepoint - Admission	Total (n=118)	Moderate (n=61)	Severe (n=57)	p-value	Discharge (n=95)	Death (n=23)	p-value	No IMV (n=90)	IMV (n=28)	p-value
Age (year) Mean $\pm$ SD (Range)	58 $\pm$ 13 (26-93)	57 $\pm$ 15 (26-93)	59 $\pm$ 10 (39-87)	0.38	57 $\pm$ 13 (26-93)	64 $\pm$ 10 (39-87)	0.03*	57 $\pm$ 14 (39-87)	62 $\pm$ 10 (26-93)	0.09
Male – n. (%)	64 (54%)	28 (46%)	36 (63%)	0.06	47 (49%)	17 (74%)	0.03*	43 (48%)	21 (75%)	0.016*
Weight (kg) - Mean $\pm$ SD	83 $\pm$ 17	83 $\pm$ 18	81 $\pm$ 15	0.68	84 $\pm$ 18	76 $\pm$ 9	0.08	83 $\pm$ 17	81 $\pm$ 16	0.64
Comorbidities - n. (%)										
Hypertension	63 (53%)	36 (59%)	27 (47%)	0.26	52 (55%)	11 (48%)	0.64	49 (54%)	14 (50%)	0.82
Diabetes	32 (27%)	19 (31%)	13 (23%)	0.40	25 (26%)	7 (30%)	0.79	26 (29%)	6 (21%)	0.47
Obesity (BMI $\geq$ 30)	28 (24%)	13 (21%)	15 (26%)	0.66	24 (25%)	4 (17%)	0.58	21 (23%)	7 (25%)	1.0
Dyslipidemia	21 (18%)	11 (18%)	10 (17%)	1.0	19 (20%)	2 (9%)	0.24	17 (19%)	4 (14%)	0.77
CHF	17 (14%)	10 (16%)	7 (12%)	0.60	12 (13%)	5 (22%)	0.32	11 (12%)	6 (21%)	0.35
Pulmonary	14 (12%)	6 (10%)	8 (14%)	0.57	9 (9%)	5 (22%)	0.14	9 (10%)	5 (18%)	0.31
CKD	13 (11%)	10 (16%)	3 (5%)	0.07	10 (10%)	3 (13%)	1.0	11 (12%)	2 (7%)	0.52
Hypothyroidism	9 (8%)	5 (8%)	4 (7%)	1.0	7 (7%)	2 (8%)	1.0	8 (9%)	1 (4%)	0.45
Cancer (active)	6 (5%)	1 (2%)	5 (9%)	0.10	3 (3%)	3 (13%)	0.08	4 (4%)	2 (7%)	0.62
Smoking – current	1 (1%)	1 (2%)	0 (0%)	1.0	1 (1%)	0 (0%)	1.0	1 (1%)	0 (0%)	1.0
Smoking – history	20 (17%)	7 (11%)	13 (23%)	0.13	12 (13%)	8 (36%)	0.01*	11 (12%)	9 (33%)	0.018*
Alcohol intake	13 (11%)	7 (11%)	6 (11%)	1.0	11 (12%)	2 (9%)	1.0	9 (10%)	4 (15%)	0.73

Note: IMV, invasive mechanical ventilation; BMI, body mass index; CKD, chronic kidney disease; CHF, chronic heart failure; n., number; SD, standard deviation; *p<0.05.

The signals and symptoms of COVID-19 reported by the patient on admission were described as: desaturation (95.8%), dyspnea (71.2%), cough (59.3%), fever (50.8%), fatigue (30.5%), myalgia (27.1%), hyporexia (17.8%), headache (12.7%), rhinorrhea (12.7%), chest pain (11.9%), nausea (9.3%), odynophagia (8.5%), anosmia (6.8%), dysgeusia (6.8%), diarrhea (6.8%), vomiting (4.2%) and abdominal pain (1.7%). We found similar characteristics in the majority of the variables analyzed in all categories, except for moderate versus severe COVID-19: desaturation (moderate (91.8%), severe (100%), p 0.05), dyspnea (moderate (60.7%), severe (82.5%), p 0.01), fatigue (moderate (41%), severe COVID (19.3%), p 0.01) and chest pain (moderate (18%), severe COVID (5.3%), p 0.04); and IMV versus no need of IMV: fatigue (IMV (14.3%), no IMV (35.6%), p 0.03).

Clinical characteristics in saliva collection

In all of the 118 participants, saliva samples were collected once, within the first days of hospitalization (mean 3.0 ± 3.2 days). At the timepoint of saliva collection, we found a significant difference between all categories for SOFA score, amount of oxygen used and arterial oxygen pressure. Moderate and severe COVID-19, and discharge versus death also presented changes in liver function (total and direct bilirubin), presented in Table 2. Indirect bilirubin and renal function (urea and creatinine) and days of symptoms also showed a significant difference between the discharge and death category, and the IMV versus no IMV presented significant difference in the days of symptoms.

At the time of saliva collection, 55.9% of patients were classified as having a SOFA score equal to 2, the minimum SOFA score was 0 and the maximum was 7, and the tool was able to differentiate between moderate and severe COVID patients only for those patients with a SOFA score ≤ 1 (moderate COVID (32.8%), severe COVID (3.5%), $p < 0.001$). In the discharge versus death analysis, the SOFA score was efficient in discriminating the categories (SOFA ≤ 1 , discharge (22.1%), death (4.3%), $p < 0.001$; SOFA = 2, discharge (61.1%), death (34.8%), $p < 0.001$; SOFA ≥ 3 , discharge (16.8%), death (60.9%), $p < 0.001$). For IMV and no IMV group, SOFA score also showed significant difference (SOFA ≤ 1 , no IMV (23.3%), IMV (3.6%), $p < 0.004$; SOFA = 2, no IMV (57.8%), IMV (50.0%), $p < 0.004$; SOFA ≥ 3 , no IMV (18.9%), IMV (46.4%), $p < 0.004$).

Table 2. Clinical Data in Saliva Collection.

Timepoint - Saliva collection - Mean $\pm$ SD	Moderate (n=61)	Severe (n=57)	p-value	Discharge (n=95)	Death (n=23)	p-value	No IMV (n=90)	IMV (n=28)	p-value
Days of symptoms	11.8 $\pm$ 4.1	12.6 $\pm$ 4.4	0.35	12.6 $\pm$ 4.1	10.6 $\pm$ 4.5	0.04*	12.7 $\pm$ 4.2	10.8 $\pm$ 4.1	0.04*
SOFA score	1.9 $\pm$ 1.3	2.6 $\pm$ 1.3	0.013*	2.0 $\pm$ 1.0	3.4 $\pm$ 1.7	<0.001*	2.1 $\pm$ 1.2	2.9 $\pm$ 1.4	0.003*
FiO ₂	28.4 $\pm$ 6.2	43.0 $\pm$ 16.2	<0.001*	34.0 $\pm$ 11.9	41.3 $\pm$ 20.1	0.01*	33.4 $\pm$ 11.4	42.2 $\pm$ 19.3	0.004*
PaO ₂	74.9 $\pm$ 24.7	65.4 $\pm$ 13.2	0.01*	72.4 $\pm$ 20.7	62.7 $\pm$ 17.1	0.04*	73.1 $\pm$ 21.9	62.5 $\pm$ 12.1	0.01*
Total bilirubin	0.4 $\pm$ 0.3	0.7 $\pm$ 0.7	0.02*	0.4 $\pm$ 0.3	1.1 $\pm$ 0.9	<0.001*	0.5 $\pm$ 0.5	0.8 $\pm$ 0.6	0.12
Direct bilirubin	0.3 $\pm$ 0.2	0.6 $\pm$ 0.6	0.05	0.3 $\pm$ 0.2	0.8 $\pm$ 0.9	0.001*	0.3 $\pm$ 0.5	0.6 $\pm$ 0.5	0.14
Indirect bilirubin	0.2 $\pm$ 0.2	0.2 $\pm$ 0.1	0.78	0.1 $\pm$ 0.1	0.3 $\pm$ 0.2	0.02*	0.1 $\pm$ 0.1	0.2 $\pm$ 0.2	0.12
Urea	43.3 $\pm$ 28.9	49.1 $\pm$ 44.8	0.40	41.3 $\pm$ 29.2	65.9 $\pm$ 57.4	0.005*	42.5 $\pm$ 31.4	57.6 $\pm$ 51.4	0.069
Creatinine	1.3 $\pm$ 1.4	1.4 $\pm$ 1.9	0.76	1.1 $\pm$ 1.2	2.3 $\pm$ 2.9	<0.01*	1.2 $\pm$ 1.3	1.8 $\pm$ 2.6	0.10

Note: O₂, oxygen; l/min, liter per minute; SOFA, sepsis-related organ failure assessment; PaO₂, arterial oxygen tension; FiO₂, fraction of inspired oxygen; SD, standard deviation; * $p < 0.05$.

Clinical characteristics at the endpoint of hospitalization

At the end of hospitalization, all outcomes were evaluated, and we found statistical differences between categories in hospital stay, intensive care stay, acute hemodialysis, invasive and non-invasive ventilation and tracheostomy (Table 3). Among patients who were discharged from hospital, 96.7% were classified as moderate COVID and 61.4% were severe COVID on hospitalization, $p < 0.001$, with 1.6% death of the moderate COVID category and 38.6% deaths in the severe COVID, $p < 0.001$. The majority of patients submitted to non-invasive ventilation 70.5% were discharged from hospital. It was observed that of 118 patients studied, 23 (19.5%) died, and among the 28 who required IMV, 75.0% died, $p < 0.001$.

Among medications used in the treatment of COVID-19, the following stand out: corticosteroids, oseltamivir, nitazoxanide, ivermectin, hydroxychloroquine, zinc and tocilizumab. Corticosteroids were the most used medication in our study (moderate COVID (95.1%), severe COVID (91.2%), $p = 0.48$) with no difference between categories and a statistical difference was observed for tocilizumab, (moderate COVID (1.6%), severe COVID (10.5%), $p = 0.05$), and for oseltamivir (no IMV (6.7%), IMV (21.4%), $p = 0.03$).

Table 3. Outcome of the clinical characteristics of patients with COVID-19 analyzed at the end of hospitalization.

Timepoint - End of hospitalization	Total (n=118)	Moderate (n=61)	Severe (n=57)	P- value	Discharge (n=95)	Death (n=23)	P- value	No IMV (n=90)	IMV (n=28)	P- value
Hospital stay (days) - Mean $\pm$ SD	14.1 $\pm$ 13.9	9.4 $\pm$ 8.8	19.1 $\pm$ 16.6	<0.001*	11.8 $\pm$ 11.6	23.3 $\pm$ 18.7	<0.001*	10.3 $\pm$ 8.0	26.3 $\pm$ 20.7	<0.001*
Intensive care stay (days) - Mean $\pm$ SD	5.9 $\pm$ 12.4	0.7 $\pm$ 1.7	11.5 $\pm$ 16.1	<0.001*	2.8 $\pm$ 6.8	18.8 $\pm$ 20.2	<0.001*	1.4 $\pm$ 2.9	20.5 $\pm$ 18.8	<0.001*
IMV - n. (%)	28 (24%)	0 (0%)	28 (49%)	<0.001*	7 (7%)	21 (91%)	<0.001*			
NIV - n. (%)	61 (52%)	16 (26%)	45 (79%)	<0.001*	43 (45%)	18 (78%)	0.005*	39 (43%)	22 (78%)	0.001*
Tracheostomy- n. (%)	8 (7%)	0 (0%)	8 (14%)	0.002*	1 (1%)	7 (30%)	<0.001*	0 (0%)	8 (29%)	<0.001*
Chest CT (% of involvement) n. (%)										
0-24%	5 (5%)	3 (7%)	2 (4%)	0.73	4 (5%)	1 (7%)	0.81	3 (4%)	2 (9%)	0.60
25-50%	32 (35%)	17 (38%)	15 (33%)		28 (37%)	4 (27%)		26 (37%)	6 (29%)	
51-100%	54 (59%)	25 (56%)	29 (63%)		44 (58%)	10 (67%)		41 (59%)	13 (62%)	

Note: n., number; SD, standard deviation; IMV, invasive mechanical ventilation; NIV, non-invasive ventilation; CT, computer tomography; *p<0.05.

Support vector machine algorithms

The mean infrared raw spectra of saliva between 1800-800 cm^{-1} of COVID-19 patients with different outcomes was depicted indicating salivary components as proteins, lipids, DNA/RNA and carbohydrates (Figure 2).

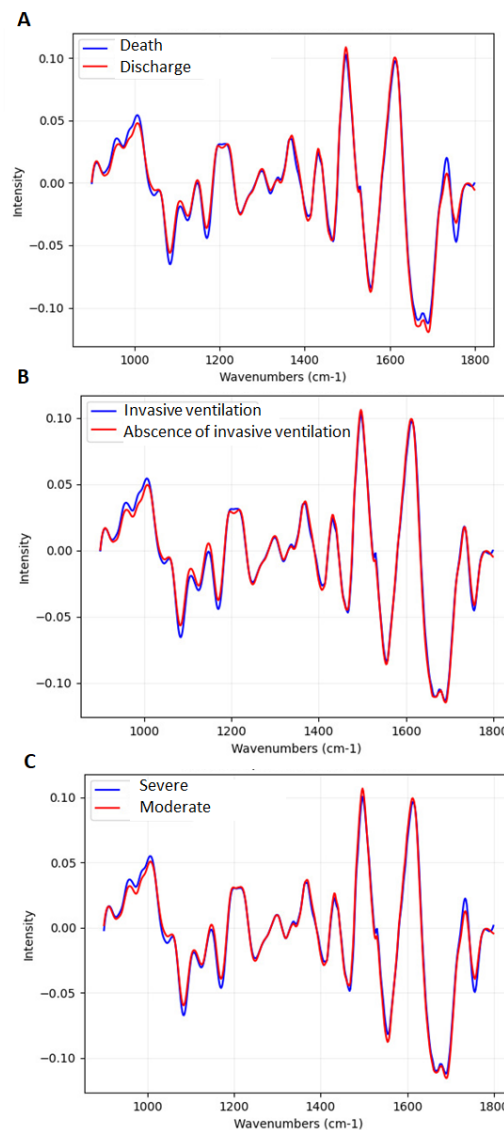


Figure 2. Representative average of ATR-FTIR spectra (800–1800 cm^{-1}) in saliva of COVID-19 patients with different outcomes.

Table 4 indicates the discrimination performance of salivary infrared spectra of COVID-19 patients in accordance with severity (moderate or severe COVID-19), discharge or death and presence or absence of invasive mechanical ventilation. The pre-processing of algorithms followed Rubberband plus Amide I normalization, and analyzed spectrum regions were 3050-2800 and 1800-900 cm⁻¹.

Table 4. Discrimination of SVM Algorithms Applied to Classify Outcomes in COVID-19 Inpatients.

Algorithm	Predictive Value	Sensitivity (Worst outcome)	Specificity (Better outcome)	Accuracy
SVM	Severity	0,68	0,52	0,60
	Death	0,60	0,64	0,64
	Invasive mechanical ventilation	0,46	0,72	0,66

Discussion

This study aimed to develop predictive algorithms for COVID-19 support need and hospital stratification by disease course and outcomes using saliva ATR-FTIR and machine learning. Key outcome markers that can determine COVID-19 disease severity were established²⁵, being considered relevant outcomes (i) the progression of the disease to moderate or severe COVID-19²¹, (ii) the possibility of death³⁰ and (iii) the need for invasive mechanical ventilation as severity predictors³¹.

The development of a simple, sensitive, reagent-free method that uses only a small volume of self-collected biofluid such as saliva can be valuable, especially in places with limited resources^{32, 33}. The ATR-FTIR technique can be considered a good investment considering a large-scale use, and a cost-effective technique,

without the need for reagents, neither the purchase of kits to maintain the analyzes in this device. Some models are portable, easy to use and with quick analysis, allowing travel for the location of interest^{28, 34}. Furthermore, another factor that makes saliva the ideal fluid for viral detection is the presence of angiotensin converting enzyme II (ACE 2) receptors in the salivary glands, which are fundamental in the process of viral coupling and replication in host cells, making these glands the perfect target for SARS-CoV-2³⁵.

ATR-FTIR saliva spectra of COVID-19 patients and healthy controls have been used for predictive algorithms development and diagnosis with high specificity and sensitivity through different statistical models^{12, 14-17}. However, there is a demand in the literature to fill gaps related to the use of ATR-FTIR for prognostic purposes¹⁶. This study provides a follow-up outcome data, being pioneer in the use of saliva for prognostic purposes on COVID-19 and SVM.

SVM model has been shown to be superior to other regression techniques as it finds nonlinear global solutions and high dimensional input support vectors, which are representative of the problem, in addition to high accuracy prediction and less spectral noise³⁶. The application of plasma samples on ATR-FTIR associated with clinical parameters (age, sex, diabetes mellitus and hypertension) was suggested as a severity triage tool for COVID-19 at the point of admission by Banerjee et al. In addition, the plasma independent test set achieved discrepant values compared to our study with 94.1% sensibility and 69.2% specificity²⁰.

Worse predictive values in the algorithm using saliva can be justified by the various factors that influence saliva amount and composition, as it does not always reflect the concentrations of molecules available on plasma. Salivary composition may be influenced by circadian rhythm, method of collection, flow rate, type and size

of salivary gland, diet, drugs, age, sex, blood type, and physiological status³⁷. However, so far, the use of prognostic salivary algorithms has not yet been described in the literature to allow comparison with our model.

Several clinical and immunological parameters are used to predict severity in COVID-19, among them, the increase in circulating inflammatory cytokines, depletion of platelets, lymphocytes and T cells, and even the viral load in saliva⁹. Associated with clinical parameters, our findings showed an older age and higher mortality among men, also related to smoking history, and worst hepatic and renal impairment which was related to worst prognosis. A meta-analysis corroborates our findings, as age is presented as a prognostic factor related to hospitalization, with a linear dose-response association with mortality. Sex was a prognostic factor for IMV and smoking was presented as a prognostic factor of clinical deterioration with a consistent association risk of mortality²¹. Another fact that may explain this outcome is the longer duration of the virus in the respiratory samples of patients with severe disease (21 days), compared to 14 days in mild disease, especially in patients older than 60 years and in male patients³⁸. Older patients (aged >60 years) also presented with more systemic symptoms, extensive radiological ground-glass lung changes with more consolidated lung lesions, lymphopenia, thrombocytopenia, and increased C-reactive protein and lactate dehydrogenase levels^{39, 40}.

