Internal and Emergency Medicine (2024) 19:1439–1458
https://doi.org/10.1007/s11739-024-03547-1

EM - ORIGINAL

Novel COVID-19 biomarkers identified through multi-omics data analysis: *N*-acetyl-4-*O*-acetylneuraminic acid, *N*-acetyl-L-alanine, *N*-acetyltriptophan, palmitoylcarnitine, and glycerol 1-myristate

Alexandre de Fátima Cobre[1] · Alexessander Couto Alves[2] · Ana Raquel Manuel Gotine[3] · Karime Zeraik Abdalla Domingues[1] · Raul Edison Luna Lazo[1] · Luana Mota Ferreira[4] · Fernanda Stumpf Tonin[5] · Roberto Pontarolo[4]

Received: 1 November 2023 / Accepted: 16 January 2024 / Published online: 28 February 2024

Abstract

This study aims to apply machine learning models to identify new biomarkers associated with the early diagnosis and prognosis of SARS-CoV-2 infection.Plasma and serum samples from COVID-19 patients (mild, moderate, and severe), patients with other pneumonia (but with negative COVID-19 RT-PCR), and healthy volunteers (control) from hospitals in four different countries (China, Spain, France, and Italy) were analyzed by GC–MS, LC–MS, and NMR. Machine learning models (PCA and PLS-DA) were developed to predict the diagnosis and prognosis of COVID-19 and identify biomarkers associated with these outcomes.A total of 1410 patient samples were analyzed. The PLS-DA model presented a diagnostic and prognostic accuracy of around 95% of all analyzed data. A total of 23 biomarkers (e.g., spermidine, taurine, L-aspartic, L-glutamic, L-phenylalanine and xanthine, ornithine, and ribothimidine) have been identified as being associated with the diagnosis and prognosis of COVID-19. Additionally, we also identified for the first time five new biomarkers (N-Acetyl-4-O-acetylneuraminic acid, N-Acetyl-L-Alanine, N-Acetyltriptophan, palmitoylcarnitine, and glycerol 1-myristate) that are also associated with the severity and diagnosis of COVID-19. These five new biomarkers were elevated in severe COVID-19 patients compared to patients with mild disease or healthy volunteers.The PLS-DA model was able to predict the diagnosis and prognosis of COVID-19 around 95%. Additionally, our investigation pinpointed five novel potential biomarkers linked to the diagnosis and prognosis of COVID-19: N-Acetyl-4-O-acetylneuraminic acid, N-Acetyl-L-Alanine, N-Acetyltriptophan, palmitoylcarnitine, and glycerol 1-myristate. These biomarkers exhibited heightened levels in severe COVID-19 patients compared to those with mild COVID-19 or healthy volunteers.

Keywords COVID-19 · Diagnosis · Prognosis · Biomarker · Machine learning

✉ Roberto Pontarolo
pontarolo@ufpr.br

Alexandre de Fátima Cobre
alexandrecobre@gmail.com

Alexessander Couto Alves
a.coutoalves@surrey.ac.uk

Ana Raquel Manuel Gotine
anaraquelmanuel@gmail.com

Karime Zeraik Abdalla Domingues
karimezeraik@gmail.com

Raul Edison Luna Lazo
raulluna@ufpr.br

Luana Mota Ferreira
luanamota@ufpr.br

Fernanda Stumpf Tonin
stumpf.tonin@ufpr.br

1 Universidade Federal do Paraná, Curitiba, Brazil

2 School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK

3 Public Health College, Universidade de São Paulo, São Paulo, Brazil

4 Department of Pharmacy, Universidade Federal do Paraná, Campus III, Av. Pref. Lothário Meissner, 632, Jardim Botânico, Curitiba, PR 80210-170, Brazil

5 H&TRC - Health & Technology Research Centre, ESTeSL, Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Lisbon, Portugal

Abbreviations

ANN	Artificial neural network
EPO	External parameter orthogonalization
GC–MS	Gas chromatography coupled to mass spectrometry
GLSW	Generalized least squares weighting
LC–MS	Liquid chromatography coupled with mass spectrometry
NMR	Nuclear magnetic resonance
OSC	Orthogonal signal correction
PLS-DA	Discriminant analysis by partial least squares
PCA	Principal component analysis
RF	Random forest
RMSEC	Root mean square error of calibration
RMSECV	Root mean square cross-validation error
RT-PCR	Reverse transcription polymerase chain reaction
SVM	Support vector machine
VIP	Variable importance in projection

Introduction

The global pandemic caused by the novel coronavirus, SARS-CoV-2, has spurred a profound urgency for innovative diagnostic and prognostic approaches to combat the spread and impact of COVID-19 [1, 2]. Harnessing the power of cutting-edge technologies like Liquid Chromatography–Mass Spectrometry (LC–MS) [3], Gas Chromatography–Mass Spectrometry (GC–MS) [4], and Nuclear Magnetic Resonance (NMR) [5], this study delves into a comprehensive analysis of omics data obtained from COVID-19 patients across diverse geographical regions, including Italy, China, Spain, and France.

The Partial Least Squares Discriminant Analysis (PLS-DA) stands as the leading machine learning (ML) algorithm for analyzing data across omics sciences [6–8]. Expert-endorsed and cited extensively, it excels due to its superior predictive performance and capacity to handle multidimensional data efficiently [6–8]. Its ability to predict using smaller dimensions, termed latent variables, facilitates analysis without pre-filtering, resolving multicollinearity issues, and identifying crucial biomarkers. Compared to other models like deep learning or artificial neural networks, PLS-DA requires fewer training samples, making it more practical for omics data, where variables often outnumber samples [9].

Integrating machine learning techniques to this multi-omics dataset not only elucidates distinctive molecular signatures but also unveils novel biomarkers crucial for the early diagnosis and prognosis of COVID-19 infection [2]. This interdisciplinary investigation aims to deepen our understanding of the disease's intricacies and offer invaluable insights into potential markers for swift and accurate identification of COVID-19 cases, thereby aiding in effective clinical management and timely interventions [10]. In this paper, we systematically explore multi-omics data via machine learning algorithms (PCA and PLS-DA) derived from varied populations, transcending geographical boundaries, to unveil pivotal biomolecular indicators that could revolutionize the landscape of COVID-19 diagnosis and prognosis.

Material and methods

The data used to carry out this study were collected from the Metaboanalyst platform database [11], a public database containing experimental data from various diseases, including COVID-19.

The following inclusion criteria were adopted for the selection of datasets: (i) samples from COVID-19 patients should be analyzed by nuclear magnetic resonance (NMR), liquid chromatography coupled with mass spectrometry (LC–MS), or gas chromatography coupled to mass spectrometry (GC–MS); (ii) datasets should present the identified metabolites (characterized); (iii) patients should be appropriately categorized according to the result of the COVID-19 test (positive or negative) or according to the severity of the disease (mild, moderate, and severe).

Thus, we just collected and evaluated seven different public metabolomics and lipidomics data sets from a cohort of COVID-19 patients (diagnosed by RT-PCR) from four different countries: France, Spain, Italy, and China. These data (metabolomics and lipidomics) were generated by analyzing the plasma of patients at different stages of the disease using three different analytical techniques: NMR, GC–MS, and LC–MS. The quick access links to the databases used in this study can be accessed below. The summary of all omics data used for the development of this study is presented in Table 1.