The SOFA score is another tool that has been used in an attempt to predict mortality in COVID-19. Rapid change in SOFA score ≥ 2 points after infection is considered a clinical criterion for sepsis-associated organ dysfunction⁴¹. A study with patients admitted to an intensive care unit showed that only the maximum SOFA score would be a good predictor of mortality and that the increase in the SOFA score would be smaller in surviving patients⁴².

On the other hand, authors argue that severe COVID-19 is a not limited lung injury but a multi-organ involvement disease, and a multi-organ failure assessment tool as SOFA score showed a significant association ($p < 0.001$) with in-hospital mortality^{41, 43}. Studies indicate that the optimal point for discrimination between mortality and survival is a SOFA of 5^{41, 44}. At the time of saliva collection, 55.9% of our patients were classified as having a SOFA score equal to 2, similar to another study⁴¹, and 95.7% with $\text{SOFA} \geq 2$ died, $p < 0.001$. In our study, the SOFA score proved to be a good predictor of mortality in all categories evaluated, with better results for discharge and death analysis (discharge - mean 2.0 ± 1.0 versus death - mean 3.4 ± 1.7 , $p < 0.001$). There is currently no consensus in the literature about the initial SOFA score being a predictor of mortality in patients with COVID-19. Work remains in identifying a gold standard prognostic biomarker that can be used to guide clinical practice in COVID-19.

In summary, we found that the characteristics of the participants at the time of admission, personal history, signals and symptoms of COVID-19, and computed tomography data were not able to predict disease severity, hospital outcome or progression to IMV. It was possible to establish the profile of patients and verify the influence of male gender, older age, smoking history, liver and kidney dysfunction, greater need for oxygen supplementation in patients with more severity and worse outcome. Salivary algorithms based on the ATR-FTIR spectrum, proved to be possible alternatives of prognostic value in COVID-19. In a context of limited resources, knowing that a patient is at risk of becoming severely ill in 68% would already be a reason to refer him to a reference center where he could have adequate follow-up. In this context, the salivary ATR-FTIR spectroscopy coupled with SVM classification could provide a novel alternative for biomedical decision or monitoring.

In a perspective of improving the predictive value of algorithms, it is suggested to test associations with clinically relevant variables verified in this study, such as the SOFA score, age, gender and amount of O₂ used.

Some limitations must be considered when interpreting the results of this study. In order to avoid performing unnecessary exams, the results of routine exams of patients were used, which were requested according to the clinical need. Furthermore, control group analyzes were not performed, and could be useful to verify the algorithm predictive and prognostic value in differentiating outcomes in COVID-19.

Conclusion

This study has shown that salivary algorithms based on the ATR-FTIR spectrum coupled with SVM classification, proved to be possible alternative of prognostic value in COVID-19. Other studies are needed in an attempt to create associations of other independent variables with saliva to improve the predictive power of the algorithms.

Data availability statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. WHO. WHO Coronavirus (COVID-19) Dashboard; 2022. Available from: <https://covid19.who.int/>. Accessed 2022 Nov, 25.

2. Yesudhas, D., Srivastava, A., Gromiha, M.M. COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics. *Infection*. 49(2), 199-213 (2021).
3. Hodgson, S.H. et al. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. *Lancet Infect Dis*. 21(2), e26-e35 (2021).
4. Kantarcioglu, B. et al. An Update on the Pathogenesis of COVID-19 and the Reportedly Rare Thrombotic Events Following Vaccination. *Clin Appl Thromb Hemost*. 27, 10760296211021498 (2021).
5. de Bruin, S. et al. Clinical features and prognostic factors in Covid-19: A prospective cohort study. *EBioMedicine*. 67, 103378 (2021).
6. Wu, S. et al. Identification and Validation of a Novel Clinical Signature to Predict the Prognosis in Confirmed Coronavirus Disease 2019 Patients. *Clin Infect Dis*. 71(12), 3154-3162 (2020).
7. Glangetas, A. et al. Deep learning diagnostic and risk-stratification pattern detection for COVID-19 in digital lung auscultations: clinical protocol for a case-control and prospective cohort study. *BMC Pulm Med*. 21(1), 103 (2021).
8. Incerti, D. et al. Prognostic model to identify and quantify risk factors for mortality among hospitalised patients with COVID-19 in the USA. *BMJ Open*. 11(4), e047121 (2021).
9. Silva, J. et al. Saliva viral load is a dynamic unifying correlate of COVID-19 severity and mortality. *medRxiv [Preprint]*. 10, 2021.01.04.21249236 (2021).
10. Caixeta, D.C. et al. One-Year Update on Salivary Diagnostic of COVID-19. *Front Public Health*. 9, 589564 (2021).
11. Wood, B.R. et al. Infrared Based Saliva Screening Test for COVID-19. *Angew Chem Int Ed Engl*. 60(31), 17102-17107 (2021).
12. Barauna, V.G. et al. Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity. *Anal Chem*. 93(5), 2950-2958 (2021).
13. Butler, H.J. et al. Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer. *Nat Commun*. 10(1), 4501 (2019).
14. Martinez-Cuazitl, A. et al. ATR-FTIR spectrum analysis of saliva samples from COVID-19 positive patients. *Sci Rep*. 11(1), 19980 (2021).
15. Nascimento, M.H.C. et al. Noninvasive Diagnostic for COVID-19 from Saliva Biofluid via FTIR Spectroscopy and Multivariate Analysis. *Anal Chem*. 94(5), 2425-2433 (2022).
16. Kazmer, S.T. et al. Pathophysiological Response to SARS-CoV-2 Infection Detected by Infrared Spectroscopy Enables Rapid and Robust Saliva Screening for COVID-19. *Biomedicine*. 10(2), 351 (2022).
17. Nogueira, M.S. et al. Rapid diagnosis of COVID-19 using FT-IR ATR spectroscopy and machine learning. *Sci Rep*. 11(1), 15409 (2021).
18. Zhang, L. et al. Fast Screening and Primary Diagnosis of COVID-19 by ATR-FT-IR. *Anal Chem*. 93(4), 2191-2199 (2021).
19. Kitane, D.L. et al. A simple and fast spectroscopy-based technique for Covid-19 diagnosis. *Sci Rep*. 11(1), 16740 (2021).
20. Banerjee, A. et al. Rapid Classification of COVID-19 Severity by ATR-FTIR Spectroscopy of Plasma Samples. *Anal Chem*. 93(30), 10391-10396 (2021).

21. Bellou, V. et al. Prognostic factors for adverse outcomes in patients with COVID-19: a field-wide systematic review and meta-analysis. *Eur Respir J.* 59(2), 2002964 (2022).
22. Yonar, D. et al. Rapid diagnosis of malignant pleural mesothelioma and its discrimination from lung cancer and benign exudative effusions using blood serum. *Biochim Biophys Acta Mol Basis Dis.* 1868(10),166473 (2022).
23. Paschotto, D.R., Pupin, B., Bhattacharjee, T.T., Soares, L.E.S. Saliva Preparation Method Exploration for ATR-FTIR Spectroscopy: Towards Bio-fluid Based Disease Diagnosis. *Anal Sci.* 36(9), 1059-1064 (2020).
24. Mikkonen, J.J. et al. Fourier Transform Infrared Spectroscopy and Photoacoustic Spectroscopy for Saliva Analysis. *Appl Spectrosc.* 70(9), 1502-10 (2016).
25. Del Valle, D.M. et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat Med.* 26(10), 1636-1643 (2020).
26. Maitra, I. et al. Establishing spectrochemical changes in the natural history of oesophageal adenocarcinoma from tissue Raman mapping analysis. *Anal Bioanal Chem.* 412(17), 4077-4087 (2020).
27. Franca, T., Goncalves, D., Cena, C. ATR-FTIR spectroscopy combined with machine learning for classification of PVA/PVP blends in low concentration. *Vibrational Spectroscopy.* 120, 103378 (2022).
28. Caixeta, D.C. et al. Salivary molecular spectroscopy: A sustainable, rapid and non-invasive monitoring tool for diabetes mellitus during insulin treatment. *PLoS One.* 15(3), e0223461 (2020).
29. Passos, J.O.S. et al. Spectrochemical analysis in blood plasma combined with subsequent chemometrics for fibromyalgia detection. *Sci Rep.* 10(1), 11769 (2020).
30. Adderley, N.J. et al. Development and external validation of prognostic models for COVID-19 to support risk stratification in secondary care. *BMJ Open.* 12(1), e049506 (2022).
31. Melamed, R. et al. Prognostic factors and outcomes in COVID-19 patients requiring prolonged mechanical ventilation: a retrospective cohort study. *Ther Adv Respir Dis.* 16, 17534666221086415 (2022).
32. Chandna, A. et al. Facilitating Safe Discharge Through Predicting Disease Progression in Moderate Coronavirus Disease 2019 (COVID-19): A Prospective Cohort Study to Develop and Validate a Clinical Prediction Model in Resource-Limited Settings. *Clin Infect Dis.* 75(1), e368-e379 (2022)
33. Aljohani, F.D. et al. Prognostic factors for predicting severity and mortality in hospitalized COVID-19 patients. *J Clin Lab Anal.* 36(3), e24216 (2022).
34. Elkins, K.M. Rapid presumptive "fingerprinting" of body fluids and materials by ATR FT-IR spectroscopy. *J Forensic Sci.* 56(6),1580-7 (2011).
35. Fernandes, L.L., et al. Saliva in the Diagnosis of COVID-19: A Review and New Research Directions. *J Dent Res.* 99(13),1435-1443 (2020).
36. Thissen, U., Pepers, M., Üstün, B., Melssen, W.J., Buydens, L.M.C. Comparing support vector machines to PLS for spectral regression applications. *Chemometrics and Intelligent Laboratory Systems.* 73(2), 169–179 (2004).
37. Pfafe, T., Cooper-White, J., Beyerlein, P., Kostner, K., Punyadeera, C. Diagnostic potential of saliva: current state and future applications. *Clin Chem.* 57(5), 675-87 (2011).
38. Zheng, S. et al. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. *BMJ.* 369, m1443 (2020).

39. Chan, J.F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. 395(10223), 514-523 (2020).
40. Song, F. et al. Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia. *Radiology*. 295(1), 210-217 (2020).
41. Yang, Z., Hu, Q., Huang, F., Xiong, S., Sun, Y. The prognostic value of the SOFA score in patients with COVID-19: A retrospective, observational study. *Medicine (Baltimore)*. 100(32), e26900 (2021).
42. Gruyters, I. et al. Predictive value of serial evaluation of the Sequential Organ Failure Assessment (SOFA) score for intensive care unit mortality in critically ill patients with COVID-19: a retrospective cohort study. *Anaesthesiol Intensive Ther*. 54(1), 3-11 (2022).
43. Puhr-Westerheide, D. et al. Sequential Organ Failure Assessment Outperforms Quantitative Chest CT Imaging Parameters for Mortality Prediction in COVID-19 ARDS. *Diagnostics (Basel)*. 12(1), 10 (2021).
44. Fayed, M. et al. Sequential Organ Failure Assessment (SOFA) Score and Mortality Prediction in Patients With Severe Respiratory Distress Secondary to COVID-19. *Cureus*. 14(7), e26911 (2022).

Acknowledgements

To all colleagues and patients who contributed directly or indirectly to the completion of this work. To Health Sciences Program, Faculty of Medicine, Federal University of Uberlandia, Minas Gerais, Brazil. The authors also express their gratitude for research funding to the National Institute of Science and Technology in Theranostics and Nanobiotechnology (CNPq Process N.: 465669/2014-0) and FAPEMIG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

D.L.Z.S., Y.C.P.M., and T.M.C. conceived the study idea, and with L.O.S. and R.S.S. performed interpretation, manuscript writing, and final approval. D.L.Z.S and N.C.M.N.M. performed collection, and L.O.S., A.T.F.S., M.G.C. and M.M.M.

performed data analysis. All authors reviewed and approved the final version of the manuscript.

Conceptualization: D.L.Z.S, Y.C.P.M., and T.M.C.

Data curation: D.L.Z.S, L.O.S., N.C.M.N.M.

Formal analysis: D.L.Z.S, L.O.S., M.G.C., A.T.F.S., and M.M.M.

Funding acquisition: T.M.C.

Investigation: D.L.Z.S., Y.C.P.M., T.M.C., and R.S.S.

Methodology: D.L.Z.S., Y.C.P.M., T.M.C.

Writing – original draft: D.L.Z.S., L.O.S., Y.C.P.M., T.M.C., and R.S.S.

Writing – review e editing: D.L.Z.S., L.O.S., T.M.C., R.S.S., and Y.C.P.M.

Competing interests

The authors declare no competing interests.

Saliva Metabolomics for COVID-19 Prognosis

Short title: Saliva Metabolomics for COVID-19 Prognosis

Debora Lara Zuza Scheucher¹, Luciana Oliveira e Silva¹, Natalia Cristina Martins Nunes Mendes¹, Hebreia Oliveira Almeida Souza², Mario Machado Martins³, Robinson Sabino-Silva⁴, Yara Cristina de Paiva Maia⁵, Thulio Marquez Cunha^{5*}.

¹ Health Sciences Program, Faculty of Medicine, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil.

² Postgraduate Program in Genetics and Biochemistry, Institute of Genetics and Biochemistry, Federal University of Uberlandia (UFU), Uberlandia, Minas Gerais, Brazil.

³ Laboratory of Nanobiotechnology, Institute of Genetics and Biochemistry, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil.

⁴ Innovation Center in Salivary Diagnostic and Nanobiotechnology, Department of Physiology, Institute of Biomedical Sciences, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil.

⁵ Faculty of Medicine, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil.

* Corresponding author

E-mail: thcunha@yahoo.com.br (TMC)

Abstract

Early identification of patients who will progress to the severe form of COVID-19 is still a challenge. Here, we explored the saliva metabolome of patients infected with SARS-CoV-2 to search for prognostic biomarkers and to improve the knowledge of metabolomic disturbance in this infection. We analyzed the saliva metabolome of 107 patients infected with SARS-CoV-2 by HPLC-MS early in hospitalization (mean 3.0 ± 3.2 days). We first classified patients into categories (i) moderate vs severe COVID-19, (ii) discharge vs death, and (iii) no invasive mechanical ventilation (IMV) vs IMV outcomes observed at the end of hospitalization. Metabolomic profile of COVID-19 patients showed an involvement of purines, amide, lipid, sulfoxide and ketones. Saliva metabolome allowed us to identify metabolites related to greater disease severity in the evaluated outcomes which might be used in selection of potential saliva biomarker for severity evaluation early in hospitalization and open new therapeutic perspectives against COVID-19.

Introduction

In December 2019, with the emergence of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), researchers and healthcare teams around the world began a race against time to unravel the structure, mechanism and the pathophysiological effects of the virus in the human body. Despite causing mild symptoms in 80% of infected patients, the coronavirus disease (COVID-19) resulted in more than 636 million confirmed cases and 6.6 million deaths globally as of November 28, 2022 [1-2]. The ability to early identify infected patients who will progress to a severely ill disease, need for invasive mechanical ventilation or death

could assist healthcare professionals in the decision-making process, and also increase knowledge of the metabolites involved in the pathogenesis of COVID-19 and its management [3,4].

Studies involving multiomic approaches search for diagnostic and prognostic biomarkers for COVID-19, and try to fill in the gaps of the metabolic disturbances caused by the viral infection [5-7]. In this context, saliva is a biological fluid that reflects the physiological state of the body, including endocrine, nutritional, metabolic and emotional conditions. Circulating biomolecules that originate from a local or even systemic disease can be transported to the salivary glands, which consequently will modify the composition of saliva that can be used for early diagnosis, prognosis and severity stratification of the disease [8-10].

The clinical applicability of metabolomic in COVID-19 was reported in several studies with diagnostic purposes, disease evolution, need of hospitalization and comparison between groups with mild, moderate, or severe disease via plasma, urine, serum, saliva and sebum [4, 7, 11-14]. Plasma metabolomic analysis showed dysregulation in biofluids after SARS-CoV-2 infection, with a disproportionate reduction in nutrient circulation [11]. Abnormally high levels of ketone bodies in serum analysis, was related to liver injury and dyslipidemia, with impairment in hepatic glutathione synthesis and oxidative stress [15]. Furthermore, groups of amino acids, carbohydrates, fatty acids, and glycerophospholipids showed differences according to the severity of disease, with impairment in the tricarboxylic acid (TCA) cycle, macrophage dysregulation, tryptophan-kynurenine metabolism activation, platelet degranulation, massive metabolic suppression and complement system pathways disturbances in COVID-19 patient sera and plasma samples [6, 11, 13, 16-18].

Saliva metabolomics are still under explored in the study of viral infections [10, 19]. COVID-19 studies in this field showed differences in oral microbiota after SARS-CoV-2 infection [7, 20-21], and Pozzi *et al.*, identified salivary metabolites that allowed assessing the need of hospitalization [7]. However, among the few studies found with saliva metabolomics, none investigated which metabolites are likely presented in saliva of patients who progressed to death or need of invasive mechanical ventilation (IMV).

In the present study, we tested the hypothesis that SARS-CoV-2 induces characteristic molecular changes that can be detected in saliva of severe COVID-19 patients, and may be used in selection of potential saliva biomarker for severity evaluation early in hospitalization and open new therapeutic perspectives against COVID-19. Thus, the objective of this study was to analyze the saliva metabolites from COVID-19 patients for disease stratification by (i) the disease progression - moderate or severe COVID, (ii) the risk of death, and (iii) the need of IMV, performing the metabolomics of saliva collected in early stages of hospitalization.

Materials and methods

Ethical aspects and biosafety statement

In this study, all methods were carried out following relevant ethical guidelines and in accordance with the Declaration of Helsinki. All experimental procedures were approved by the Institutional Review Board of the UNA Faculty of Uberlandia (protocol number 4.602.081). Written informed consent was obtained from all the participants of the study. Samples were identified by numbers and time of collection avoiding any relation to the name of the patient or other information that could lead to

personal identification. Clinical data were restricted to the smallest possible number of scientific team. All samples used for metabolomic analyses were managed according to laboratory security protocols for COVID-19. All materials were cleaned before and after each experiment to prevent any source of contamination.

Study design, sample collection and data source

This observational prospective study was conducted in wards and intensive care units (ICU) for COVID-19 patients at the Clinical Hospital of Uberlandia (HC-UFU, Uberlandia, Minas Gerais, Brazil), a public hospital, and at the Uberlandia Medical Center (UMC, Uberlandia, Minas Gerais, Brazil), a private hospital. Recruitment followed a convenience sampling method between June 17, 2020 and March 17, 2021. Patients were prospectively enrolled and included a total of 107 adult inpatients, older than 18 years-old, with a positive SARS-CoV-2 RT-PCR (Real-Time Polymerase Chain Reaction) test collected via nasopharyngeal swab. Exclusion criteria were orotracheal intubation prior or at the time of saliva collection, patients classified as mild COVID-19, and vaccinated patients.

Participants were followed during the hospitalization period, and timepoints were defined as admission, saliva collection and at the end of hospitalization. The disease severity classification was performed at the end of hospitalization, and the positive COVID-19 patients were categorized into: (i) moderate versus severe COVID-19; (ii) discharge versus death; and (iii) invasive mechanical ventilation (IMV) versus no need of IMV (Figure 1).

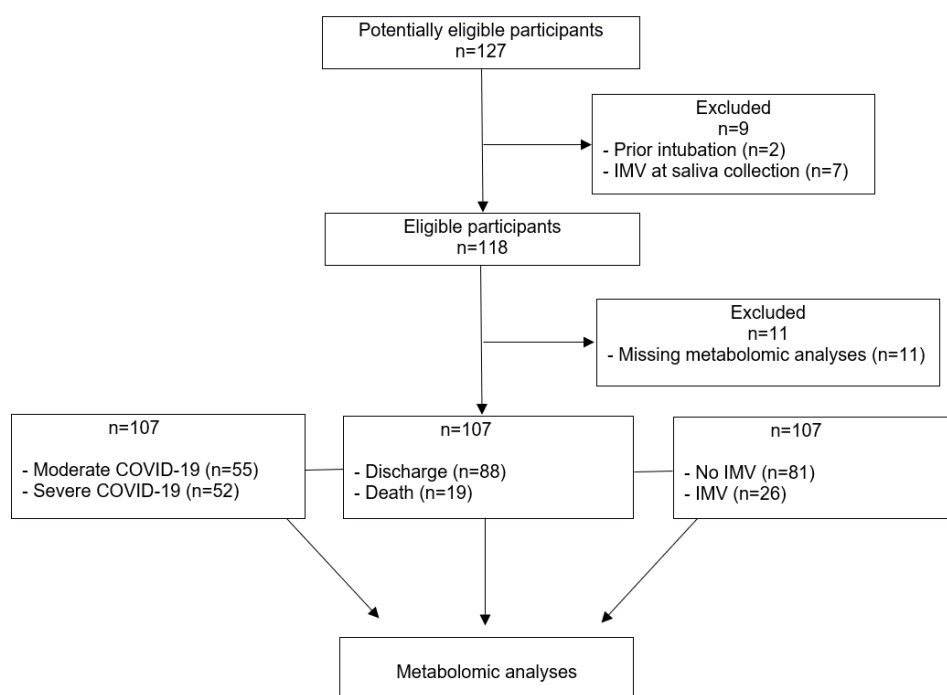


Fig 1. Study flowchart. IMV, invasive mechanical ventilation.

A saliva sample was obtained from each patient within a few days after hospital admission (mean 3.0 ± 3.2 days) according to security protocols, which was performed by an unstimulated saliva collection followed by a spitting method [22]. Participants were asked to abstain from eating, drinking or performing oral hygiene for at least 1 h before sample collection [23]. During collection, participants were instructed to maintain saliva in the mouth, and when appropriated, spit saliva into a 15 ml Falcon™ sterile tube until complete a minimum volume of 1 ml. The sample was immediately stored in an ice-chamber for around 1-h and after stored at -80°C until sample processing.

Routine laboratory measurements and blood counts were obtained as part of standard medical care by electronic database. Additional data as vital signs, O_2 therapy, imaging reports for chest imaging, diagnostic outcomes, medications used,

aggravating factors (need for tracheostomy, IMV, non-invasive ventilation (NIV), acute hemodialysis (HD), acute stroke) and outcomes were also recorded.

Variables

Our data included four different categories: 1) demographic data at admission (age, sex, smoking status, alcohol intake, COVID-19 symptoms on hospital admission, date of symptoms onset, date of SARS-CoV-2 PCR-positive result), 2) clinical data at the time of sample collection (heart rate, temperature, respiratory rate, O₂ saturation, systolic blood pressure, diastolic blood pressure, disease status, imaging exams, blood count, clinical biochemistry profile), 3) general clinical information (body mass index (BMI) on admission, medications in use – prior and during hospitalization, O₂ therapy start date, date of orotracheal intubation, outcome - discharge or death), 4) pre-existing medical conditions (comorbidities) (chronic kidney disease (CKD), dyslipidemia, hypothyroidism, pulmonary disease, hypertension, obesity, diabetes, cardiac disease, cancer). All categories were obtained from observational and prospective data collection and from the patient's electronic medical records.

COVID-19 disease severity classification

A severity scale for COVID-19 was based on a previous study [24] and modified as follows: 1) mild COVID-19, based on mild hypoxemia ($>94\%$) O₂ saturation, without need for O₂ therapy; 2) moderate COVID-19, based on hypoxemia ($\leq 94\%$) O₂ saturation, and need for O₂ therapy by nasal cannula (up to 5 l/min), or pneumonia on imaging; 3) severe COVID-19, based on the use of non-rebreather mask (NRB) (>6 l/min), or invasive mechanical ventilation (IMV). Intubation status and O₂ therapy modality were obtained by clinical notes and evaluation of each patient during

hospitalization. SOFA score was also calculated for all patients (<https://www.mdcalc.com/sequential-organ-failure-assessment-sofa-score>). Mild COVID-19 patients were not included in this study, as they did not require hospitalization.

Metabolome analysis

For metabolites extraction, 100 µl of saliva was added with 1000 µl of methanol, ultrapure, spectroscopic grade methanol (HPLC grade), in a 2 ml eppendorf. This mixture was homogenized by vortexing for 5 min. In sequence, it was centrifuged for 15 min at 13000 g and the supernatant, containing metabolites of diverse chemical natures, transferred to another eppendorf which was concentrated in a vacuum concentrator for 30 min and lyophilized. The material was resuspended in 400 µl of methanol, HPLC grade, and filtered through a syringe tip filter with a pore size of 0.22 micrometers.

Analyzes were carried out using a high-performance liquid chromatography hyphenated to the mass spectrometer (HPLC/MS), in an HPLC (Agilent™, model Infinity 1260) hyphenated to the Agilent 6520B high-resolution Q-TOF mass spectrometer with electrospray ionization source (ESI). Chromatographic parameters were: Agilent Poroshell column with a 2.1 mm internal diameter, 5 cm long, 2.7 µm particles; mobile solutions used in the gradient elution (mobile phase) were: (A) water acidified with 0.1% (v/v) formic acid, and (B) methanol, with the gradient: 10% of B (0 min), 98% of B (0-4 min), 98% of B (4-10 min). Ionization parameters were: nebulizer pressure of 25 psi, drying gas flow rate at 8 l/min at a temperature of 220°C and an energy of 4.5 kV was applied to the capillary.

Statistical analysis

Agilent MassHunter Workstation software of qualitative method analysis v. 10.0 was used to process the raw data. The “Molecular feature extraction (MFE)” tool was used to extract the mass spectra and convert them to “.CEF” extension. Agilent Mass Profiler Professional (MPP) v. B.13.1.1 software was used to filter and analyze the extracted molecular compounds. The filters used were: minimum absolute abundance = 5000 counts and all permitted loads. The analysis parameters were: retention time tolerance of 0.15 min; mass tolerance window 15 ppm + 2mDa. Molecular compounds present in 75% of at least one group with a coefficient of variation below 25% were retained for further data analysis.

Statistical analyzes were performed with log2 transformed values. Unpaired T-test and Benjamini-Hochberg were applied as correction tests with $p < 0.05$ and fold change equal to or greater than 2.00. Metabolites were identified by using the METLIN 2019 database (included in the MPP).

Patient characteristics were summarized using standard descriptive statistics: mean value/standard deviation for continuous variables and count/percent for categorical variables. The normality of the data was assessed using Kolmogorov-Smirnov test and graphically through histograms. The non-normal distribution data were transformed with Z score. General Linear Model (GLZM) was used to compare clinical and metabolic profile with adequation of distribution of each dependent variable. Categorical variables were evaluated with Chi Square test; Exact’s Fisher.

Results

Patients

A total of 107 adult inpatients were included, all of them with a positive SARS-COV-2 RT-PCR. Patient's characteristics on timepoints admission, saliva collection and at the endpoint of hospitalization are summarized in Table 1. Regarding race/ethnicity, 79.4% were White, 18.7% Hispanic and 1.9% African American, with no differences between categories. The most prevalent comorbidities were hypertension (52%), followed by obesity (39%) and diabetes (26%), and among diabetic patients, 3.7% were insulin dependent on admission. 3.7% of chronic kidney disease (CKD) patients were undergoing hemodialysis prior to hospitalization. No differences were found on admission between categories, except for age (older patients had higher mortality), gender [73% of males required IMV (p.0.042)], and for smoking history for all categories.

Table 1. Demographics and Baseline Characteristics of COVID-19 Patients. Data on timepoints: admission, saliva collection and at the end of hospitalization.

Timepoint - Admission	Total (n=107)	Moderate (n=55)	Severe (n=52)	p value	Discharge (n=88)	Death (n=19)	p value	No IMV (n=81)	IMV (n=26)	p value
Age (year) Mean $\pm$ SD (Range)	58 $\pm$ 14 (26 - 93)	57 $\pm$ 15 (26 - 93)	59 $\pm$ 11 (39 - 87)	0.36	57 $\pm$ 14 (26 - 93)	64 $\pm$ 11 (39 - 87)	0.04	57 $\pm$ 14 (26 - 93)	62 $\pm$ 10 (39 - 87)	0.11
Male - n. (%)	59 (55%)	27 (49%)	32 (61%)	0.19	45 (51%)	14 (74%)	0.082	40 (49%)	19 (73%)	0.042
Weight (kg) - Mean $\pm$ SD	83 $\pm$ 17	84 $\pm$ 18	81 $\pm$ 15	0.54	84 $\pm$ 18	78 $\pm$ 10	0.23	84 $\pm$ 17	81 $\pm$ 17	0.69
Comorbidities - n. (%)										
Hypertension	56 (52%)	31 (56%)	25 (48%)	0.39	47 (53%)	9 (47%)	0.80	43 (53%)	13 (50%)	0.82
Obesity (BMI $\geq$ 30)	33 (39%)	20 (43%)	13 (35%)	0.49	30 (43%)	3 (21%)	0.23	27 (42%)	6 (30%)	0.43
Diabetes	28 (26%)	17 (31%)	11 (21%)	0.25	23 (26%)	5 (26%)	1.0	23 (28%)	5 (19%)	0.44
Dyslipidemia	19 (18%)	10 (18%)	9 (17%)	0.90	17 (19%)	2 (10%)	0.51	15 (18%)	4 (15%)	1.0
CHF	15 (14%)	9 (16%)	6 (11%)	0.47	11 (12%)	4 (21%)	0.46	10 (12%)	5 (19%)	0.51
Pulmonary diseases	13 (12%)	6 (11%)	7 (13%)	0.68	8 (9%)	5 (26%)	0.052	8 (10%)	5 (19%)	0.29

CKD	12 (11%)	9 (16%)	3 (6%)	0.12	10 (11%)	2 (10%)	1.0	10 (12%)	2 (8%)	0.72
Hypothyroidism	7 (6%)	4 (7%)	3 (6%)	1.0	6 (7%)	1 (5%)	1.0	6 (7%)	1 (4%)	1.00
Cancer (active)	4 (4%)	0 (0.0%)	4 (8%)	0.052	2 (2%)	2 (10%)	0.14	2 (2%)	2 (8%)	0.24
Smoking – current	1 (1%)	1 (2%)	0 (0%)	0.36	1 (1%)	0 (0%)	0.32	1 (1%)	0 (0%)	0.18
Smoking – history	16 (15%)	4 (7%)	12 (23%)	0.03	9 (10%)	7 (37%)	0.001	8 (10%)	8 (31%)	0.01
Alcohol intake	12 (11%)	6 (11%)	6 (11%)	0.58	10 (11%)	2 (10%)	0.20	8 (10%)	4 (15%)	0.165
Timepoint - Saliva collection - Mean ± SD										
Days of symptoms	12.2 ± 4.1	11.8 ± 3.8	12.7 ± 4.4	0.23	12.5 ± 4.1	11.0 ± 4.3	0.17	13.0 ± 2.2	12.5 ± 3.9	0.10
SOFA score	2.2 ± 1.2	1.9 ± 1.2	2.5 ± 1.2	0.40	2.0 ± 1.1	3.0 ± 1.5	0.001	2.0 ± 1.1	2.8 ± 1.4	0.003
PaO ₂ /FiO ₂ - mmHg	224.1 ± 106.3	283.8 ± 106.5	165.6 ± 66.7	<0.001	237.4 ± 107.3	173.5 ± 87.8	0.01	244.2 ± 109.3	173.7 ± 79.9	0.004
Timepoint - End of hospitalization										
Hospital stay (days) - Mean ± SD	13.9 ± 14.0	9.3 ± 8.9	18.8 ± 16.7	<0.001	12.1 ± 12.0	22.7 ± 119.2	0.02	10.1 ± 8.0	26.0 ± 20.7	<0.001
Intensive care stay (days) - Mean ± SD	5.8 ± 12.4	0.7 ± 1.6	11.3 ± 16.1	<0.001	2.9 ± 7.0	19.6 ± 20.7	<0.001	1.3 ± 2.8	20.0 ± 18.9	<0.001
Acute HD - n. (%)	4 (4%)	2 (4%)	2 (4%)	0.95	2 (2%)	2 (10%)	0.14	2 (2%)	2 (77%)	0.24
Acute Stroke - n. (%)	1 (1%)	1 (2%)	0 (0%)	0.36	1 (1%)	0 (0%)	0.14	1 (1%)	0 (0%)	0.18
IMV - n. (%)	26 (24%)	0 (0%)	26 (50%)	0.000	7 (6%)	19 (100%)	<0.001			
NIV - n. (%)	55 (51%)	13 (24%)	42 (81%)	0.000	39 (44%)	16 (84%)	0.002	34 (42%)	21 (81%)	0.001
Tracheostomy - n.(%)	7 (6%)	0 (0%)	7 (13%)	0.005	1 (1%)	6 (32%)	<0.001	0 (0%)	7 (27%)	<0.001
Chest CT (% of involvement) n. (%)										
0-24%	5 (6%)	3 (7%)	2 (5%)	0.77	4 (5%)	1 (8%)	0.70	3 (4%)	2 (10%)	0.49
25-50%	32 (37%)	17 (39%)	15 (35%)		28 (38%)	4 (31%)		26 (39%)	6 (32%)	
51-100%	49 (57%)	23 (53%)	26 (60%)		41 (56%)	8 (61%)		38 (57%)	11 (58%)	

n. (%), number; SD, standard deviation; BMI, body mass index; CHF, chronic heart failure; CKD, chronic kidney disease; SOFA, sepsis-related organ failure assessment; PaO₂, arterial oxygen tension; FiO₂, fraction of inspired oxygen; HD, Hemodialysis; IMV, invasive mechanical ventilation; NIV, non-invasive ventilation; CT, computer tomography (worst CT evaluated in the timepoint end of hospitalization).

The signals and symptoms of COVID-19 reported by the patient on admission were described as: desaturation (95.3%), dyspnea (71.0%), cough (57.9%), fever (51.4%), fatigue (29.9%), myalgia (28.0%), hyporexia (17.8%), headache (14.0%), rhinorrhea (13.1%), chest pain (12.1%), nausea (10.3%), odynophagia (8.4%), dysgeusia (7.5%), diarrhea (7.5%), anosmia (6.5%), vomiting (4.7%) and abdominal pain (1.9%). We found similar characteristics in majority of the variables analyzed in all categories, except for moderate versus severe COVID-19: desaturation (moderate (90.9%), severe (100%), p 0.02), dyspnea (moderate (58.2%), severe (84.6%), p 0.003), and fatigue (moderate (40%), severe (19.2%), p 0.01).

At the timepoint of saliva collection, we found a significant difference for days of symptoms, SOFA score, and oxygenation index ($\text{PaO}_2/\text{FiO}_2$). There was a significant difference in liver function for: discharge and death [total bilirubin (p 0.02) and indirect bilirubin (p 0.01)], and need of IMV categories [total bilirubin (p 0.01) and direct bilirubin (p 0.03)] presented in S1 Table. At the timepoint end of hospitalization, we found statistical difference for hospital stay, intensive care stay, invasive and non-invasive ventilation and tracheostomy.