Clinical conditions evaluated in the study

This study analyzed COVID-19, non-COVID-19, and healthy patient samples. The non-COVID-19 group were patients who had negative RT-PCR test results for COVID-19 but who suffered from other pneumonias. Table 1 shows the clinical conditions evaluated in each study database.

Study design

Figure 1 shows a flowchart used to investigate biomarkers associated with the diagnosis and prognosis of COVID-19, including the investigation of these biomarkers in the pathophysiology of the disease. It is important to highlight that this flowchart was inspired by previous studies published in

Table 1 Multi-omics datasets and their respective countries

Study	Dataset	Country	Bioanalytics method	Sample type	Sample collection period	Condition	Total N
Bruzzone [14]	1	Spain	NMR	Plasma	2020	Healthy	280
			NMR	Plasma	2020	COVID-19	261
Shi [15]	2	China	GC–MS	Plasma	2020	Healthy	57
			GC–MS	Plasma	2020	COVID-19	60
			GC–MS	Plasma	2020	Non-COVID-19	30
Albóniga [16]	3	Spain	LC–MS	Plasma	Not reported	Healthy	133
			LC–MS	Plasma	Not reported	COVID-19	254
Barberis [17]	4	Italy	GC–MS	Plasma	March–April, 2020	Healthy	26
			GC–MS	Plasma	March–April, 2020	Non-COVID-19	32
			GC–MS	Plasma	March–April, 2020	COVID-19	103
Blasco [18]	5	France	LC–MS	Plasma	April, 2020	Healthy	45
			LC–MS	Plasma	April, 2020	COVID-19	55
Caterino [19]	6	Italy	LC–MS	Plasma	Not reported	Healthy	9
			LC–MS	Plasma	Not reported	Mild	20
			LC–MS	Plasma	Not reported	Moderate	16
			LC–MS	Plasma	Not reported	Severe	16
Barberis [20]	7	Italy	GC–MS	Serum	March, 2020	Healthy	24
			GC–MS	Serum	March, 2020	COVID-19	21

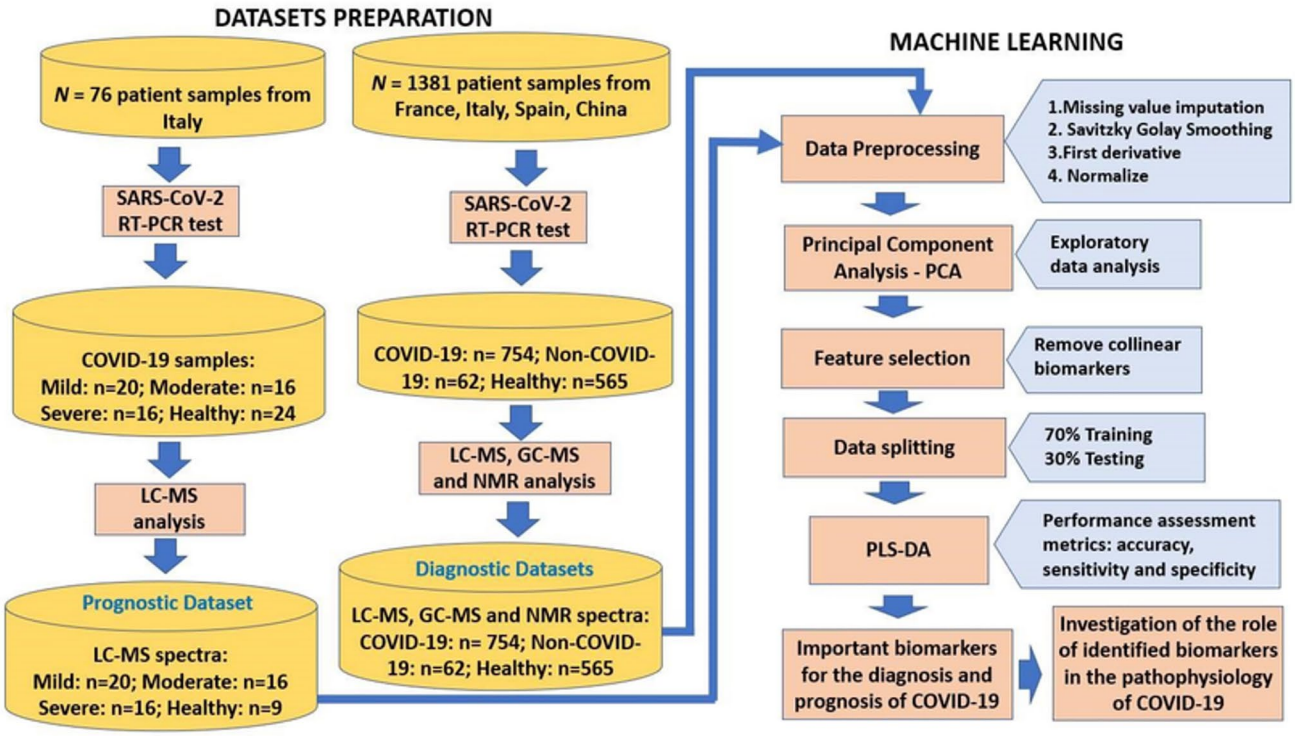


Fig. 1 Flowchart adapted from [12, 13] demonstrating the steps followed in the study. To investigate diagnostic biomarkers of COVID-19, we collected omics datasets ($n=1381$) in six distinct datasets of COVID-19, non-COVID-19, and healthy patients from Italy, France, Spain, and China. To investigate prognostic biomarkers for COVID-19, we collected another dataset ($n=76$) that contained COVID-19 patients of different severities (mild, moderate, and severe) and healthy groups. The seven datasets were separately and independently pre-processed and analyzed via machine learning models (PCA and PLS-DA) to identify diagnostic and prognostic biomarkers. Finally, the role of important biomarkers identified in the pathophysiology of COVID-19 was investigated

the literature [12, 13]. This flowchart is summarized in the following three steps:

First step Initially, blood samples (plasma and serum) from patients in the COVID-19, non-COVID-19 groups, and the control group (healthy volunteers) were analyzed by LC–MS, GC–MS, and NMR techniques. These three analytical techniques allow the detection of thousands of metabolites in plasma.

Second step We performed a multivariate analysis using two machine learning models (PCA and PLS-DA) to compare all metabolites between the COVID-19 and control groups. From this comparison, we identified the most relevant biomarkers for the diagnosis and prognosis of COVID-19.

Third step After identifying the main biomarkers for the diagnosis and prognosis of COVID-19 through machine learning, we compared the role of these biomarkers with the pathophysiological mechanisms associated with COVID-19 based on publications available in the literature.

Description of the clinical conditions of the seven different datasets

Dataset I (Spain dataset): NMR analysis

Dataset I (dataset from Spain) refer to plasma samples from COVID-19 patients ($n=261$) diagnosed with RT-PCR and from healthy volunteers ($n=280$) with negative COVID-19 RT-PCR test results [14]. Samples from healthy volunteers were collected during routine examinations (check-ups) before the pandemic (2018–2019). The two groups of samples were analyzed by NMR. The gender distribution in both patient groups was statistically the same (116 women in the COVID-19 group *vs* 146 women in the healthy group, $p=1.00$); on the other hand, the mean age in both patient groups was statistically different (COVID-19 = 65 years *vs* healthy = 45 years, $p<0.001$). The average hospitalization time of COVID-19 patients was 14 days. Nearly 9.34% ($n=24$) of the patients died, and the most frequent comorbidities were hypertension ($n=116$, 45.14%), cardiovascular events ($n=68$, 26.46%), and diabetes ($n=64$, 24.90%). The most reported symptoms were fever ($n=177$, 68.87%), fatigue ($n=148$, 57.59%), and dry cough ($n=135$, 52.53%) [14]. Table 1 shows the division of *dataset I* into two subsets (training and test) for machine learning analysis.

Dataset II (China dataset): GC–MS analysis

Dataset II (dataset from China) consisted of plasma samples from three groups of patients: COVID-19 patients diagnosed by RT-PCR test ($n=57$), a group of healthy volunteers with a negative result of the COVID-19 RT-PCR test ($n=60$), and non-COVID-19 patients who had COVID-19-like symptoms but who had a negative result of the COVID-19 RT-PCR test [15]. The following inclusion criteria were adopted to define a non-COVID-19 patient: (i) reduced lymphocyte or white blood cell count at the onset of the infection, (ii) fever or respiratory symptoms, and (iii) imaging manifestations of pneumonia. Median age was similar among COVID-19, non-COVID-19, and healthy patients, at 51 years (IQR, 38–59), 50.5 years (IQR, 37.5–68.8), and 52 years (44.3–59), respectively. Males were more prevalent in the COVID-19 [$n=47$ (59.5%)] and healthy volunteers [$n=38$ (55.9%)] groups but were less prevalent in the non-COVID-19 group [$n=11$ (36.7%) %)]. The most important comorbidities among COVID-19 and non-COVID-19 patients were hypertension [$n=19$ (24.1%) *vs* $n=6$ (20%)], and diabetes [$n=11$ (13.9%) *vs* $n=2$ (6.7%)]. For the COVID-19 group, 32 patients (40.5%) had moderate disease and 47 (59.5%) developed severe disease [15]. Table 1 shows the division of *dataset II* into two subsets (training and testing) for machine learning analysis.

Dataset III (Spain dataset): LC–MS analysis

Dataset III (dataset from Spain) consisted of plasma samples from COVID-19 patients ($n=254$) diagnosed with RT-PCR and a group of healthy volunteers ($n=133$) with a negative COVID-19 test result [16]. COVID-19 patients were grouped into the following categories: (i) severe COVID-19: patients who developed severe respiratory syndrome; (ii) moderate COVID-19: patients who have opportunities on chest X-rays and need oxygen therapy; (iii) mild COVID-19: patients who were asymptomatic one week after infection with SARS-CoV-2, and who also did not need oxygen therapy. All samples were analyzed by capillary electrophoresis coupled with a time-of-flight mass spectrometer (CE-MS) with an electrospray analyzer [16]. The median age of the healthy group was lower [42 years (IQR, 36−51)] than the COVID-19 group [71 years (55−85)], but within the COVID-19 group, there was a significant age difference: mild COVID-19 57 years (50−75), moderate COVID-19 85 years (75−89), severe COVID-19 71 years (59−92). Females were more predominant among patients in the healthy volunteers' group, but for COVID-19 patients, there was no significant difference (52%). Hypertension (52%) was the most prevalent comorbidity in the COVID-19 group [16]. Table 1 shows the division of dataset III into two subsets (training and testing) for machine learning analysis.

Dataset IV (Italy dataset): GC–MS analysis

Dataset VII belongs to Italy and is composed of COVID-19 patients ($n=103$) diagnosed with RT-PCR, non-COVID-19 patients ($n=32$) who had other pneumonia but symptoms like COVID-19 with negative COVID-19 RT-PCR results, and a

group of healthy volunteers ($n=26$) [17]. Plasma samples from all three groups of patients were collected and analyzed by GC–MS, which identified 1108 metabolites, of which 556 belonged to the lipid class (467 identified by the positive ionization mode and 89 by the negative ionization mode). In the COVID-19 group, 84 of the patients had moderate disease, and 19 patients had severe disease. As for the non-COVID-19 group, 20 patients had moderate disease, and 12 patients had severe disease. The mean ages (in years) for COVID-19, non-COVI-19, and healthy patients were 67.3 ± 18.0, 69.6 ± 8.9, and 50.1 ± 5.3, respectively. The percentage of females in the COVID-19, non-COVID-19, and healthy groups were, respectively, 40.7% ($n=42$), 59.3% ($n=19$), and 57.7% ($n=15$). The time (in days) from admission to diagnosis was 5.8 ± 7.2 in the COVID-19 group and 7.7 ± 6.5 in the non-COVID-19 group. The time (in days) from diagnosis to severity was 6.5 ± 7.3 for patients with severe COVID-19 and 1.8 ± 4.9 for non-COVID-19 patients. In the COVID-19 group, the most frequent comorbidities were hypertension ($n=38$), cardiovascular disease ($n=38$), diabetes ($n=17$), and digestive system ($n=16$). The most reported symptoms in the COVID-19 group were fever ($n=52$), cough ($n=34$), and dyspnea ($n=27$) [17].

Dataset V (France dataset): LC–MS analysis

Dataset V is formed by a cohort of COVID-19 patients ($n=55$) diagnosed with RT-PCR in a hospital in France (Tours University Hospital) and samples from healthy volunteers ($n=45$) [18]. Plasma samples from both groups of patients were collected and analyzed by LC–MS. The mean age (in years) between the COVID-19 and healthy group was statistically similar (77.5 ± 16.0 *vs* 75.9 ± 17.5, $p=0.83$). The sex distribution between the COVID-19 and healthy groups was also similar (female 51% *vs* female 49%, $p=1.00$). Hypertension (56.8%), cardiovascular events (55.6%), renal failure (46.7%), and smoking (42.3%) were the most prevalent comorbidities in the COVID-19 group. The time (in days) from performing the RT-PCR test to collecting the plasma sample in the COVID-19 and healthy group was statistically similar (3.6 ± 2.6 *vs* 2.6 ± 1). The most frequent symptoms were dyspnea (64.8%), fever (66.7%), cough (61.1%), and diarrhea (14.8%), which are the most frequent symptoms reported by the COVID-19 group. A total of 66.7% of the COVID-19 group were hospitalized, and 11.1% died [18]. Table 1 shows the division of dataset V into two subsets (training and testing) for machine learning analysis.

Dataset VI (Italy dataset): LC–MS analysis

Dataset VI comprises a cohort of COVID-19 patients ($n=56$) distributed according to disease severity: mild COVID-19 ($n=20$), moderate COVID-19 ($n=16$), and COVID-19 severe ($n=20$) from two hospitals in Italy (University Hospital Federico II and Cotugno Hospital, Naples-Italy) [19]. The database also contained samples of healthy volunteers who constituted the control group ($n=9$). Plasma samples from all patient groups were collected and analyzed by LC–MS and 630 metabolites were detected, 483 of which belonged to the lipid class. The median age of the COVID-19 group was 58 years, with 70% ($n=36$) men and 30% ($n=16$) women. The median age of the control group was 46 years old; 40% were men and 60% were women. Data on patients' comorbidities were not available [19]. Table 1 shows the division of the VI dataset into two subsets (training and testing) for machine learning analysis.