Timepoint - Saliva collection - Mean $\pm$ SD	Moderate (n=55)	Severe (n=52)	p value	Discharge (n=88)	Death (n=19)	p value	No IMV (n=81)	IMV (n=26)	p value
O_2 (l/min)	1.8 $\pm$ 1.5	5.4 $\pm$ 3.7	<0.001	3.2 $\pm$ 2.9	5.2 $\pm$ 4.6	0.02	3.1 $\pm$ 2.8	5.0 $\pm$ 4.3	0.008
FiO_2	28.3 $\pm$ 6.2	43.5 $\pm$ 16.8	<0.001	34.0 $\pm$ 12.2	43.3 $\pm$ 21.6	0.01	33.5 $\pm$ 11.7	42.6 $\pm$ 20.0	0.005
PaO_2	75.3 $\pm$ 24.2	64.3 $\pm$ 12.3	0.007	71.9 $\pm$ 20.6	61.5 $\pm$ 13.7	0.04	72.8 $\pm$ 21.4	62.0 $\pm$ 12.2	0.01
GPT	63.0 $\pm$ 64.0	44.0 $\pm$ 23.6	0.27	57.0 $\pm$ 52.4	35.9 $\pm$ 12.7	0.38	58.2 $\pm$ 55.7	42.2 $\pm$ 21.7	0.41
GOT	60.6 $\pm$ 57.5	45.3 $\pm$ 22.8	0.36	55.1 $\pm$ 47.1	41.7 $\pm$ 17.6	0.58	55.1 $\pm$ 50.0	47.3 $\pm$ 18.0	0.69

GGT	213.6 ± 178.7	159.0 ± 75.0	0.26	195.9 ± 150.4	150.3 ± 79.2	0.38	196.1 ± 158.6	168.1 ± 87.2	0.41
ALP	100.7 ± 57.0	78.3 ± 22.4	0.16	90.9 ± 48.2	87.3 ± 31.4	0.86	92.7 ± 50.9	83.7 ± 26.2	0.62
Total bilirubin	0.4 ± 0.3	0.6 ± 0.6	0.18	0.5 ± 0.3	0.9 ± 0.7	0.02	0.4 ± 0.2	0.8 ± 0.6	0.01
Direct bilirubin	0.3 ± 0.2	0.4 ± 0.4	0.11	0.3 ± 0.3	0.6 ± 0.6	0.10	0.2 ± 0.1	0.6 ± 0.5	0.03
Indirect bilirubin	0.2 ± 0.2	0.2 ± 0.1	0.72	0.1 ± 0.1	0.3 ± 0.2	0.01	0.1 ± 0.1	0.2 ± 0.2	0.15
Urea	42.6 ± 29.5	43.4 ± 33.0	0.89	41.1 ± 29.7	52.4 ± 36.7	0.16	41.0 ± 30.8	49.4 ± 31.8	0.24
Creatinine	1.2 ± 1.4	1.3 ± 1.8	0.90	1.1 ± 1.2	1.9 ± 2.9	0.06	1.1 ± 1.2	1.6 ± 2.4	0.20

S1 Table. Clinical Data on timepoint saliva collection. O₂, oxygen; l/min, liter per minute; PaO₂, arterial oxygen tension; FiO₂, fraction of inspired oxygen; GPT, glutamate-pyruvate transaminase (also called ALT, alanine aminotransferase); GOT, glutamic-oxaloacetic transaminase (also called AST, aspartate aminotransferase); GOT, glutamic-oxaloacetic transaminase; GGT, gamma-glutamyl transferase; ALP, alkaline phosphatase.

All patients classified as moderate and 63.5% of severe COVID were discharged from hospital ($p < 0.001$), with 36.5% deaths in the severe COVID category, $p < 0.001$. The majority of patients submitted to non-invasive ventilation (73.1%) were discharged from hospital. It is observed that from 107 patients studied, 19 (17.8%) died, and among the 26 (24%) who required IMV, 73.1% died, $p < 0.001$.

Metabolomic profile in COVID-19 saliva

The 107 patients were categorized into (i) moderate and severe COVID-19, (ii) discharge and death, and (iii) no IMV vs IMV categories and their saliva samples were submitted for metabolomics analysis. Altogether, in the moderate vs severe COVID-19 analysis, 42 out of 1733 entities were significantly differentiated after applying filtration using frequency (appeared in more than 75% of samples in at least one category of samples), with 7 metabolites remaining after applying $p\text{-value} < 0.05$ and fold change > 2.0 ; in the discharge vs death analysis, 43 out of 1733 compounds

were identified, with 2 metabolites remaining after applying the filters mentioned above; and in the no IMV vs IMV analysis, 35 out of 1733 entities were significantly differentiated after applying filtration methods (Table 2).

Table 2. List of differentiative metabolites among COVID-19 saliva patients. Categories: moderate and severe COVID-19 (7 compounds); discharge and death (2 compounds); and no IMV and IMV (4 compounds) at $p < 0.05$ and fold change > 2.0 .

No.	Compounds	p (Corr)	FC (Moderate vs Severe COVID-19)
1	6-Thiouric acid	0.01	-2.969494
2	17-Phenyl trinor prostaglandin E2 serinol amide	<0.001	144.28325
3	25-Hydroxy-cholesterol(d3)	0.01	20.5328
4	Undecylprodigiosin	<0.001	18.430672
5	(1S,Z)-3-(2R,E)-6-Hydroxy-2,5-dimethylnon-4-en-1-ylidene)-1-methyloctahydro-2H-quinolizin-1-ol	0.02	14.625197
6	Methyl 1-(methylsulfinyl) propyl disulfide	0.01	-8.487999
7	Digitoxigenin 3-[glucosyl-(1-6)-glucosyl-(1-4)-2,6-dideoxyribohexoside]	<0.001	1139.8263
No.	Compounds	p (Corr)	FC (Discharge vs Death)
1	6-Thiouric acid	0.001	5.8959403
2	17-phenyl trinor prostaglandin E2 serinol amide	0.04	-80.30943
No.	Compounds	p (Corr)	FC (no IMV vs IMV)
1	3-Octanone	0.003	19.51051
2	6-Thiouric acid	0.002	4.447275
3	Undecylprodigiosin	0.005	-9.37128
4	Methyl 1-(methylsulfinyl) propyl disulfide	0.03	9.018281

No., number; Corr, corrected; vs, versus; FC, fold change; IMV, invasive mechanical ventilation.

Metabolites that were statistically different, represented by the average of the normalized intensity of the compound were presented in a volcano plot and a heatmap among all categories (moderate vs severe COVID-19; discharge vs death; no IMV vs IMV) of COVID-19 saliva analysis (Figure 2).

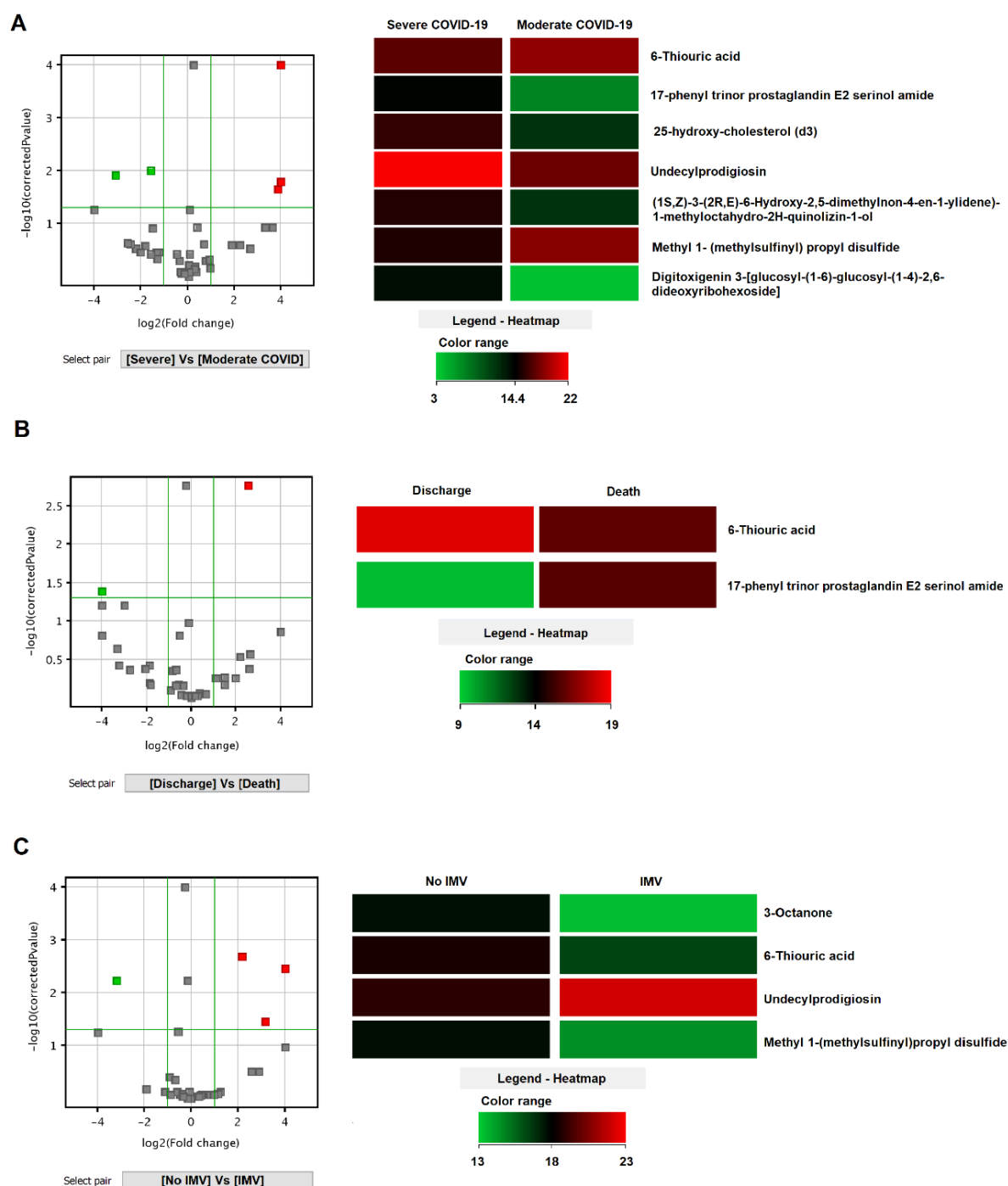


Fig 2. Volcano plot and heatmap of COVID-19 patients. (A) Moderate vs severe COVID-19: volcano plot based on fold change and adjusted p-value, highlighting 7 metabolites in the heatmap by colors; (B) Discharge vs death: volcano plot based on fold change and adjusted p-value, highlighting 2 metabolites in the heatmap by colors; (C) No IMV vs IMV: volcano plot based on fold change and adjusted p-value, highlighting 4 metabolites in the heatmap by colors.

Figure 3 represents the normalized intensity values in Log 2 of metabolites among all categories (moderate vs severe COVID-19; discharge vs death; no IMV vs

IMV) of COVID-19 saliva analysis. Individual heatmaps of moderate vs severe COVID, discharge vs death and no IMV vs IMV are presented in S1 Figure.

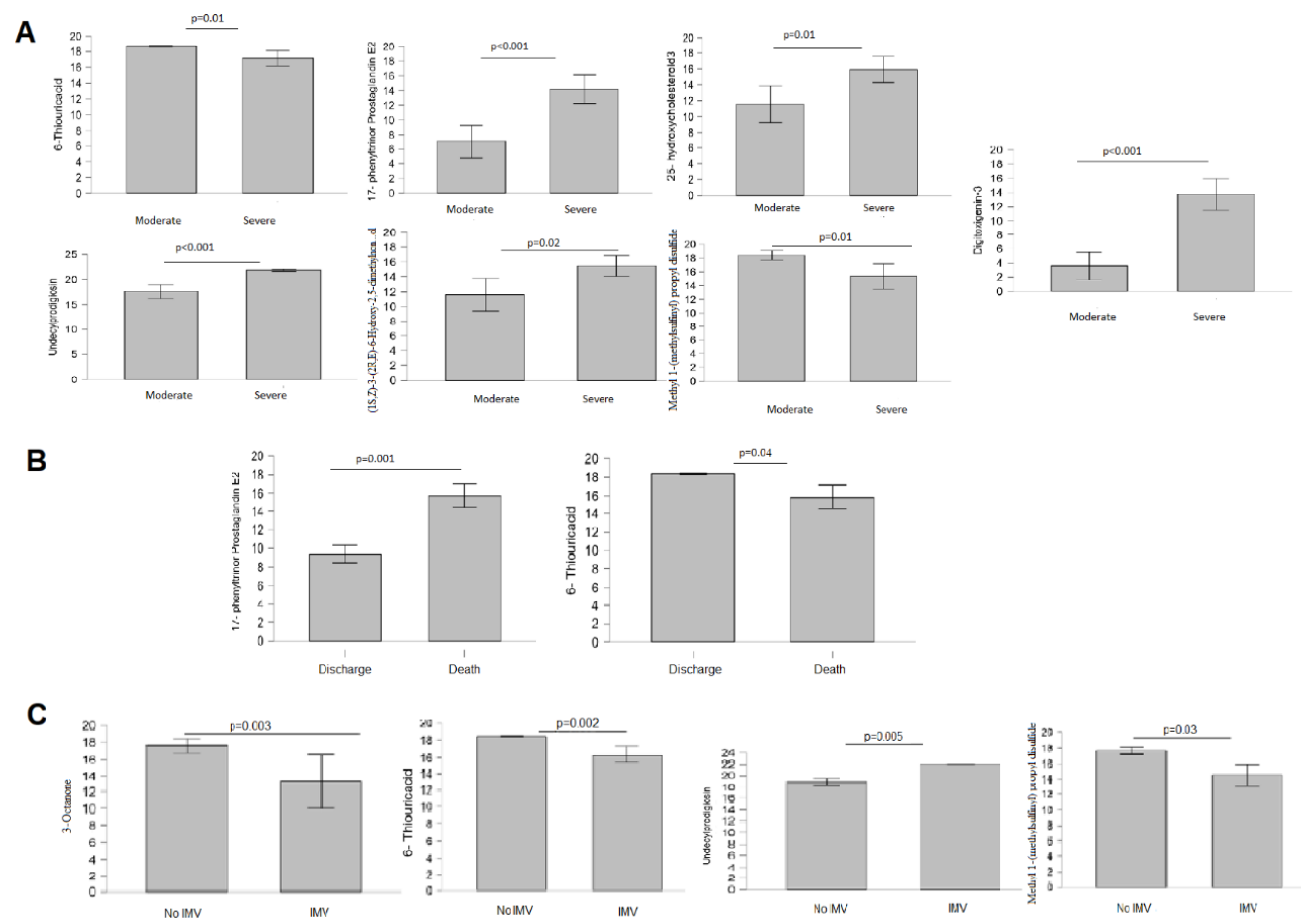
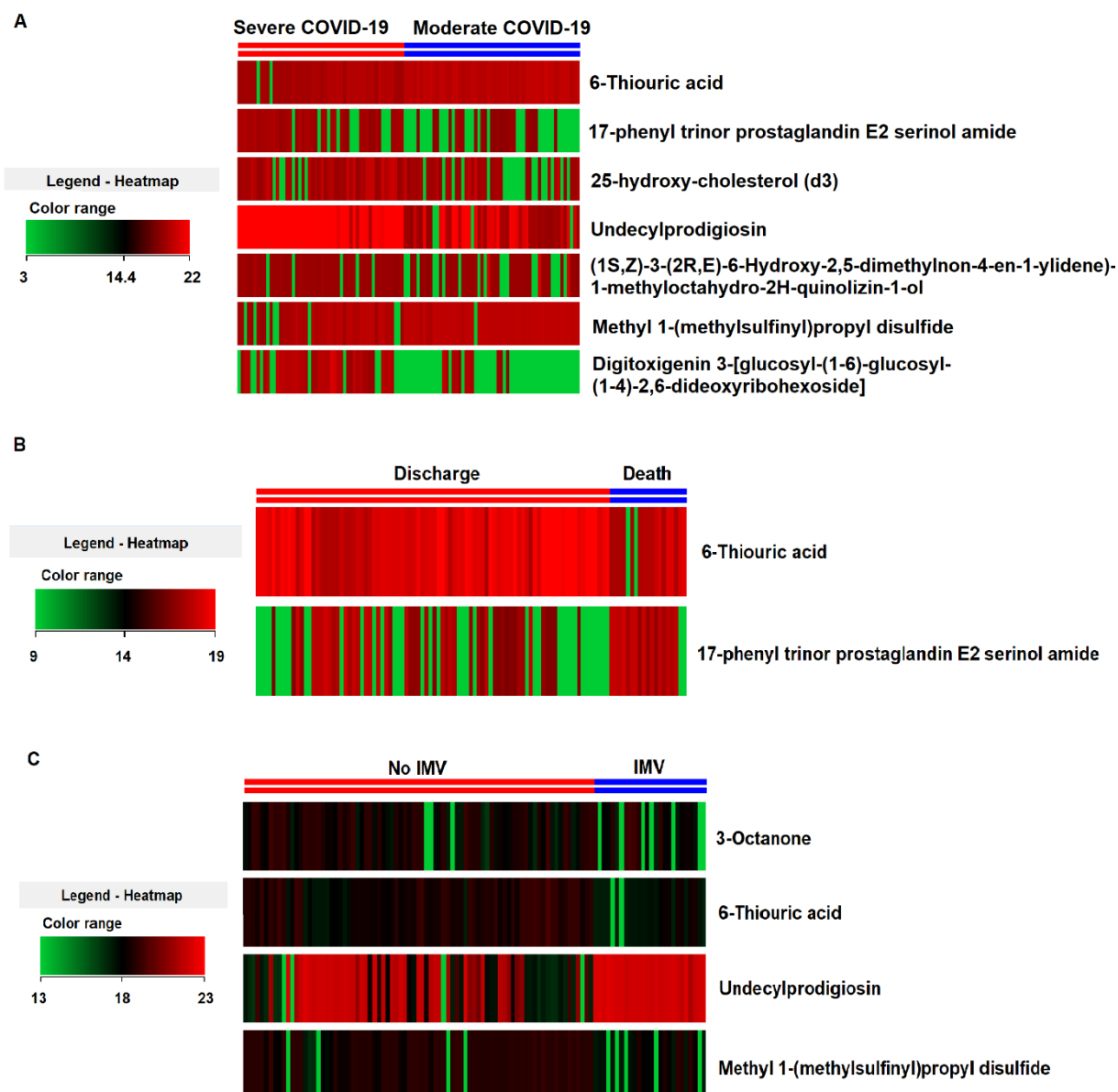


Fig 3. Normalized intensity values in Log 2 of COVID-19 saliva metabolites. (A) Moderate and severe COVID-19; (B) Discharge vs death; (C) No IMV vs IMV.



S1 Fig. Individual heatmap of severe COVID-19 versus moderate COVID-19, discharge versus death and no IMV versus IMV.