Dataset VII (Italy dataset): GC–MS analysis

Database IV consisted of a prospective cohort formed by two groups: (i) COVID-19 patients ($n=24$) who were healthcare workers at a hospital in northern Italy (Novara University Hospital); (ii) a group of healthy volunteers ($n=21$) [20]. Serum samples from these patients were collected and analyzed by GC–MS, in which 322 metabolites were detected and quantified, aiming to investigate which metabolites are important for the diagnosis. The mean ages (in years) between the COVID-19 and healthy groups were, respectively, 38.9 ± 10.9 and 36.5 ± 10.1 years. The number of women between the COVID-19 and healthy groups was equal ($n=16$). The mean time from collection between collection of serum samples and diagnosis was 13.3 ± 5.1 days. Smoking was the only comorbidity reported between the COVID-19 ($n=2$) and healthy ($n=3$) groups. Fever ($n=13$) and anosmia ($n=12$) were the most important symptoms among COVID-19 patients. Table 1 shows the division of dataset IV into two training and testing subsets for machine learning analyses [20].

Data pre-processing

Pre-processing is a critical step in metabolomics. This aims to remove the experimental noise contained in the spectra, correct the baseline problems of the acquired spectra, and other interfering problems that, for some reason, it was not possible to eliminate during the process of obtaining the spectra. This allows converting raw spectral data into clean and more adjusted spectra for chemometric analyses. For each dataset, the pre-processing methods were optimized by choosing the combination of scaling, normalization, filtering, and transformation that maximizes goodness of fit (greater accuracy and lower values of training errors, e.g., RMSE). The following preprocessing options were explored: (i) Scaling and centering: mean center, multiway center, median center, square root mean

scale, multiway scale, group scale, Log decay scaling, and auto-scale; (ii) Normalization: multiplicative scatter correction (MSC-mean); standard normal variate (SNV); (iii) Variable filtering: generalized least squares weighting (GLSW), derivative (Savitzky – Golay), orthogonal signal correction (OSC), baseline (specified points), smoothing (Savitzky – Golay), baseline (weighted least square) and external parameter orthogonalization (EPO); (iv) Transformations: Log10 and absolute value [21–28]. Data filtering is particularly important in human omics sciences (metabolomics, lipidomics, etc.) due to the variability of the human population compared to studies performed with laboratory animals generally under controlled conditions [29]. This high variability is noticeable when experimental data from patients of different regions or in comparative studies where different instrumental analysis methods are used [30, 31]. In our case, the samples were analyzed in different countries (Italy, China, Spain, and France) and by different analytical techniques (NMR, GC–MS, and LC–MS). Thus, the filtering techniques applied in the present study were used to minimize these biological variations among patients.

Principal component analysis - PCA

All machine learning models were developed in SOLO software (Eigenvector Research, Copenhagen). The Principal Component Analysis (PCA) was applied to detect outlying samples and assess the degree of separation in PC space between cases and control individuals [32]. For the severity data, we assessed the degree of separation among mild COVID-19, moderate COVID-19, and severe COVID-19, and healthy individuals. For diagnostic data, PCA was employed to verify possible discrimination between COVID-19, non-COVID-19 patients, and healthy volunteers. The detection of outlier samples was further investigated by constructing the graph of leverage *vs* student residuals, where samples with high leverage values and student residuals simultaneously were identified as potential outliers and removed from the dataset [33].

Partial least squares discriminant analysis - PLS-DA

The first step in developing the PLS-DA model was the investigation of outlier samples. The detection of outlier samples was further investigated by constructing the leverage vs student residuals graph, where samples with high leverage values and student residuals simultaneously were identified as potential outliers and removed from the dataset [33]. The datasets of each of the four countries were randomly divided into two subsets: the training subset (70% of samples) and the test subset (30% of samples), Table 1. Calibration and validation were performed automatically using the Kennard-Stone algorithm [34]. The estimation of cross-validation error (RMSECV) and root mean square error of calibration (RMSEC) was performed using the Venetian Blind cross-validation method [35]. Thus, the number of latent variables of the PLS-DA model were chosen considering lower values of RMSEC and RMSECV [33]. The performance of the PLS-DA model was evaluated using sensitivity (Eq. 1), specificity (Eq. 2), and accuracy (Eq. 3) [36, 37].

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where: TP = true positive, FP = false positive, TN = true negative, FN = false negative.

Variable importance in projection (VIP)

VIP was applied to the PLS-DA models to identify relevant disease biomarkers. VIP quantifies the importance of individual variables (biomarkers) in the model prediction. Inspection of the VIP plot was employed to identify important variables visually. In the PLS-DA model, a VIP score is calculated based on the weighted sum of quadratic correlations between the original data variable and the latent variables of the PLS-DA model. The weight corresponds to the percentage of variances explained by each specific latent variable. Original variables with weights greater than 1 in the VIP chart are considered statistically significant and important for differentiating between the different classes of the samples under study [38, 39].

Comparing biomarker levels through univariate analysis

Univariate analysis was utilized to examine significant differences in biomarker levels among patients with mild, moderate, and severe COVID-19, as well as the healthy group. In this analysis, we initially assessed the normality of each biomarker (variable) using the Shapiro–Wilk test. When a biomarker exhibited a normal distribution, group comparisons were conducted using the one-way ANOVA test. Conversely, for biomarkers that deviated from a normal distribution, comparisons among the four groups were performed using the Kruskal–Wallis test. Significance levels of $p < 0.05$ were considered statistically significant.

Machine configuration used in machine learning analyzes

For the development of the machine learning models (PCA and PLS-DA) the following machine (computer) configuration was used: (i) Operating system: Windows; (ii) Machine name: DESKTOP-1A56SJ7; (iii) Processor: Intel(R) Core(TM) i5-4200U CPU @ 1.60 GHz 1.60 GHz; (iv) Installed RAM: 8.00 GB; (v) System type: 64-bit operating system, × 64-based processor.

Results

Pre-processing and data standardization

The LC–MS, GC–MS and NMR spectra used in machine learning analyzes (PCA and PLS-DA) were previously standardized through pre-processing using specific statistical techniques. The process involved imputation of missing data by column (biomarker) median, Savitzky-Golay smoothing, application of the first derivative and normalization (log10). Savitzky-Golay smoothing aimed to improve the signal/noise ratio, resulting in clearer and easier to interpret spectra. The use of the first derivative corrected shifts in the spectra in the baseline. Normalization (log10) was implemented to standardize the data and reduce biological variations between samples, which can be caused using different analytical techniques (LC–MS, GC–MS, and NMR) and the diversity of geographic origins of the samples.

PCA analysis

Visualization of the PC scores plot did not reveal clear outliers (Figs. 2 and 3). There was a reasonable separation in PC space between COVID-19 and healthy controls, as well as among COVID-19, non-COVID-19, and healthy samples. As for the severity data, mild, moderate, and severe COVID-19 patients were clearly separated.

PLS-DA model: pre-processing optimization

Figure S1 in supplementary material shows the graphs of leverage *vs* student residuals from PLS-DA model for the

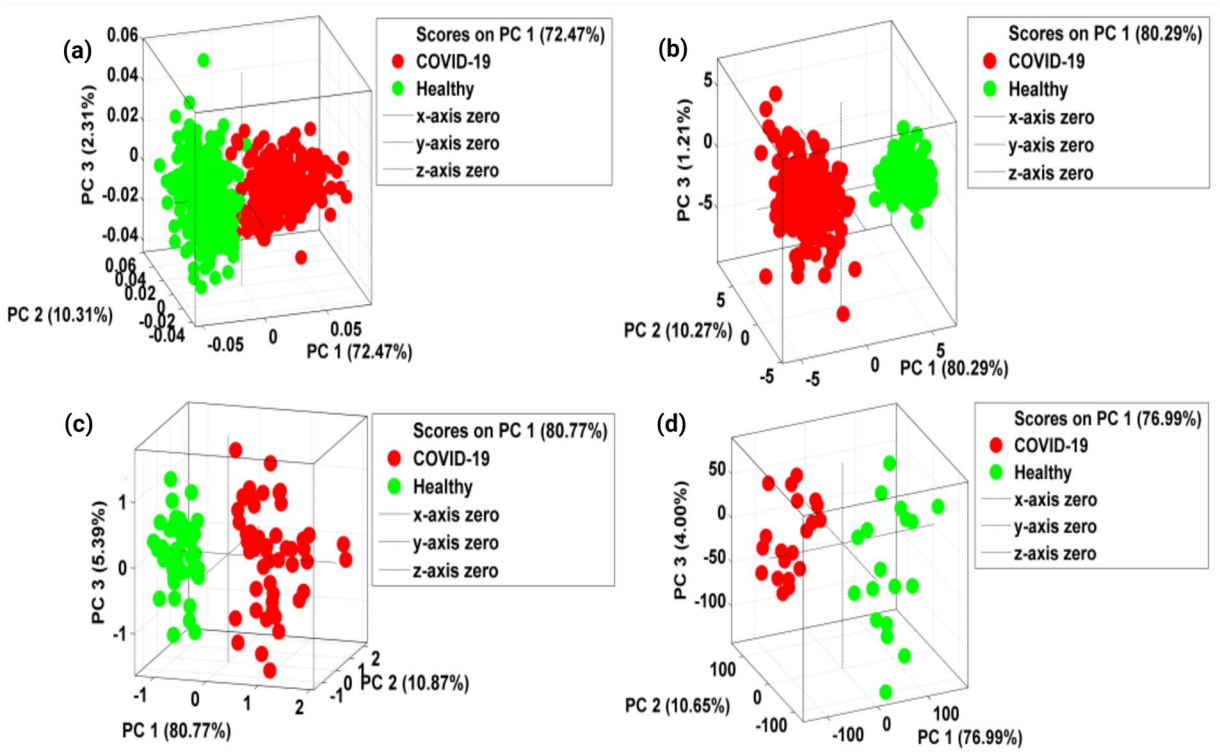
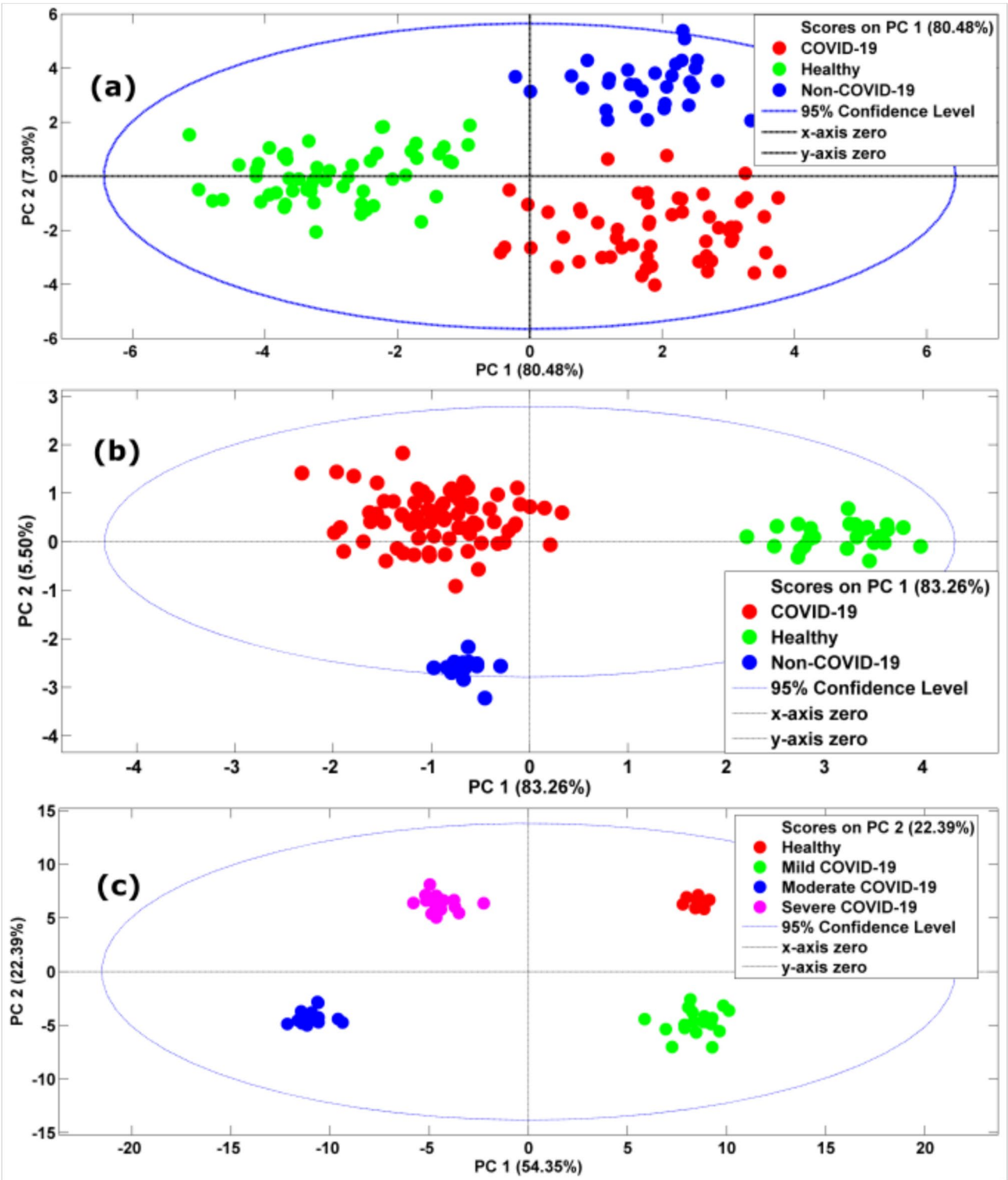