Discussion

To our knowledge, this study is the first to evaluate the saliva metabolome profile of COVID-19 patients categorizing them into moderate and severe COVID, discharge and death and by the need for IMV. Our findings revealed the main saliva metabolites related to disease clinical severity and helped us to identify its relation to COVID-19.

Characteristics of the study population

The population in this study comprised inpatients stratified in three categories analyzed at the end of hospitalization period. None of the variables analyzed in the three timepoints (admission, time of saliva collection and end of hospitalization) were able to predict the evaluated outcomes defined as severity, death or need of IMV. On admission, we found worst outcome for smoking history, males and older patients whose mortality was higher. BMI and diabetes were similar in both groups, thus limiting the putative repercussions of differential complications of diabetes on metabolome pattern [25-27]. Altogether, the matching of all categories suggested the absence of confusion factors that could alter metabolomics findings in our study.

There is a consensus that hyperinflammation results in a severe condition in COVID-19 [24, 28, 29] and is related to clinical (older age, male sex, previous comorbidities as hypertension, diabetes and obesity, among others) [30-32] and biological factors (inflammation and coagulation disorders evidenced by dysregulation of macrophage, platelet degranulation, complement system pathways, and massive metabolic suppression) [13, 33-34]. Greater severity of COVID-19 and mortality in patients with diabetes, obesity or older age indicate that the virus may trigger metabolic dysregulation, and consequently, change the circulating metabolites in the body, and should be further investigated [6, 32], since no other isolated parameter was able to predict outcome.

Data on saliva collection revealed the ability of SOFA score to discriminate between discharge and death, and no IMV versus IMV categories early in hospitalization. However, there is no consensus in the literature that the SOFA score applied at the beginning of hospitalization is a good parameter for determining

prognosis. Rapid change in SOFA score ≥ 2 points after infection is considered a clinical criterion for sepsis-associated organ dysfunction [35]. On the other hand, authors argue that severe COVID-19 is not a limited lung injury but a multi-organ involvement disease, and a multi-organ failure assessment tool as SOFA score showed a significant association ($p < 0.001$) with in hospital mortality [35, 36]. We believe that SOFA score is a useful tool that can be better explored as a complementary prognostic method in metabolomic analysis.

Saliva metabolite profiling of COVID-19 patients

The 6-Thiouric acid, chemical formula ($C_5H_4N_4O_2S$), belongs to the class of organic compounds known as purinones. Thiopurine analogs, have been reported to inhibit SARS-CoV papain-like proteases (PLpro) and is considered an important antiviral target [37, 38]. The PLpro, produced by coronaviruses, have significant functional implications in the innate immune response during SARS-CoV-2 infection, and it participates in the viral replication process and pathogenesis mainly by converting viral polyprotein into a functional replicase complex [39]. The PLpro together with 3C-like protease (3CLpro) proteolytically cleave the coronaviral polyprotein and release non-structural proteins. 6-mercaptopurine (6MP) and 6-thioguanine (6TG) are purine analogues, effective in the treatment of patients with acute lymphoblastic or myeloblastic leukemia [40], that has a thiocarbonyl moiety which is the active pharmacophore for the inhibition of PLpro [37] 6-MP is inactivated by xanthine oxidase into 6-thiouric acid [41]. According to our results, 6-thiouric acid has a protective function in the body, being downregulated in severe COVID-19 patients (-2.96 times) and overregulated in the discharge (5.89 times) and no-IMV (4.44 times) categories.

The 17-phenyl trinor prostaglandin E2 serinol amide, chemical formula ($C_{26}H_{37}NO_6$), is a stable analog of prostaglandin E2 (PGE2) [42]. Derived from arachidonic acid (AA) metabolism, the PGE2 participate in the initiation of inflammation and regulation of immune response, hematopoiesis, and macrophage facilitation [43, 44]. PGE2 is abundantly detected in many tissues and has four receptor subtypes (EP1-4), responsible for different pathophysiological actions such as increasing vascular permeability, cell proliferation and migration, and control of vascular smooth muscle tone [45, 46]. PGE2 has been shown to exacerbate inflammatory responses and immune diseases [45], and participate in the release process of interleukin 6 (IL-6) and tumor necrosis factor alpha (TNF α) [42, 46]. The influx of inflammatory cells (monocytes and neutrophils) and release of inflammatory cytokines and reactive oxygen species are related to ground-glass opacities on computer tomography (CT), pulmonary edema of alveolar spaces, and acute respiratory distress syndrome (ARDS) in COVID-19 patients [44]. Our results showed that 17-phenyl trinor prostaglandin E2 serinol amide was overregulated (144.28 times) in the severe COVID compared to moderate category, also appearing downregulated in the discharge category (-80.30 times).

Cholesterol is a sterol synthesized by mammalian cells and forms an essential component of the cell membrane, responsible for controlling membrane permeability and integrity of cellular signaling. In addition, it also acts as an important precursor for several molecules, such as steroid hormones, bile acid, vitamin D, and oxysterols. Oxysterols are important for regulating cholesterol homeostasis and has multiple immunoregulatory functions, important for immunity and inflammation [47]. The oxysterol 25-Hydroxy-cholesterol (25-HC), is a potent SARS-CoV-2 inhibitor and is described in the literature as a marker for severe illness of COVID-19 as it appears

in high concentrations in fatal cases [48]. 25-HC showed the inhibitory activity against infections by SARS-CoV-2 by modifying the cholesterol composition on the plasma membrane, blocking viral membrane fusion through the suppression of cholesterol export, and also modifying cholesterol intracellular content to reduce viral replication [47, 49]. However, it also enhances the reactive oxygen species (ROS) production within a short period of time in human neutrophils resulting in a pro-inflammatory effect [50]. In our study, the 25-HC was overregulated (20.53 times) in the severe COVID compared to moderate category.

Undecylprodigiosin (UP), chemical formula ($C_{25}H_{35}N_3O$), is biosynthesized by *Streptomyces sp.* strains, displaying antioxidant, anti-microbial, anti-malarial, immunosuppressive and anticancer properties [51, 52]. Ribosome may be the potential target of UP in cancer cells, but its mechanisms of action are still unclear [53]. No studies were found relating UP with SARS-CoV-2, COVID-19 or saliva. Our study revealed that UP is overregulated (18.43 times) in severe COVID-19 patients and downregulated (-9.37 times) in no IMV patients.

The metabolite (1S,Z)-3-(2R,E)-6-Hydroxy-2,5-dimethylnon-4-en-1-ylidene)-1-methyloctahydro-2H-quinolizin-1-ol was overregulated in severe COVID-19 patients (14.62), but no studies were found with the compound.

Methyl 1-(methylsulfinyl) propyl disulfide, chemical formula ($C_5H_{12}OS_3$), belongs to the class of organic compounds known as sulfoxides, and has been detected in several different foods, such as onion-family vegetables. Based on a literature review, very few articles have been published on Methyl 1-(methylsulfinyl) propyl disulfide and no relation was found with COVID-19. In our study, the compound was downregulated (-8.48 times) in severe COVID-19 patients compared to moderate, and overregulated in no IMV category (9.01 times).

Digitoxigenin 3-[glucosyl-(1-6)-glucosyl-(1-4)-2,6-dideoxyribohexoside], chemical formula ($C_{41}H_{64}O_{17}$), belongs to the class of organic compounds known as cardenolide glycosides and derivatives, and can be found in tea, herbs, spices, and green vegetables. Cardenolide glycosides are often toxic and result in heart-arresting through inhibition of the enzyme $Na^+/K^+ - ATPase$, which is responsible for maintaining the sodium and potassium ion gradients across the cell membranes [54]. The metabolite was overregulated (1139.82 times) in severe COVID-19 patients in our study.

3-Octanone, chemical formula ($C_8H_{16}O$), belongs to the class of ketones, and is a main bioactive constituent in *Schizonepeta tenuifolia*, a traditional Chinese herb found in a respiratory detox shot (RDS) with antioxidant, anti-inflammatory, and antiviral activities with ability to block the infection of SARS-CoV-2 [55]. However, the components contributing to its antioxidant and anti-inflammatory activities remain unclear [56]. In our study, 3-Octanone was overregulated (19.51 times) in no IMV category.

Study limitations

Some limitations must be considered when interpreting the results of this study. Control group analyzes were not performed, and could be used to verify the presence of different metabolites between healthy and sick participants, and better elucidate the metabolic pathways that are dysregulated in COVID-19.

Conclusion

To our knowledge, this study is the first to evaluate the saliva metabolome profile of COVID-19 patients categorizing them into moderate and severe COVID, discharge and death and by the need for IMV. Saliva metabolomic analysis can provide an important tool that remains to be explored in prospective studies, but to date, we report a significant saliva metabolome profile of COVID-19 patients with involvement of purines, amide, lipid, sulfoxide, glycoside and ketones. Saliva metabolome allowed us to identify metabolites related to greater disease severity in the evaluated outcomes which might be used in selection of potential saliva biomarker for severity evaluation early in hospitalization and open new therapeutic perspectives against COVID-19.

Acknowledgments

To all colleagues and patients who contributed directly or indirectly to the completion of this work.

References

1. WHO. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). 2020. Available from: [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)).
2. WHO. WHO Coronavirus (COVID-19) Dashboard. 2022. Available from: <https://covid19.who.int/>. Accessed 2022 Nov, 28.
3. Chandna A, Mahajan R, Gautam P, Mwandigha L, Gunasekaran K, Bhusan D, et al. Facilitating Safe Discharge Through Predicting Disease Progression in Moderate Coronavirus Disease 2019 (COVID-19): A Prospective Cohort Study to Develop and Validate a Clinical Prediction Model in Resource-Limited Settings. *Clin Infect Dis*. 2022 Aug 24;75(1):e368-e379. doi: 10.1093/cid/ciac224
4. Bi X, Liu W, Ding X, Liang S, Zheng Y, Zhu X, et al. Proteomic and metabolomic profiling of urine uncovers immune responses in patients with COVID-19. *Cell Rep*. 2022 Jan 18;38(3):110271. doi: 10.1016/j.celrep.2021.110271

5. Battaglini D, Lopes-Pacheco M, Castro-Faria-Neto HC, Pelosi P, Rocco PRM. Laboratory Biomarkers for Diagnosis and Prognosis in COVID-19. *Front Immunol.* 2022; 13:857573. doi: 10.3389/fimmu.2022.857573
6. Blasco H, Bessy C, Plantier L, Lefevre A, Piver E, Bernard L, et al. The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism. *Sci Rep.* 2020 Oct 8;10(1):16824. doi: 10.1038/s41598-020-73966-5
7. Pozzi C, Levi R, Braga D, Carli F, Darwich A, Spadoni I, et al. A 'Multiomic' Approach of Saliva Metabolomics, Microbiota, and Serum Biomarkers to Assess the Need of Hospitalization in Coronavirus Disease 2019. *Gastro Hep Adv.* 2022;1(2):194-209. doi: 10.1016/j.gastha.2021.12.006
8. Pfaffe T, Cooper-White J, Beyerlein P, Kostner K, Punyadeera C. Diagnostic potential of saliva: current state and future applications. *Clin Chem.* 2011 May;57(5):675-87. doi: 10.1373/clinchem.2010.153767. Epub 2011 Mar 7. PMID: 21383043.
9. Fernandes LL, Pacheco VB, Borges L, Athwal HK, de Paula Eduardo F, Bezinelli L, et al. Saliva in the Diagnosis of COVID-19: A Review and New Research Directions. *J Dent Res.* 2020 Dec;99(13):1435-1443. doi: 10.1177/0022034520960070
10. Costa Dos Santos Junior G, Pereira CM, Kelly da Silva Fidalgo T, Valente AP. Saliva NMR-Based Metabolomics in the War Against COVID-19. *Anal Chem.* 2020 Dec 15;92(24):15688-15692. doi: 10.1021/acs.analchem.0c04679
11. Su Y, Chen D, Yuan D, Lausted C, Choi J, Dai CL, et al. Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. *Cell.* 2020 Dec 10;183(6):1479-1495.e20. doi: 10.1016/j.cell.2020.10.037
12. Song JW, Lam SM, Fan X, Cao WJ, Wang SY, Tian H, et al. Omics-Driven Systems Interrogation of Metabolic Dysregulation in COVID-19 Pathogenesis. *Cell Metab.* 2020 Aug 4;32(2):188-202.e5. doi: 10.1016/j.cmet.2020.06.016
13. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell.* 2020 Jul 9;182(1):59-72.e15. doi: 10.1016/j.cell.2020.05.032
14. Spick M, Lewis HM, Frampas CF, Longman K, Costa C, Stewart A, et al. An integrated analysis and comparison of serum, saliva and sebum for COVID-19 metabolomics. *Sci Rep.* 2022 Jul 13;12(1):11867. doi: 10.1038/s41598-022-16123-4
15. Bruzzone C, Bizkarguenaga M, Gil-Redondo R, Diercks T, Arana E, García de Vicuña A, et al. SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. *iScience.* 2020 Oct 23;23(10):101645. doi: 10.1016/j.isci.2020.101645
16. Thomas T, Stefanoni D, Reisz JA, Nemkov T, Bertolone L, Francis RO, et al. COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. *JCI Insight.* 2020 Jul 23;5(14):e140327. doi: 10.1172/jci.insight.140327
17. Wu D, Shu T, Yang X, Song JX, Zhang M, Yao C, et al. Plasma metabolomic and lipidomic alterations associated with COVID-19. *Natl Sci Rev.* 2020 Jul;7(7):1157-1168. doi: 10.1093/nsr/nwaa086
18. Song Z, Bao L, Deng W, Liu J, Ren E, Lv Q, et al. Integrated histopathological, lipidomic, and metabolomic profiles reveal mink is a useful animal model to mimic the pathogenicity of severe COVID-19 patients. *Signal Transduct Target Ther.* 2022 Jan 28;7(1):29. doi: 10.1038/s41392-022-00891-6

19. Sapkota D, S land TM, Galtung HK, Sand LP, Gianneccchini S, To KKW, et al. COVID-19 salivary signature: diagnostic and research opportunities. *J Clin Pathol*. 2020 Aug 7;jclinpath-2020-206834. doi: 10.1136/jclinpath-2020-206834
20. Iebba V, Zanutta N, Campisciano G, Zerbato V, Di Bella S, Cason C, et al. Profiling of Oral Microbiota and Cytokines in COVID-19 Patients. *Front. Microbiol*. 2021;12:671813. doi: 10.3389/fmicb.2021.671813
21. Hao M, Wang D, Xia Q, Kan S, Chang L, Liu H, et al. (2022) Pathogenic Mechanism and Multi-omics Analysis of Oral Manifestations in COVID-19. *Front. Immunol*. 2022;13:879792. doi: 10.3389/fimmu.2022.879792
22. Paschotto DR, Pupin B, Bhattacharjee TT, Soares LES. Saliva Preparation Method Exploration for ATR-FTIR Spectroscopy: Towards Bio-fluid Based Disease Diagnosis. *Anal Sci*. 2020 Sep 10;36(9):1059-1064. doi: 10.2116/analsci.20P029
23. Mikkonen JJ, Raittila J, Rieppo L, Lappalainen R, Kullaa AM, Myllymaa S. Fourier Transform Infrared Spectroscopy and Photoacoustic Spectroscopy for Saliva Analysis. *Appl Spectrosc*. 2016 Sep;70(9):1502-10. doi: 10.1177/0003702816654149
24. Del Valle DM, Kim-Schulze S, Huang HH, Beckmann ND, Nirenberg S, Wang B, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat Med*. 2020 Oct;26(10):1636-1643. doi: 10.1038/s41591-020-1051-9
25. Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. *Diabetes Metab Res Rev*. 2020 Mar 31;36(7):e3319. doi: 10.1002/dmrr.3319.
26. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*. 2020 May 26;323(20):2052-2059. doi: 10.1001/jama.2020.6775. Erratum in: *JAMA*. 2020 May 26;323(20):2098.
27. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis*. 2020 May;94:91-95. doi: 10.1016/j.ijid.2020.03.017
28. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet*. 2020 Mar 28;395(10229):1033-1034. doi: 10.1016/S0140-6736(20)30628-0
29. Kaur S, Bansal R, Kollimuttathuillam S, Gowda AM, Singh B, Mehta D, et al. The looming storm: Blood and cytokines in COVID-19. *Blood Rev*. 2021 Mar;46:100743. doi: 10.1016/j.blre.2020.100743
30. Izcovich A, Ragusa MA, Tortosa F, Lavena Marzio MA, Agnoletti C, Bengolea A, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. *PLoS ONE*. 2020;15(11): e0241955. <https://doi.org/10.1371/journal.pone.0241955>
31. Incerti D, Rizzo S, Li X, Lindsay L, Yau V, Keebler D, et al. Prognostic model to identify and quantify risk factors for mortality among hospitalised patients with COVID-19 in the USA. *BMJ Open*. 2021 Apr 7;11(4):e047121. doi: 10.1136/bmjopen-2020-047121
32. Bellou V, Tzoulaki I, van Smeden M, Moons KGM, Evangelou E, Belbasis L. Prognostic factors for adverse outcomes in patients with COVID-19: a field-wide systematic review and meta-analysis. *Eur Respir J*. 2022 Feb 3;59(2):2002964. doi: 10.1183/13993003.02964-2020