Fig. 2 PCA model for discriminating samples from COVID-19 patients (red circles) and healthy volunteers (green circles). In (**a**) a PCA model is illustrated referring to the samples of COVID-19 patients ($n = 261$) and healthy volunteers ($n = 280$) from Spain analyzed NMR (dataset 1). In (**b**) a PCA model is illustrated referring to samples from COVID-19 patients ($n = 254$) and healthy volunteers ($n = 133$) from Spain analyzed by LC–MS (dataset 3). In (**c**) a PCA model is illustrated referring to samples from COVID-19 patients ($n = 55$) and healthy volunteers ($n = 45$) from France analyzed by LC–MS (dataset 5). In (**d**) a PCA model is illustrated referring to samples from COVID-19 patients ($n = 21$) and healthy volunteers ($n = 24$) from Italy analyzed by GC–MS (dataset 7). In all datasets, the PCA model was able to discriminate between COVID-19 samples and samples from healthy volunteers (color figure online)



samples from the four countries studied. From Figure S1, although some samples had high leverage values, they were not considered outliers because they are within ± 3 standard deviations. Thus, no sample was excluded from the dataset. Several pre-processing methods (alone or in combination) and cross-validation were tested during the training of PLS-DA models. The combination of three pre-processing methods (GLSW filtering, normalize, and auto-scale) and the Venetian blind cross-validation were the ones that had the best predictive performance (higher values of accuracy, sensitivity, and specificity and lower RMSECV) in all different datasets studied. Figures S2 (supplementary material)

◂**Fig. 3** PCA model for discrimination of diagnostic and severity data samples. For diagnostic data, samples from COVID-19 patients, healthy volunteers, and non-COVID-19 are represented by old, green, and blue colored circles, respectively. In (**a**) a PCA model is illustrated referring to samples from COVID-19 patients ($n=60$), healthy volunteers ($n=57$), and non-COVID-19 ($n=30$) from Spain analyzed by GC–MS (dataset 2). In (**b**) a PCA model is illustrated referring to the samples of COVID-19 patients ($n=103$), healthy volunteers ($n=26$), and non-COVID-19 ($n=40$) from Italy analyzed by GC–MS (dataset 4). For the severity data, in (**c**) a PCA model is illustrated referring to samples of severe COVID-19 patients ($n=16$), moderate COVID-19 ($n=16$), mild COVID-19 ($n=20$), and volunteers healthy subjects ($n=9$) from Italy analyzed by LC–MS (dataset 6). Samples from healthy patients, severe COVID-19, moderate COVID-19, mild COVID-19, and healthy volunteers are represented by pink, blue, green, and red circles, respectively (dataset 6). In all datasets, the PCA model was able to discriminate between COVID-19 samples and samples from healthy volunteers

shows graphs of the number of latent variables selected for the training models as a function of the low values of the RMSECV training errors.

For the NMR plasma analysis data (data from Spain, dataset 1) the PLS-DA model was trained using five latent variables (Figure S2, supplementary material). For the LC–MS analyzes of the Spain (dataset 3), France (dataset 5), and Italy (dataset 6) the PLS-DA models were trained using seven, two, and two latent variables, respectively. For plasma samples from Italy (dataset 2) analyzed by GC–MS, the model was trained using three LV. Still in the GC–MS analyses, dataset 4 and dataset 7 both from Italy, were trained using three LV, respectively. All plots of RMSECV *vs* number of latent variables are shown in Figure S2 (supplementary material). The performance of all PLS-DA models in predicting the diagnosis and severity of COVID-19 in all countries studied are summarized in Table 2.

Variable importance in projection (VIP): COVID-19 diagnosis prediction

The VIP graphs of the most important variables in diagnosing and predicting severity are shown in Figs. 4 and 5, respectively.

The box-plot plots showing the range of biomarkers associated with severity are shown in Fig. 6 (LC–MS data from Italy, dataset 6). On the other hand, the boxplots of biomarkers associated with diagnosis for all other countries are shown in figures S8-S10 in Supplementary Material.

Patients from Spain (dataset 1, NMR analyses) with COVID-19 had higher levels of L-phenylalanine and pyruvic acid than the group of healthy volunteers. For the China GC–MS analyses (dataset 2), COVID-19 patients had higher levels of succinic acid, L-phenylalanine, lactic acid, glutamic acid, fumaric acid, and D-fructose than healthy volunteers or patients with other pneumonia (non-covid-19); however, COVID-19 patients had lower levels of 1,3-dihydroxypropan-2-yl palmitate and citric acid than healthy or non-covid-19 individuals.

In LC–MS analysis of Spain dataset (dataset 3), COVID-19 patients had higher levels of glycerol-3-phosphate, retinal, and cysteine-S-sulfate than healthy volunteers. On the other hand, the biomarkers sarcosine, levulinic acid, ribothymidine, and iminodiacetic acid were at lower levels in COVID-19 patients than in healthy volunteers. For data from France analyzed by LC–MS (dataset 4), only the biomarker 1,2 dioleoylglycerol had higher levels in COVID-19 patients than in healthy volunteers. Already, the following biomarkers were at lower levels in COVID-19 patients than in healthy or non-COVID-19 volunteers: 2-hydroxypent-3-enetrile, 1-decanol, 3-hydroxypyridine sulfate, benzothiazole, and L-lactic acid. For the diagnostic data from the samples analyzed by LC–MS (dataset 5), COVID-19 patients had higher levels of 2-hydroxybutyric acid, cytosine, asparagine, isoleucine, and N-acetyl glucosamine-1-phosphate than healthy volunteers. However, COVID-19 patients had low levels of indoxyl sulfate, glycerol myristate, and N-acetyl tryptophan (Fig. 5A). Finally, for the diagnostic data from Italy analyzed by the GC–MS (dataset 7), the following biomarkers were at higher levels in COVID-19 patients than in healthy volunteers: dihomo-gamma-linoleic acid, 9-Decenoic acid, campesterol, and erythromo-1,4-lactone.

According to the VIP (Fig. 5C) graph and the Kruskal–Wallis test (Fig. 6), the following biomarkers were at higher levels in patients with severe COVID-19 than in patients with mild COVID-19: spermidine ($p=0.018$), taurine ($p=0.048$), L-aspartic ($p=0.000$), L-glutamic ($p=0.000$), L-phenylalanine ($p=0.000$), and xanthine ($p=0.019$). Severe COVID-19 patients had lower levels of the following biomarkers when compared to mild COVID-19 patients: 5 aminopentanoic acid ($p=0.000$), dihydropiandrosterone sulfate ($p=0.003$), dodecenoyylcarinitine ($p=0.014$), and L-palmitoylcarnitine ($p=0.011$).

Table 3 shows the common biomarkers found in the different datasets that were important in predicting the diagnosis and severity of COVID-19. It is important to highlight that some of these biomarkers had already been previously identified by our group in previous studies involving other patients. of COVID-19.