33. Ramlall V, Thangaraj PM, Meydan C, Foox J, Butler D, Kim J, et al. Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. *Nat Med*. 2020 Oct;26(10):1609-1615. doi: 10.1038/s41591-020-1021-2
34. Smati S, Tramunt B, Wargny M, Gourdy P, Hadjadj S, Cariou B. COVID-19 and Diabetes Outcomes: Rationale for and Updates from the CORONADO Study. *Curr Diab Rep*. 2022 Feb;22(2):53-63. doi: 10.1007/s11892-022-01452-5
35. Yang Z, Hu Q, Huang F, Xiong S, Sun Y. The prognostic value of the SOFA score in patients with COVID-19: A retrospective, observational study. *Medicine (Baltimore)*. 2021 Aug 13;100(32):e26900. doi: 10.1097/MD.00000000000026900
36. Puhr-Westerheide D, Reich J, Sabel BO, Kunz WG, Fabritius MP, Reidler P, et al. Sequential Organ Failure Assessment Outperforms Quantitative Chest CT Imaging Parameters for Mortality Prediction in COVID-19 ARDS. *Diagnostics (Basel)*. 2021 Dec 22;12(1):10. doi: 10.3390/diagnostics12010010
37. Chou CY, Chien CH, Han YS, Prebenda MT, Hsieh HP, Turk B, et al. Thiopurine analogues inhibit papain-like protease of severe acute respiratory syndrome coronavirus. *Biochem Pharmacol*. 2008 Apr 15;75(8):1601-9. doi: 10.1016/j.bcp.2008.01.005
38. Mirza MU, Ahmad S, Abdullah I, Froeyen M. Identification of novel human USP2 inhibitor and its putative role in treatment of COVID-19 by inhibiting SARS-CoV-2 papain-like (PLpro) protease. *Comput Biol Chem*. 2020 Dec;89:107376. doi: 10.1016/j.compbiolchem.2020.107376
39. Shin D, Mukherjee R, Grewe D, Bojkova D, Baek K, Bhattacharya A, et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*. 2020 Nov;587(7835):657-662. doi: 10.1038/s41586-020-2601-5
40. Pui CH, Evans WE. Acute lymphoblastic leukemia. *N Engl J Med*. 1998 Aug 27;339(9):605-15. doi: 10.1056/NEJM199808273390907
41. Sorouraddin MH, Khani MY, Amini K, Naseri A, Asgari D, Rashidi MR. Simultaneous Determination of 6-Mercaptopurine and its Oxidative Metabolites in Synthetic Solutions and Human Plasma using Spectrophotometric Multivariate Calibration Methods. *Bioimpacts*. 2011;1(1):53-62. doi: 10.5681/bi.2011.008
42. Zhang YY, Liu WN, You XJ, Gu H, Xu C, Ni X. Prostaglandin E₂ receptors differentially regulate the output of proinflammatory cytokines in myometrial cells from term pregnant women. *Sheng Li Xue Bao*. 2019 Apr 25;71(2):248-260.
43. Kawahara K, Hohjoh H, Inazumi T, Tsuchiya S, Sugimoto Y. Prostaglandin E₂-induced inflammation: Relevance of prostaglandin E receptors. *Biochim Biophys Acta*. 2015 Apr;1851(4):414-21. doi: 10.1016/j.bbalip.2014.07.008
44. Das UN. Essential Fatty Acids and Their Metabolites in the Pathobiology of Inflammation and Its Resolution. *Biomolecules*. 2021 Dec 14;11(12):1873. doi: 10.3390/biom11121873
45. Tsuge K, Inazumi T, Shimamoto A, Sugimoto Y. Molecular mechanisms underlying prostaglandin E₂-exacerbated inflammation and immune diseases. *Int Immunol*. 2019 Aug 23;31(9):597-606. doi: 10.1093/intimm/dxz021
46. Gomez I, Foudi N, Longrois D, Norel X. The role of prostaglandin E₂ in human vascular inflammation. *Prostaglandins Leukot Essent Fatty Acids*. 2013 Aug;89(2-3):55-63. doi: 10.1016/j.plefa.2013.04.004
47. Foo, C.X.; Bartlett, S.; Ronacher, K. Oxysterols in the Immune Response to Bacterial and Viral Infections. *Cells*. 2022; 11:201. <https://doi.org/10.3390/cells11020201>

48. Zu S, Deng YQ, Zhou C, Li J, Li L, Chen Q, et al. 25-Hydroxycholesterol is a potent SARS-CoV-2 inhibitor. *Cell Res.* 2020 Nov;30(11):1043-1045. doi: 10.1038/s41422-020-00398-1
49. Lan Q, Wang C, Zhou J, Wang L, Jiao F, Zhang Y, et al. 25-Hydroxycholesterol-Conjugated EK1 Peptide with Potent and Broad-Spectrum Inhibitory Activity against SARS-CoV-2, Its Variants of Concern, and Other Human Coronaviruses. *Int J Mol Sci.* 2021 Nov 1;22(21):11869. doi: 10.3390/ijms222111869
50. Alba G, Reyes-Quiróz ME, Sáenz J, Geniz I, Jiménez J, Martín-Nieto J, et al. 7-Keto-cholesterol and 25-hydroxy-1 cholesterol rapidly enhance ROS production in human neutrophils. *Eur J Nutr.* 2016 Dec;55(8):2485-2492. doi: 10.1007/s00394-015-1142-4
51. Ramesh C, Anwesh M, Vinithkumar NV, Kirubakaran R, Dufossé L. Complete Genome Analysis of Undecylprodigiosin Pigment Biosynthesizing Marine *Streptomyces* Species Displaying Potential Bioactive Applications. *Microorganisms.* 2021 Oct 28;9(11):2249. doi: 10.3390/microorganisms9112249
52. Santos-Beneit F, Cenicerós A, Nikolaou A, Salas JA, Gutierrez-Merino J. Identification of Antimicrobial Compounds in Two *Streptomyces* sp. Strains Isolated From Beehives. *Front Microbiol.* 2022 Feb 3;13:742168. doi: 10.3389/fmicb.2022.742168
53. Liu P, Wang YY, Qi X, Gu Q, Geng M, Li J. Undecylprodigiosin induced apoptosis in P388 cancer cells is associated with its binding to ribosome. *PLoS One.* 2013 Jun 14;8(6):e65381. doi: 10.1371/journal.pone.0065381. Erratum in: *PLoS One.* 2020 Jul 14;15(7):e0236282.
54. Agrawal AA, Petschenka G, Bingham RA, Weber MG, Rasmann S. Toxic cardenolides: chemical ecology and coevolution of specialized plant–herbivore interactions. *New Phytologist.* 2012;194:28-45. Doi:10.1111/j.1469-8137.2011.04049.x
55. Hetrick B, Yu D, Olanrewaju AA, Chilin LD, He S, Dabbagh D, et al. A traditional medicine, respiratory detox shot (RDS), inhibits the infection of SARS-CoV, SARS-CoV-2, and the influenza A virus in vitro. *Cell Biosci.* 2021 May 29;11(1):100. doi: 10.1186/s13578-021-00609-1
56. Bai X, Liu L, Zhang J, Chen L, Wu T, Aisa HA, et al. Spectrum-effect relationship between GC-QTOF-MS fingerprint and antioxidant, anti-inflammatory activities of *Schizonepeta tenuifolia* essential oil. *Biomed Chromatogr.* 2021 Jul;35(7):e5106. doi: 10.1002/bmc.5106

Supporting information

S1 Table. Clinical Data on Saliva Collection – supplementary data. O₂, oxygen; l/min, liter per minute; PaO₂, arterial oxygen tension; FiO₂, fraction of inspired oxygen; GPT, glutamate-pyruvate transaminase; GOT, glutamic-oxaloacetic transaminase; GGT, gamma-glutamyl transferase; ALP, alkaline phosphatase.

S1 Fig. Individual heatmap of severe COVID-19 versus moderate COVID-19, discharge versus death and no IMV versus IMV.

6 CONCLUSION

This study reports the first application of salivary ATR-FTIR spectroscopy and machine learning with a predictive and prognostic value in COVID-19 patients. In summary, we found that the characteristics of the participants at the time of admission, personal history, symptoms of COVID-19, and computed tomography data were not able to predict severity of disease, hospital outcome or progression to IMV. It was possible to establish the profile of patients and verify the influence of male gender, older age, smoking history, liver and kidney dysfunction, greater need for oxygen supplementation in patients with more severity and worse outcome.

Among the three categories studied, the salivary infrared spectra by SVM algorithm proved to be a possible alternative of prognostic value in COVID-19. Specially in a context of limited resources, knowing that a patient is at risk of becoming severely ill in 68% would already be a reason to refer him to a reference center where he could have adequate follow-up. In this context, the salivary ATR-FTIR spectroscopy coupled with SVM classification could provide a novel alternative for biomedical decision or monitoring (Article 1).

This was also the first study to evaluate the saliva metabolome profile of COVID-19 patients categorizing them into moderate and severe COVID, discharge and death and by the need for IMV. Saliva metabolomic allowed us to identify metabolites related to greater disease severity in the evaluated outcomes which might be used in selection of potential saliva biomarker for severity evaluation early in hospitalization and open new therapeutic perspectives against COVID-19 (Article 2).

More extensive studies are needed to verify these preliminary results. One perspective is to increase the sample size in an attempt to increase the algorithm performance. Another perspective is to perform a control group analyses to allow a better training of algorithms by having more data available for the machine learning model. A control group also is required to better differentiate the metabolites among healthy and COVID-19 patients, favoring the discrimination of metabolic pathways that may be related to disease outcomes.

7 PERSPECTIVES

As a perspective, it is intended to carry out additional analyzes with the database of this thesis and to elaborate articles with the following objectives:

- Analyze the influence of clinical parameters on salivary algorithms developed from the ATR-FTIR spectrum.
- Analyze the ATR-FTIR wavenumbers that differentiate categories moderate vs severe, discharge vs death, and IMV and no IMV.
- Analyze the metabolites present in the exhaled gas of patients with COVID-19 and identify biomarkers that can be used in the diagnosis of the disease by breath.
- Analyze saliva samples that were collected longitudinally at three different moments from each patient, and verify possible associations and differences between days of symptoms, time of collection and outcomes by using ATR-FTIR and metabolomic platforms.

8 PATIENT REPORT

“As a healthcare professional, I had no choice in being sent to work on the front line of COVID-19. It was a necessity of the service as the number of infected people grew, and there was still no vaccine available. It wasn't a personal choice, as I was pre-diabetic and I didn't receive any training to perform my job safely. And despite correctly using personal protective equipment, I ended up contracting COVID-19.

I looked for help at the hospital where I worked for three times, being hospitalized with 75% of pulmonary involvement after 8 days of symptoms. I developed the severe form of the disease and was hospitalized for a long period using a tracheostomy, invasive mechanical ventilation, I had pressure injuries and lost all body movements. I lost my memory of the entire period of hospitalization, dreamed all the time about my children and felt constant anxiety and fear. One day I woke up in the ICU and heard an employee say that I couldn't do it, and that they should let me die. I panicked and was very scared, but I felt the constant presence of someone higher telling me that I wasn't going to die.

I don't have good memories of the period when I was in the ward and I felt negligence of care for myself and my family by the nursing team, and the care was somewhat inhumane. But I'm grateful to everyone who took care of me. It was a relief to be discharged from the hospital, but I became totally dependent on my children for self-care. And it was with a lot of effort, help and prayers that I was able to return to work. The return was traumatic and I was not reassigned from function or work sector, even though I still felt a lot of pain in my legs and numb arms. On my first day back to work my coordinators put me in charge of the care of an intubated patient and I felt sick when I saw the patient. And after the end of medical certificates, they chose to fire me.

Anyway, it's been 2 years since I was released from this nightmare I lived. It was a very difficult time for my family and for me, and I still don't feel fully recovered. I still find myself crying, anxious and afraid of I don't know what, but I have faith that better days will come.”

9 REFERENCES

- ABRÃO, A. L. *et al.* Salivary proteomics: A new adjuvant approach to the early diagnosis of familial juvenile systemic lupus erythematosus. **Medical Hypotheses**, v. 89, p. 97-100, Apr. 2016. <https://doi.org/10.1016/j.mehy.2016.02.010>
- AI, T. *et al.* Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. **Radiology**, vol. 296, n.2, p. E32-E40, Aug. 2020. <https://doi.org/10.1148/radiol.2020200642>
- ANDREW CHAN, K. L.; KAZARIAN, S. G. Attenuated total reflection Fourier-transform infrared (ATRFTIR) imaging of tissues and live cells. **Chemical Society Reviews**, v. 45, n. 7, p. 1850-1864, Apr. 2016. <https://doi.org/10.1039/C5CS00515A>
- ANDREWS, N. *et al.* Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. **The New England Journal of Medicine**, v. 386, n. 16, p. 1532-1546, Apr. 2021. <https://doi.org/10.1056/NEJMoa2119451>
- BAKER, M. J. *et al.* Using Fourier transform IR spectroscopy to analyze biological materials. **Nature Protocols**, v. 9, n. 8, p. 1771-1791, Aug. 2014. <https://doi.org/10.1038/nprot.2014.110>
- BANERJEE, A. *et al.* Rapid Classification of COVID-19 Severity by ATR-FTIR Spectroscopy of Plasma Samples. **Analytical Chemistry**, v. 93, n. 30, p. 10391-10396, Aug. 2021. <https://doi.org/10.1021/acs.analchem.1c00596>
- BARAUNA, V. G. *et al.* Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity. **Analytical Chemistry**, v. 93, n. 5, p. 2950-2958, Feb. 2021. <https://doi.org/10.1021/acs.analchem.0c04608>
- BARBERIS, E. *et al.* Large-Scale Plasma Analysis Revealed New Mechanisms and Molecules Associated with the Host Response to SARS-CoV-2. **International Journal of Molecular Sciences**, v. 21, n. 22, Nov. 2020. <https://doi.org/10.3390/ijms21228623>
- BARBERIS, E. *et al.* Precision Medicine Approaches with Metabolomics and Artificial Intelligence. **International Journal of Molecular Sciences**, v. 23, n. 19, p. 11269, Sep. 2022. <https://doi.org/10.3390/ijms231911269>
- BARTH, A. Infrared spectroscopy of proteins. **Biochim Biophys Acta**, v. 1767, n. 9, p. 1073-1101, Sep. 2007. <https://doi.org/10.1016/j.bbabbio.2007.06.004>
- BATTAGLINI, D. *et al.* Laboratory Biomarkers for Diagnosis and Prognosis in COVID-19. **Frontiers in Immunology**, v. 13, p. 857573, Apr. 2022. <https://doi.org/10.3389/fimmu.2022.857573>

BAUM, B. J.; YATES, J. R. 3rd; SRIVASTAVA, S; WONG, D. T.; MELVIN, J. E. Scientific frontiers: emerging technologies for salivary diagnostics. **Advances in Dental Research**, v. 23, n. 4, p. 360-368, 2011. <https://doi.org/10.1177/0022034511420433>

BEEKES, M.; LASCH, P.; NAUMANN, D. Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research. **Vet Microbiology**, v. 123, n. 4, p. 305-319, Aug. 2007. <https://doi.org/10.1016/j.vetmic.2007.04.010>

BEL'SKAYA, L. V.; SARF, E. A.; MAKAROVA, N. A. Use of fourier transform ir spectroscopy for the study of saliva composition. **Journal of Applied Spectroscopy**, v.85, n.3, Jul. 2018. <https://doi.org/10.1007/s10812-018-0670-0>

BEL'SKAYA, L. V.; SARF, E. A.; SOLOMATIN, D. V. Application of FTIR Spectroscopy for Quantitative Analysis of Blood Serum: A Preliminary Study. **Diagnostics**, v. 11, p. 2391, 2021. <https://doi.org/10.3390/diagnostics11122391>

BI, X. *et al.* Proteomic and metabolomic profiling of urine uncovers immune responses in patients with COVID-19. **Cell Reports**, v. 38, n. 3, p. 110271, Jan. 2022. <https://doi.org/10.1016/j.celrep.2021.110271>

BLASCO, H.; BESSY, C.; PLANTIER, L. *et al.* The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism. **Scientific Reports**. v. 10, n. 1, Oct. 2020. <https://doi.org/10.1038/s41598-020-73966-5>

BORDBAR, M. M. *et al.* A colorimetric electronic tongue for point-of-care detection of COVID-19 using salivary metabolites. **Talanta**, v. 246, p. 123537, Aug. 2022. <https://doi.org/10.1016/j.talanta.2022.123537>

BUNACIU, A. A.; HOANG, V. D.; ABOUL-ENEIN, H. Y. Applications of FT-IR Spectrophotometry in Cancer Diagnostics. **Critical Reviews in Analytical Chemistry**, v. 45, n. 2, p. 156-165, Jan. 2015. <https://doi.org/10.1080/10408347.2014.904733>

BRUZZONE, C. *et al.* SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. **iScience**, v. 23, n. 10, p. 101645, Oct. 2020. <https://doi.org/10.1016/j.isci.2020.101645>

BUTLER, H. J. *et al.* Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer. **Nature Communications**, v. 10, n. 1, Oct. 2019. <https://doi.org/10.1038/s41467-019-12527-5>

CAIXETA, D. C. *et al.* Salivary molecular spectroscopy: A sustainable, rapid and non-invasive monitoring tool for diabetes mellitus during insulin treatment. **PLoS ONE**, v. 15, n. 3, p. e0223461, Mar. 2020. <https://doi.org/10.1371/journal.pone.0223461>

CAIXETA, D. C. *et al.* One-Year Update on Salivary Diagnostic of COVID-19. **Frontiers in Public Health**, v. 9, p. 589564, 2021. <https://doi.org/10.3389/fpubh.2021.589564>

CARPENTER, G. H. The secretion, components, and properties of saliva. **Annual Review of Food Science and Technology**, v. 4, p. 267-276, 2013. <https://doi.org/10.1146/annurev-food-030212-182700>

CASCELLA, M. *et al.* Features, Evaluation, and Treatment of Coronavirus (COVID-19). In: **StatPearls** [Internet]. Treasure Island (FL): StatPearls Publishing, Jan. 2022. Bookshelf ID: NBK554776

CASTAÑÉ, H. *et al.* Machine learning and semi-targeted lipidomics identify distinct serum lipid signatures in hospitalized COVID-19-positive and COVID-19-negative patients. **Metabolism**, v. 131, p. 155197, Jun. 2022. <https://doi.org/10.1016/j.metabol.2022.155197>

CELAYA-PADILLA, J. M. *et al.* Kynurenine and Hemoglobin as Sex-Specific Variables in COVID-19 Patients: A Machine Learning and Genetic Algorithms Approach. **Diagnostics**, v. 11, p. 2197, Nov. 2021. <https://doi.org/10.3390/diagnostics11122197>

CHILAMAKURI, R.; AGARWAL, S. COVID-19: Characteristics and Therapeutics. **Cells**, v. 10, n. 2, p. 206, Jan. 2021. <https://doi.org/10.3390/cells10020206>

CHOI, J. Y.; SMITH, D. M. SARS-CoV-2 Variants of Concern. **Yonsei Medical Journal**, v. 62, n. 11, p. 961-968, Nov. 2021. <https://doi.org/10.3349/ymj.2021.62.11.961>

CLEMENS, G. *et al.* Vibrational spectroscopic methods for cytology and cellular research. **Analyst**, v. 139, n. 18, p. 4411-4444, Sep. 2014. <https://doi.org/10.1039/C4AN00636D>

SANTOS JUNIOR, G. C.; PEREIRA, C. M.; FIDALGO, T. K.; VALENTE, A. P. Saliva NMR-Based Metabolomics in the War Against COVID-19. **Analytical Chemistry**, v. 92, n. 24, p. 15688-15692, 2020. <https://doi.org/10.1021/acs.analchem.0c04679>

CUCINOTTA, D.; VANELLI, M. WHO declares COVID-19 a pandemic. **Acta Biomed**, v. 91, n. 1, p. 157-160, 2020. <https://doi.org/10.23750/abm.v91i1.9397>

D'ALESSANDRO, A. *et al.* Biological and Clinical Factors Contributing to the Metabolic Heterogeneity of Hospitalized Patients with and without COVID-19. **Cells**, v. 10, p. 2293, Sep. 2021. <https://doi.org/10.3390/cells10092293>

DAS, T. *et al.* Analysis of the ex-vivo transformation of semen, saliva and urine as they dry out using ATR-FTIR spectroscopy and chemometric approach. **Scientific Reports**, v. 11, n. 1, p. 11855, Jun. 2021. <https://doi.org/10.1038/s41598-021-91009-5>

DAWES, C.; WONG, D. T. Role of Saliva and Salivary Diagnostics in the Advancement of Oral Health. **Journal of Dental Research**, v. 98, n. 2, p. 133-141, Feb. 2019. <https://doi.org/10.1177/0022034518816961>

DELAFIORI, J. *et al.* Covid-19 Automated Diagnosis and Risk Assessment through Metabolomics and Machine Learning. **Analytical Chemistry**, v. 93, n. 4, p. 2471-2479, Feb. 2021. <https://doi.org/10.1021/acs.analchem.0c04497>

DEPCIUCH, J. *et al.* Phospholipid-protein balance in affective disorders: Analysis of human blood serum using Raman and FTIR spectroscopy. A pilot study. **Journal of Pharmaceutical and Biomedical Analysis**, v. 131, p. 287-296, Nov. 2016. <https://doi.org/10.1016/j.jpba.2016.08.037>

ELLIS, D. I.; GOODACRE, R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. **Analyst**, v. 131, n. 8, p. 875-885, Aug. 2006. <https://doi.org/10.1039/b602376m>

ELKINS, K. M. Rapid presumptive "fingerprinting" of body fluids and materials by ATR FT-IR spectroscopy. **Journal of Forensic Sciences**, v. 56, n. 6, p. 1580-1587, Nov. 2011. <https://doi.org/10.1111/j.1556-4029.2011.01870.x>

ENKE, C. G.; NIEMAN, T. A. Signal-to-Noise Ratio Enhancement by Least-Squares Polynomial Smoothing. **Analytical Chemistry**, v. 48, n. 8, p. 705A-712A, Jul. 1976. <https://doi.org/10.1021/ac50002a769>

Eurosurveillance Editorial Team. Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. **Euro Surveillance**, v. 25, n. 5, p. 200131e, Feb. 2020. <https://doi.org/10.2807/1560-7917.ES.2020.25.5.200131e>

FERNANDES, L. L. Saliva in the Diagnosis of COVID-19: A Review and New Research Directions. **Journal of Dental Research**, v. 99, n. 13, p. 1435-1443, Dec. 2020. <https://doi.org/10.1177/0022034520960070>

FERREIRA, I. C.; AGUIAR, E. M.; SILVA, A. T. *et al.* Attenuated Total Reflection-Fourier Transform Infrared (ATR-FTIR) Spectroscopy Analysis of Saliva for Breast Cancer Diagnosis. **Journal of Oncology**, p. 4343590, Feb. 2020. <https://doi.org/10.1155/2020/4343590>

GALLO MARIN, B. *et al.* Predictors of COVID-19 severity: A literature review. **Reviews in Medical Virology**, v. 31, n. 1, p. 1-10, Jan. 2021. <https://doi.org/10.1002/rmv.2146>

GANGLOFF, C. *et al.* Machine learning is the key to diagnose COVID-19: a proof-of-concept study. **Scientific Reports**, v. 11, n.1, p. 17577, Aug. 2021. <https://doi.org/10.1038/s41598-021-86735-9>

GATES, P. J. *et al.* Mass spectrometry ionization techniques and applications for the analysis of organic compounds. In: CARLTON, A. T.; SILVA, C. H. T. P. (Org.).

Current Methods in Medicinal Chemistry and Biological Physics. Kerala: Research Signpost, 2007. v. 1, p. 187-206.

GAUTRET, P.; MILLION, M.; JARROT, P. A. *et al.* Natural history of COVID-19 and therapeutic options. **Expert Review of Clinical Immunology**, v. 16, n. 12, p. 1159-1184, Dec. 2020. <https://doi.org/10.1080/1744666X.2021.1847640>

GE, X. Y.; LI, J. L.; YANG, X. L. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. **Nature**, v. 503, n. 7477, p. 535-538, 2013. <https://doi.org/10.1038/nature12711>

GORBALENYA, A. E.; BAKER, S. C.; BARIC, R. S. *et al.* The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. **Nature Microbiology**, v. 5, n. 4, p. 536-544, 2020. <https://doi.org/10.1038/s41564-020-0695-z>.

GUPTA, A.; MADHAVAN, M.; SEHGAL, K. *et al.* Extrapulmonary manifestations of COVID-19. **Nature Medicine**, v. 26, p. 1017-1032, 2020. <https://doi.org/10.1038/s41591-020-0968-3>

HAO, M. *et al.* Pathogenic Mechanism and Multi-omics Analysis of Oral Manifestations in COVID-19. **Frontiers in Immunology**, v. 13, p. 879792, Jul. 2022. <https://doi.org/10.3389/fimmu.2022.879792>

HODGSON, S. H.; MANSATTA, K.; MALLETT, G. *et al.* What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. **The Lancet Infectious Diseases**, v. 21, n. 2, p. e26-e35, 2021. [https://doi.org/10.1016/S1473-3099\(20\)30773-8](https://doi.org/10.1016/S1473-3099(20)30773-8)

HOFFMANN, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. **Cell**, v. 181, p. 271-280, Apr. 2020. <https://doi.org/10.1016/j.cell.2020.02.052>

HUANG, C.; WANG, Y.; LI, X. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. **Lancet**, v. 395, n. 10223, p. 497-506, 2020. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)

HUCK, C. W. Recent Developments in Solid-Phase Extraction for Near and Attenuated Total Reflection Infrared Spectroscopic Analysis. **Molecules**, v. 21, n. 5, p. 633, May. 2016. <https://doi.org/10.3390/molecules21050633>

HUI, D. S.; I AZHAR, E.; MADANI, T. A. *et al.* The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. **International Journal of Infectious Diseases**, v. 91, p. 264-266, 2020. <https://doi.org/10.1016/j.ijid.2020.01.009>

HUMPHREY, S. P.; WILLIAMSON, R. T. A review of saliva: normal composition, flow, and function. **Journal of Prosthetic Dentistry**, v. 85, n. 2, p. 162-169, Feb. 2001. <https://doi.org/10.1067/mpr.2001.113778>

IEBBA, V. *et al.* Profiling of Oral Microbiota and Cytokines in COVID-19 Patients. **Frontiers in Microbiology**, v. 12, p. 671813, Jul. 2021.
<https://doi.org/10.3389/fmicb.2021.671813>

KANTARCIOGLU, B. *et al.* An Update on the Pathogenesis of COVID-19 and the Reportedly Rare Thrombotic Events Following Vaccination. **Clinical and Applied Thrombosis Hemostasis**, v. 27, p. 10760296211021498, Jan-Dec. 2021.
<https://doi.org/10.1177/10760296211021498>

KAZARIAN, S. G.; CHAN, K. L. ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems. **Analyst**, v. 138, n. 7, p. 1940-1951, Apr. 2013. <https://doi.org/10.1039/c3an36865c>

KAZMER, S. T. *et al.* Pathophysiological Response to SARS-CoV-2 Infection Detected by Infrared Spectroscopy Enables Rapid and Robust Saliva Screening for COVID-19. **Biomedicines**, v. 10, n. 2, p. 351, Feb. 2022.
<https://doi.org/10.3390/biomedicines10020351>

KHAUSTOVA, S. A. *et al.* Application of mid-infrared molecular spectroscopy for assessment of biochemical parameters of blood serum. **Bulletin of Experimental Biology and Medicine**, v. 148, n. 6, p. 943-947, Dec. 2009.
<https://doi.org/10.1007/s10517-010-0858-7>

KITANE, D. L. *et al.* A simple and fast spectroscopy-based technique for Covid-19 diagnosis. **Scientific Reports**, v. 11, n. 1, p. 16740, Aug. 2021.
<https://doi.org/10.1038/s41598-021-95568-5>

KUMAR, S.; THAMBIRAJA, T. S.; KARUPPANAN, K.; SUBRAMANIAM, G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. **Journal of Medical Virology**, v. 94, n. 4, p. 1641-1649, Apr. 2022.
<https://doi.org/10.1002/jmv.27526>

LEHMANN, W. D. A timeline of stable isotopes and mass spectrometry in the life sciences. **Mass Spectrometry Reviews**, v. 36, n. 1, p. 58-85, Jan. 2017.
<https://doi.org/10.1002/mas.21497>

LI, H. *et al.* SARS-CoV-2 and viral sepsis: observations and hypotheses. **Lancet**, v. 395, p. 1517-1520, 2020. [https://doi.org/10.1016/S0140-6736\(20\)30920-X](https://doi.org/10.1016/S0140-6736(20)30920-X)

LIMA, C. A. *et al.* ATR-FTIR spectroscopy for the assessment of biochemical changes in skin due to cutaneous squamous cell carcinoma. **International Journal of Molecular Sciences**, v. 16, n. 4, p. 6621-6630, Mar. 2015.
<https://doi.org/10.3390/ijms16046621>

LIPMAN, D.; SAFO, S. E.; CHEKOUO, T. Multi-omic analysis reveals enriched pathways associated with COVID-19 and COVID-19 severity. **PLoS ONE**, v. 17, n. 4, p. e0267047, Apr. 2022.
<https://doi.org/10.1371/journal.pone.0267047>

LIU, X.; LIU, C.; LIU, G.; LUO, W.; XIA, N. COVID-19: Progress in diagnostics, therapy and vaccination. **Theranostics**, v. 10, n. 17, p. 7821-7835, Jun. 2020. <https://doi.org/10.7150/thno.47987>

LUKOSE, J.; CHIDANGIL, S.; GEORGE, S. D. Optical technologies for the detection of viruses like COVID-19: Progress and prospects. **Biosensors e Bioelectronics**, v. 178, p. 113004, Apr. 2021. <https://doi.org/10.1016/j.bios.2021.113004>

MAO, B.; LIU, Y.; CHAI, Y. H. *et al.* Assessing risk factors for SARS-CoV-2 infection in patients presenting with symptoms in Shanghai, China: a multicentre, observational cohort study. **Lancet Digit Health**, v. 2, n. 6, p. e323-e330, Jun. 2020. [https://doi.org/10.1016/S2589-7500\(20\)30109-6](https://doi.org/10.1016/S2589-7500(20)30109-6)

MAJUMDER, J.; MINKO, T. Recent developments on therapeutic and diagnostic approaches for COVID-19. **AAPS Journal**, v. 23, n. 14, p. 1-22, 2021. <https://doi.org/10.1208/s12248-020-00532-2>

MALAMUD, D. Saliva as a diagnostic fluid. **Dental Clinics of North America**, v. 55, n. 1, p. 159-178, Jan. 2011. <https://doi.org/10.1016/j.cden.2010.08.004>

MARTINEZ-CUAZITL, A. *et al.* ATR-FTIR spectrum analysis of saliva samples from COVID-19 positive patients. **Scientific Reports**, v. 11, n. 1, p. 19980, Oct. 2021. <https://doi.org/10.1038/s41598-021-99529-w>

MITCHELL, M. **Machine Learning** 1ed: McGraw-Hill, Inc; 1997.

MITTAL, A. *et al.* COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. **PLoS Pathogens**, v. 16, n. 8, p. e1008762, Aug. 2020. <https://doi.org/10.1371/journal.ppat.1008762>

MOUSAVIZADEH, L.; GHASEMI, S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. **Journal of Microbiology, Immunology and Infection**, v. 54, p. 159-163, 2020. <https://doi.org/10.1016/j.jmii.2020.03.022>

MOVASAGHI, Z.; REHMAN, S.; UR REHMAN, D. I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. **Applied Spectroscopy Reviews**, v. 43, n. 2, p. 134-179, 2008. <https://doi.org/10.1080/05704920701829043>

NAGURA-IKEDA, M. *et al.* Clinical Evaluation of Self-Collected Saliva by Quantitative Reverse Transcription-PCR (RT-qPCR), Direct RT-qPCR, Reverse Transcription-Loop-Mediated Isothermal Amplification, and a Rapid Antigen Test To Diagnose COVID-19. **Journal of Clinical Microbiology**, v. 58, n. 9, p. e01438-20, Aug. 2020. <https://doi.org/10.1128/JCM.01438-20>

NASCIMENTO, M. H.; MARCARINI, W. D.; FOLLI, G. S. Noninvasive Diagnostic for COVID-19 from Saliva Biofluid via FTIR Spectroscopy and Multivariate Analysis. **Analytical Chemistry**, v. 94, n. 5, p. 2425-2433, Feb. 2022. <https://doi.org/10.1021/acs.analchem.1c04162>

NOGUEIRA, M. S. *et al.* Rapid diagnosis of COVID-19 using FT-IR ATR spectroscopy and machine learning. **Scientific Reports**, v. 11, n. 1, p. 23153, Nov. 2021. <https://doi.org/10.1038/s41598-021-93511-2>

NOREEN, R. *et al.* FTIR spectro-imaging of collagens for characterization and grading of gliomas. **Biotechnology Advances**, v. 30, n. 6, p. 1432-1446, Nov./Dec. 2012. <https://doi.org/10.1016/j.biotechadv.2012.03.009>

ORPHANOU, C. M. *et al.* The detection and discrimination of human body fluids using ATR FT-IR spectroscopy. **Forensic Science International**, v. 261, p. 82, Apr. 2016. <https://doi.org/10.1016/j.forsciint.2015.04.020>

PFAFFE, T. *et al.* Diagnostic potential of saliva: current state and future applications. **Clinical Chemistry**, v. 57, n. 5, p. 675-87, May. 2011. <https://doi.org/10.1373/clinchem.2010.153767>

POZZI, C. *et al.* A 'Multiomic' Approach of Saliva Metabolomics, Microbiota, and Serum Biomarkers to Assess the Need of Hospitalization in Coronavirus Disease 2019. **Gastro Hep Advances**, v. 1, n.2, p. 194-209, 2022. <https://doi.org/10.1016/j.gastha.2021.12.006>

PUTZIG, C. L. *et al.* Infrared Spectroscopy. **Analytical Chemistry**, v. 66, n. 12, p. 26–66, Jun. 1994. <https://doi.org/10.1021/ac00084a003>

QUARLERI, J.; GALVAN, V.; DELPINO, M. V. Omicron variant of the SARS-CoV-2: a quest to define the consequences of its high mutational load. **GeroScience**, v. 44, n.1, p. 53-56, Feb. 2022. <https://doi.org/10.1007/s11357-021-00500-4>

RICHARD, V. R. *et al.* Early Prediction of COVID-19 Patient Survival by Targeted Plasma Multi-Omics and Machine Learning. **Molecular e Cellular Proteomics**, v. 21, n. 10, p. 100277, Oct. 2022. <https://doi.org/10.1016/j.mcpro.2022.100277>

RICHARDSON, S. *et al.* Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. **JAMA**, v. 323, n. 20, p. 2098, May. 2020. <https://doi.org/10.1001/jama.2020.6775>

RODRIGUES, R. P. *et al.* Differential Molecular Signature of Human Saliva Using ATR-FTIR Spectroscopy for Chronic Kidney Disease Diagnosis. **Brazilian Dental Journal**, v. 30, n. 5, p. 437-445, Oct. 2019. <https://doi.org/10.1590/0103-6440201902228>

SABINO-SILVA, R.; JARDIM, A. C.; SIQUEIRA, W. L. Coronavirus COVID-19 impacts to dentistry and potential salivary diagnosis. **Clinical Oral Investigations**, v. 24, n. 4, p. 1619-1621, Apr. 2020. <https://doi.org/10.1007/s00784-020-03248-x>

SHEN, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. **Cell**, v. 182, n. 1, p. 59-72.e15, Jul. 2020. <https://doi.org/10.1016/j.cell.2020.05.032>

SAMUEL, A. L. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 210-229, 1959.
<https://doi.org/10.1147/rd.33.0210>

SAPKOTA, D. *et al.* COVID-19 salivary signature: diagnostic and research opportunities. **Journal of Clinical Pathology**, p. 206834, Aug. 2020.
<https://doi.org/10.1136/jclinpath-2020-206834>

SCHLEDER, G. R.; FAZZIO, A. Machine Learning na Física, Química, e Ciência de Materiais: Descoberta e Design de Materiais. **Revista Brasileira de Ensino de Física**, v. 43, Suppl 1, 2021. <https://doi.org/10.1590/1806-9126-RBEF-2020-0407>

SCHNEIDER, B. B. *et al.* Differential mobility spectrometry/mass spectrometry history, theory, design optimization, simulations, and applications. **Mass Spectrometry Reviews**, v. 35, n. 6, p. 687-737, Oct. 2016.
<https://doi.org/10.1002/mas.21453>

SENEVIRATHNA, K. *et al.* Diagnostic Value of FTIR Spectroscopy, Metabolomic Screening and Molecular Genetics in Saliva for Early Detection of Oral Squamous Cell Carcinoma (OSCC). **Medical & Clinical Research**, v. 6, n. 2, p. 435-454, Mar. 2021. <https://doi.org/10.33140/MCR.06.03.05>

SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J. **Spectrometric identification of Organic Compounds**. 7. ed., John Wiley & Sons Ed., Danvers, 13 p., 2005.