Discussion

In this study, machine learning models (PCA and PLS-DA) were developed to analyze seven multi-omics datasets of COVID-19 patients in different stages of the disease (mild, moderate, and severe) coming from four different countries (Italy, France, Spain, and China) were analyzed to identify potential new diagnostic and prognostic biomarkers. From these analyses, high levels of a total of 23 diagnostic and

Table 2 Performance results of PLS-DA models in predicting the diagnosis and severity of COVID-19 in each of the seven evaluated datasets

Dataset	Country	Bioanalytics method	Group of patients	Total samples	TP	FN	TN	FP	Sensitivity	Specificity	Accuracy	Processing time
1	Spain	MNR	Healthy	280	273	7	251	10	0.98	0.96	0.97	74 s
		MNR	COVID-19	261	249	12	269	11	0.95	0.96	0.96	
2	China	GC–MS	Healthy	57	55	2	81	9	0.96	0.90	0.93	45 s
		GC–MS	COVID-19	60	54	6	83	4	0.90	0.95	0.93	
		GC–MS	Non-COVID-19	30	29	1	108	9	0.97	0.92	0.93	
3	Spain	LC–MS	Healthy	133	125	8	268	12	0.94	0.96	0.95	89 s
		LC–MS	COVID-19	254	243	11	149	10	0.96	0.94	0.95	
4	Italy	GC–MS	Healthy	26	26	0	372	15	1.00	0.96	0.96	37 s
		GC–MS	COVID-19	103	96	7	24	2	0.93	0.92	0.93	
5	France	LC–MS	Healthy	45	43	2	51	4	0.96	0.93	0.94	25 s
		LC–MS	COVID-19	55	52	3	40	5	0.95	0.89	0.92	
6	Italy	LC–MS	Healthy	9	9	0	50	2	1.00	0.96	0.97	18 s
		LC–MS	Mild	20	20	0	38	3	1.00	0.93	0.95	
		LC–MS	Moderate	16	16	0	42	3	1.00	0.93	0.95	
		LC–MS	Severe	16	16	0	44	1	1.00	0.98	0.98	
7	Italy	GC–MS	Healthy	24	23	1	21	0	0.96	1.00	0.98	12 s
		GC–MS	COVID-19	21	21	0	22	2	1.00	0.92	0.96	

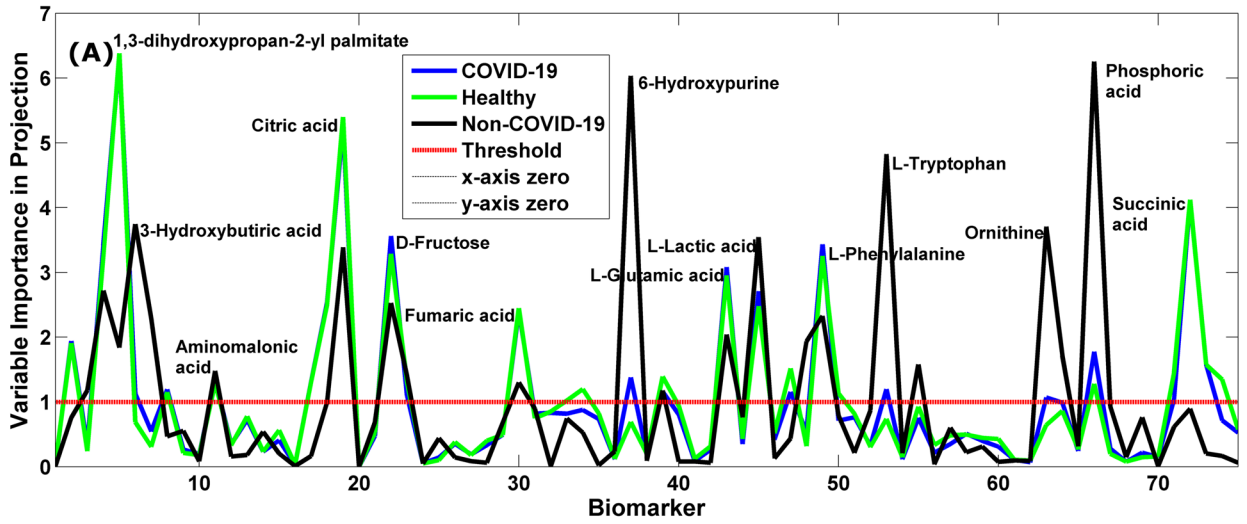
prognostic biomarkers of COVID-19 were identified. Of the 23 metabolites, three of them (N-acetyl glucosamine-1-phosphate, ornithine, and ribothymidine) had already been described in the literature for the first time as new biomarkers associated with the pathogenesis and severity of COVID-19, in a study conducted by our research group involving other patients [30]. Therefore, the present study consolidates the role of these three biomarkers in understanding the molecular and pathophysiological mechanisms of the disease.

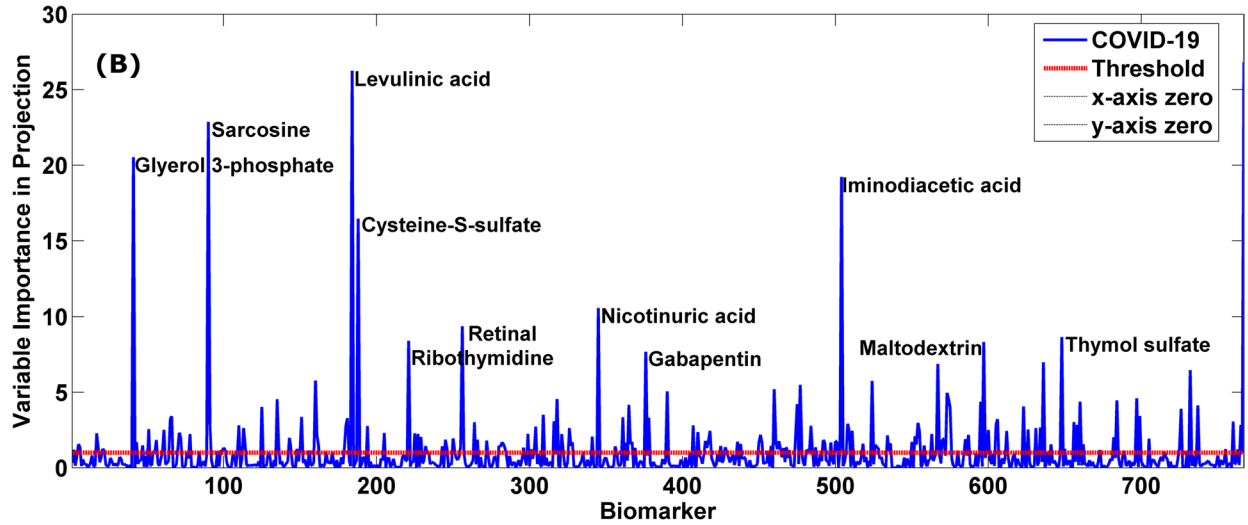
As the course of COVID-19 is very variable, caused mainly by biological variability among patients because they belong to different countries, the use of different analytical techniques: LC–MS, GC–MS and NMR, and SARS-Cov-2 mutations, the proteomics and metabolomics of these patients is very uncertain and complex, requiring the application of the PLS-DA model. This is considered a gold standard machine learning model for the analysis of metabolomics, lipidomics, and data from other omics sciences (for example, proteomics, glycomics, genomics, and transcriptomics) due to its greater predictive performance in the diagnosis and prognosis of diseases [7, 8, 30]. For example, a recent systematic review study conducted by Mendez (2018) evaluated the number of citations of studies published in the last 28 years (between 1990 and 2018) and available on the Web of Science platform, showed a growing increase in scientific publications citing PLS-DA ($n = 2242$ citations), while other algorithms [(e.g., Support Vector Machine (SVM), Artificial Neural Network (ANN), Random forest (RF), logistic regression (LR), and deep learning (DL)] were less commonly mentioned ($n = 500$ citations) [6]. Another recent study, also conducted by Mendez (2019), evaluated the predictive performance of eight different machine learning models (PLS-DA, LR, RF, principal component regression (PCR), radial basis function kernel, SVM, ANN, and non-artificial neural network) using ten different publicly available omics datasets found that the PLS-DA algorithm had the best predictive performance [8].