SONG, F. *et al.* Emerging 2019 novel coronavirus (2019-nCoV) pneumonia. **Radiology**, v. 295, n. 1, p. 210–217, Apr. 2020.
<https://doi.org/10.1148/radiol.2020200274>

SONG, J. W. *et al.* Omics-Driven Systems Interrogation of Metabolic Dysregulation in COVID-19 Pathogenesis. **Cell Metabolism**, v. 32, n. 2, p. 188-202.e5, Aug. 2020.
<https://doi.org/10.1016/j.cmet.2020.06.016>

SONG, Z. *et al.* Integrated histopathological, lipidomic, and metabolomic profiles reveal mink is a useful animal model to mimic the pathogenicity of severe COVID-19 patients. **Signal Transduction and Targeted Therapy**, v. 7, n.1, p. 29, Jan. 2022.
<https://doi.org/10.1038/s41392-022-00891-6>

SPICK, M. *et al.* An integrated analysis and comparison of serum, saliva and sebum for COVID-19 metabolomics. **Scientific Reports**, v. 12, n. 1, p. 11867, Jul. 2022.
<https://doi.org/10.1038/s41598-022-16123-4>

STEFAN, N. *et al.* Obesity and impaired metabolic health in patients with COVID-19. **Nature Reviews Endocrinology**, v. 16, n.7, p. 341-342, Jul. 2020.
<https://doi.org/10.1038/s41574-020-0364-6>

STUART, B. H. Biological Applications. In: (Ed.). **Infrared Spectroscopy: Fundamentals and Applications**: John Wiley & Sons, p. 01-93, Jul. 2004 ISBN 9780470011140. <https://doi.org/10.1002/0470011149>

STUART, B. H. Biological Applications. In: (Ed.). **Infrared Spectroscopy: Fundamentals and Applications**: John Wiley & Sons, p.137-165, Jul. 2004. ISBN 9780470011140. <https://doi.org/10.1002/0470011149.ch7>

SU, S. *et al.* Epidemiology, genetic recombination, and pathogenesis of coronaviruses. **Trends in Microbiology**, v. 24, n. 6, p. 490-502, Jun. 2016. <https://doi.org/10.1016/j.tim.2016.03.003>

SU, Y. *et al.* Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. **Cells**, v. 183, n. 6, p. 1479-1495.e20, Dec. 2020. <https://doi.org/10.1016/j.cell.2020.10.037>

THISSEN, U. *et al.* Comparing support vector machines to PLS for spectral regression applications. **Chemometrics and Intelligent Laboratory Systems**, v. 73, n. 2, p. 169–179, Oct. 2004. <https://doi.org/10.1016/j.chemolab.2004.01.002>

THOMAS, T. *et al.* COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. **Journal of Clinical Investigation Insight**, v. 5, n. 14, p. e140327, Jul. 2020. <https://doi.org/10.1172/jci.insight.140327>

TOPOL, E. J. High-performance medicine: the convergence of human and artificial intelligence. **Nature Medicine**, v. 25, n.1, p. 44-56, 2019. <https://doi.org/10.1038/s41591-018-0300-7>

UDDIN, S.; KHAN, A.; HOSSAIN, M. E.; MONI, M. A. Comparing different supervised machine learning algorithms for disease prediction. **BMC Medical Informatics and Decision Making**, v. 19, n. 1, p. 281, 2019. <https://doi.org/10.1186/s12911-019-1004-8>

UDUGAMA, B. *et al.* Diagnosing COVID-19: the disease and tools for detection. **American Chemical Society nano**, v. 14, n. 4, p. 3822–35, Apr. 2020. <https://doi.org/10.1021/acsnano.0c02624>

VANGUILDER, H. D.; VRANA, K. E.; FREEMAN, W. M. Twenty-five years of quantitative PCR for gene expression analysis. **Biotechniques**, v. 44, n. 5, p. 619-26, Apr. 2008. <https://doi.org/10.2144/000112776>

WANG, A. *et al.* Machine Learning Prediction of COVID-19 Severity Levels From Salivaomics Data. **ArXiv** [Preprint]. Jul. 2022.

WANG, X. *et al.* Multiomics: unraveling the panoramic landscapes of SARS-CoV-2 infection. **Cellular and molecular immunology**, v. 18, n. 10, p. 2313-2324, Oct. 2021. <https://doi.org/10.1038/s41423-021-00754-0>

WHOa. **Novel coronavirus (2019-nCoV): situation report**, 1. Geneva: World Health Organization; 2020. In: <<https://apps.who.int/iris/handle/10665/330760>>. Accessed in: 2021-09-08.

World Health Organization. **WHO Coronavirus (COVID-19) Dashboard**; 2022. In: <<https://covid19.who.int/>>. Accessed in: 2022-11-16.

WINICHAKOON, P. *et al.* Negative Nasopharyngeal and Oropharyngeal Swab Does Not Rule Out COVID-19. **Journal of Clinical Microbiology**, v. 58, p. e00297-20, Apr. 2020. <https://doi.org/10.1128/JCM.00297-20>

WOOD, B. R. *et al.* Infrared Based Saliva Screening Test for COVID-19. **Angewandte Chemie International Edition**, v. 60, n. 31, p. 17102-17107, Jul. 2021. <https://doi.org/10.1002/anie.202104453>

WU, D. *et al.* Plasma metabolomic and lipidomic alterations associated with COVID-19. **National Science Review**, v. 7, n. 7, p. 1157-1168, Jul. 2020. <https://doi.org/10.1093/nsr/nwaa086>

YANG, Y. *et al.* Laboratory Diagnosis and Monitoring the Viral Shedding of 2019-nCoV Infections. **medRxiv**, Feb. 2020. <https://doi.org/10.1101/2020.02.11.20021493>

ZENG, B. *et al.* Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern: a systematic review and meta-analysis. **BMC Medicine**, v. 20, n. 1, p. 200, May. 2022. <https://doi.org/10.1186/s12916-022-02397-y>

ZHANG, C. Z. *et al.* Saliva in the diagnosis of diseases. **International Journal of Oral Science**, v. 8, n. 3, p. 133-137, Sep. 2016. <https://doi.org/10.1038/ijos.2016.38>

ZHANG, L. *et al.* Fast Screening and Primary Diagnosis of COVID-19 by ATR-FT-IR. **Analytical Chemistry**, v. 93, n. 4, p. 2191-2199, Feb. 2021. <https://doi.org/10.1021/acs.analchem.0c04049>

ATTACHMENTS

ANNEX A- Proof of submission of the research project and approval by the Research Ethics Committee

FACULDADE UNA DE UBERLÂNDIA	
--	---

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Plataformas tecnológicas para o diagnóstico e controle do COVID-19.
Pesquisador: Luiz Ricardo Goulart Filho
Área Temática:
Versão: 2
CAAE: 30848620.1.0000.5704
Instituição Proponente: IMUNOSCAN ENGENHARIA MOLECULAR LTDA
Patrocinador Principal: IMUNOSCAN ENGENHARIA MOLECULAR LTDA

DADOS DO PARECER

Número do Parecer: 3.981.361

Apresentação do Projeto:
Trata-se de resposta de pendências sobre o protocolo "Plataformas tecnológicas para o diagnóstico e controle do COVID 19.

Objetivo da Pesquisa:
resposta de pendências ao CEP UNA

Avaliação dos Riscos e Benefícios:
não aplicáveis

Comentários e Considerações sobre a Pesquisa:
as pendências foram acatadas e o CEP UNA manifesta-se pela aprovação do presente estudo

Considerações sobre os Termos de apresentação obrigatória:
adequados

Recomendações:
nenhuma

Conclusões ou Pendências e Lista de Inadequações:
nenhuma

Considerações Finais a critério do CEP:
OBS.: O CEP/UNA LEMBRA QUE QUALQUER MUDANÇA NO PROTOCOLO DEVE SER INFORMADA IMEDIATAMENTE AO CEP PARA FINS DE ANÁLISE E APROVAÇÃO.

Endereço: Alameda Paulina Margonari, 59 Bairro: JARDIM KARAIBA UF: MG Município: UBERLÂNDIA Telefone: (34)3233-1505	CEP: 28.411-206 E-mail: cep@una.br
---	---

Continuação do Parecer: 3.981.361

O CEP/UNA lembra que:

a- segundo a Resolução 466/12, o pesquisador deverá arquivar por 5 anos o relatório da pesquisa e os Termos de Consentimento Livre e Esclarecido, assinados pelo sujeito de pesquisa.

b- poderá, por escolha aleatória, visitar o pesquisador para conferência do relatório e documentação pertinente ao projeto.

c- a aprovação do protocolo de pesquisa pelo CEP/UNA dá-se em decorrência do atendimento a Resolução CNS 466/12, não implicando na qualidade científica do mesmo.

Orientações ao pesquisador :

- O participante da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (Res. CNS 466/12) e deve receber uma via original do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado.
- O pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado e descontinuar o estudo somente após análise das razões da descontinuidade pelo CEP que o aprovou (Res. CNS 466/12), aguardando seu parecer, exceto quando perceber risco ou dano não previsto ao participante ou quando constatar a superioridade de regime oferecido a um dos grupos da pesquisa que requeiram ação imediata.
- O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo (Res. CNS 466/12 e 510/16). É papel de o pesquisador assegurar medidas imediatas adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP
- Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
----------------	---------	----------	-------	----------

Endereço: Alameda Paulina Margonari, 59

Bairro: JARDIM KARAIBA

CEP: 28.411-206

UF: MG

Município: UBERLÂNDIA

Telefone: (34)3233-1505

E-mail: cep@una.br

FACULDADE UNA DE
UBERLÂNDIA



Continuação do Parecer: 3.981.361

Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1535524.pdf	20/04/2020 12:32:45		Aceito
Recurso Anexado pelo Pesquisador	Resposta_CEP_UNA_20_04_2020.pdf	20/04/2020 12:32:19	Luiz Ricardo Goulart Filho	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_COVID_19.docx	20/04/2020 09:48:42	Luiz Ricardo Goulart Filho	Aceito
Declaração de Pesquisadores	Declaracoes_CEP_2020.pdf	19/04/2020 15:07:53	Luiz Ricardo Goulart Filho	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TERMO_DE_CONSENTIMENTO_LIVRE_E_ESCLARECIDO_COVID19.pdf	16/04/2020 21:03:35	Luiz Ricardo Goulart Filho	Aceito
Folha de Rosto	FolhaDeRosto_Imunoscan_2020.pdf	16/04/2020 20:32:12	Luiz Ricardo Goulart Filho	Aceito
Projeto Detalhado / Brochura Investigador	Proposta_PPP_sensores_nCov_2020.pdf	15/04/2020 23:11:30	Luiz Ricardo Goulart Filho	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

UBERLÂNDIA, 20 de Abril de 2020

Assinado por:
Denis Fabiano de Souza
(Coordenador(a))

Endereço: Alameda Paulina Margonari, 59

Bairro: JARDIM KARAIBA

UF: MG

Município: UBERLÂNDIA

CEP: 28.411-206

Telefone: (34)3233-1505

E-mail: cep@una.br

APPENDIX

APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado(a) Senhor(a),

Após concordar em conversar conosco, a respeito de nossa pesquisa, pelos primeiros profissionais que lhe cuidam diretamente, estamos convidando-lhe para participar da pesquisa intitulada “PLATAFORMAS TECNOLÓGICAS PARA O DIAGNÓSTICO E O CONTROLE DO COVID-19”, sob a responsabilidade do pesquisador Dr. Thúlio Marquez Cunha. Nessa pesquisa, nós estamos buscando saliva, sangue e ar expirado de pacientes com suspeita de diagnóstico de COVID.

O Termo/Registro de Consentimento Livre e Esclarecido está sendo obtido pelo pesquisador Dr. Thúlio Marquez Cunha e Débora Lara Zuza Scheucher no Hospital de Clínicas de Uberlândia e no Uberlândia Medical Center, no momento de sua internação ou consulta ambulatorial, e você tem todo o tempo que precisar para pensar se pode, ou não nos ajudar em nossa pesquisa. Nenhum dado ou coleta de qualquer material acontecerá previamente à concordância e assinatura deste documento. Todos os profissionais que o avaliarem são treinados e habilitados para colherem suas amostras (conforme item IV da Resolução nº 466/2012 ou Capítulo. III da Resolução nº 510/2016).

Além de colher uma amostra de aproximadamente 10 ml de sangue (tubo tampa roxa EDTA e tubo tampa amarela). Será necessário realizar a coleta de amostras de orofaringe e mucosa nasal com swabs (cotonetes), imediatamente colocados em tubos estéreis em solução adequada, de saliva em tubos secos você irá permanecer sem comer, beber água ou escovar os dentes por 1 hora e no momento da coleta irá juntar uma boa quantidade de saliva na boca e cuspir 1 ml no frasco. E para a coleta de ar você terá que fazer uma inspiração profunda e assoprar por 10 segundos em bocal conectado a um frasco de vidro que irá armazenar o ar que sai do seu pulmão.

Os pacientes com suspeita de infecção viral e pacientes com outras doenças, que servirão como controles, serão devidamente caracterizados por médicos especialistas, sendo submetidos às coletas das três amostras, como parte dos procedimentos de rotina laboratorial.

O(a) paciente poderá fazer parte de um dos dois grupos envolvidos no estudo: o grupo dos indivíduos considerados positivos para o COVID-19 com e sem manifestação de sintomas, e pelo o grupo de indivíduos considerados negativos para o COVID-19 (com ou sem outras doenças).

As amostras serão processadas no laboratório e, em seguida serão realizados testes moleculares, sorológicos e então serão utilizadas para validação em três plataformas diagnósticas em fase de testes. Todo material utilizado será estéril e descartável, sem qualquer desconforto ou riscos adicionais, além dos riscos mínimos, hematomas que podem ocorrer na retirada do sangue.

No laboratório suas amostras serão armazenadas e submetidas a exames para que possamos fabricar testes que detectam o adoecimento por COVID facilitando o diagnóstico e interrompendo a transmissão para outras pessoas.

Os dados clínicos de seu prontuário, as anotações a respeito de como está acontecendo seu tratamento, seus exames e se você se curou, ou não, também serão utilizados, mas em nenhum momento relataremos as informações obtidas de maneira que você possa ser identificado. Suas amostras armazenadas serão identificadas por códigos numéricos.

Ao longo da pesquisa, podemos também realizar um contato telefônico breve, de no máximo 15 minutos, para perguntar alguma informação que tenha ficado faltando no seu prontuário.

Todas as nossas anotações e registros terão de ser guardados em local seguro por 5 anos de acordo com Resolução nº 510/16, Capítulo VI, Art. 28: IV. Reafirmamos que em nenhum momento você será identificado. Os resultados da pesquisa serão publicados e ainda assim a sua identidade será preservada.

Você terá como ver os resultados da pesquisa, em linguagem que conseguirá entender, para que possa ver que cumprimos com o sigilo que lhe prometemos, e que sua contribuição foi importante para melhorar e salvar outras vidas.

Você não terá nenhum gasto e nenhum ganho financeiro por participar na pesquisa. Se tiver que se deslocar unicamente para oferecer informações e colher exames para realização da pesquisa, será ressarcido pelo custo extra que teve com o seu transporte.

Havendo algum dano decorrente da pesquisa, você terá direito a solicitar indenização através das vias judiciais (Código Civil, Lei 10.406/2002, Artigos 927 a 954 e Resolução CNS nº 510 de 2016, Artigo 19).

Os riscos que podem ocorrer é o de ser contaminado por algum patógeno ao permanecer em hospital/ambulatorio por um tempo um pouco maior do que somente estivesse consultando. Há também o risco de erro na hora de “pegar a veia” para retirada de um tubo de sangue. Para minimizar estes riscos, treinaremos toda a equipe e seguiremos todos os

protocolos de biossegurança preconizados pela ANVISA, desinfetaremos as mãos, os móveis que se sentará, usaremos máscara e tomaremos o menor tempo possível, para sua avaliação.

O benefício em participar deste estudo consiste na contribuição para o desenvolvimento científico de uma nova plataforma diagnóstica visando a detecção rápida, sensível e precoce do COVID-19 com o objetivo de aprimorar o Sistema Único de Saúde brasileiro.

Você é livre para deixar de participar da pesquisa a qualquer momento, sem qualquer prejuízo ou coação até o momento da divulgação dos resultados. Você também é livre para solicitar a retirada dos seus dados da pesquisa e descarte de suas amostras.

Uma via original deste Termo de Consentimento Livre e Esclarecido ficará com você, assinada e rubricada pelos pesquisadores.

Em caso de qualquer dúvida ou reclamação a respeito da pesquisa, você poderá entrar em contato com Dr. Thúlio Marquez Cunha e a pesquisadora Débora Lara Zuza Scheucher, através do telefone (34) 3218-2728, (Laboratório de função pulmonar), localizado na Av. Pará 1720, CEP 38405-320- Umuarama, Uberlândia - MG.

Para obter orientações quanto aos direitos dos participantes de pesquisa acesse a cartilha no link:

https://conselho.saude.gov.br/images/comissoes/conep/documentos/Cartilha_Direitos_Eticos_2020.pdf.

Você poderá também entrar em contato com o Comitê de Ética na Pesquisa com Seres Humanos – CEP, da Universidade Federal de Uberlândia, localizado na Av. João Naves de Ávila, nº 2121, bloco A, sala 224, campus Santa Mônica – Uberlândia/MG, 38408-100; pelo telefone (34) 3239-4131 ou pelo e-mail cep@propp.ufu.br. O CEP/UFU é um colegiado independente criado para defender os interesses dos participantes das pesquisas em sua integridade e dignidade e para contribuir para o desenvolvimento da pesquisa dentro de padrões éticos conforme resoluções do Conselho Nacional de Saúde.

_____, ____ de _____ de 20____

Assinatura do(s) pesquisador(es)

Eu aceito participar do projeto citado acima, voluntariamente, após ter sido devidamente esclarecido.

Assinatura do participante da pesquisa