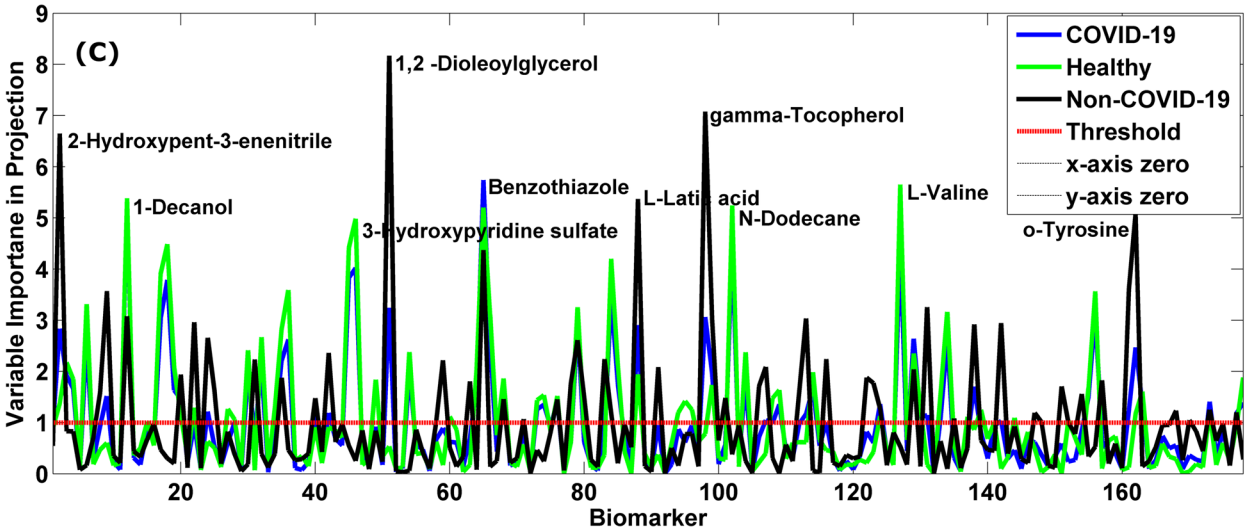
The diagnostic and prognostic accuracy of PLS-DA models developed in our study ranged from 80 to 96% AUC ROC, in agreement with previous studies available in the literature (AUC ROC 70–99%) [30, 40]. Another additional factor that increases the data's complexity is that the samples were analyzed in three different analytical techniques, NMR, LC–MS, and GC–MS, allowing the identification of a great diversity of diagnostic and prognostic biomarkers with different characteristics, physical–chemical differences, different molecular weights, and different biological functions in the patient's organism. This offers an opportunity to better understand the biochemical and pathophysiological mechanisms of infection and disease progression at the molecular level [41, 42].

In our study, from analysis of plasma samples by NMR (data from Spain) and GC–MS (data from China), patients with COVID-19 had high levels of L-phenylalanine, cytosine, asparagine, isoleucine L-aspartic, and L-glutamic than the group of healthy volunteers or patients with other pneumonia. Similar results were found in a recent study on changes in the amino acid profile in patients with SARS-CoV-2 infection found high levels of phenylalanine, glutamic acid, glutamate, tryptophan, alanine, glycine, and histidine in adult patients with COVID-19 [43]. In COVID-19, these amino acids have been associated with the severity of the disease. They are directly involved in catabolic processes because inflammatory cytokines promote the breakdown of muscle tissue, resulting in the release of amino acids, which are subsequently used in the gluconeogenesis pathway to supply the metabolic demands caused by SARS-Cov-2 infection [30, 44]. As COVID-19 patients often present with hypoxemia due to respiratory difficulties [45, 46], the pyruvate produced in the cytosol, instead of penetrating the mitochondria (respiratory chain) to produce energy, it is retained in the cytosol and converted into lactate (anaerobic respiration), with low energy production, a process catalyzed by the enzyme lactate dehydrogenase (LDH) [47, 48]. Lactate dehydrogenase is an enzyme frequently used in clinical laboratories and used as a routine test to diagnose and monitor several clinical conditions, including COVID-19 [49]. The LDH is synthesized in many tissues of the body (e.g., blood cells, muscles, heart, kidneys, and lungs) and is released into the bloodstream in cases of cell or tissue injury [50, 51]. Regarding COVID-19, some systematic reviews and meta-analyses found a strong association between high LDH levels and the most severe forms of COVID-19 [52–54]. This happens because during SARS-CoV-2 infection, lung tissue damage and inflammation can occur, which leads to the release of LDH into the bloodstream as lung cells and other tissues are damaged [50, 51].

Because of the increase in lactate levels, the emergence of metabolic acidosis causes complications in the patient's clinical condition [55]. Our study also confirmed these results where COVID-19 patients from China (Fig. 4A) and Italy (Fig. 4C) had higher lactate levels than healthy volunteers. Still, in the data from China (Fig. 4A), lower levels of citric acid were observed in COVID-19 patients than in healthy volunteers. It is known that citric acid is one of the compounds produced in the Krebs cycle (citric acid cycle) at the level of mitochondria, but as the infectious process occurs in an anaerobic state, it is expected that citric acid levels are low, as pyruvate molecules are constantly being converted to lactate, rather than producing acetyl coenzyme A, which is the initial substrate of the Krebs cycle [56]. In our study, higher levels of glutamic acid were also observed in COVID-19 patients than in healthy patients. Similar results were found in recent


(A)
1,3-dihydroxypropan-2-yl palmitate
Citric acid
3-Hydroxybutiric acid
D-Fructose
Aminomalonic acid
Fumaric acid
6-Hydroxypurine
L-Lactic acid
L-Glutamic acid
L-Phenylalanine
L-Tryptophan
Ornithine
Phosphoric acid
Succinic acid
COVID-19
Healthy
Non-COVID-19
Threshold
x-axis zero
y-axis zero
Variable Importance in Projection
Biomarker


(B)
Levulinic acid
Sarcosine
Glycerol 3-phosphate
Cysteine-S-sulfate
Retinal
Ribothymidine
Nicotinuric acid
Gabapentin
Iminodiacetic acid
Maltodextrin
Thymol sulfate
COVID-19
Threshold
x-axis zero
y-axis zero
Variable Importance in Projection
Biomarker


(C)
1,2 -Dioleoylglycerol
2-Hydroxypent-3-enenitrile
1-Decanol
Benzothiazole
3-Hydroxypyridine sulfate
gamma-Tocopherol
N-Dodecane
L-Valine
o-Tyrosine
COVID-19
Healthy
Non-COVID-19
Threshold
x-axis zero
y-axis zero
Variable Importane in Projection
Biomarker

◂**Fig. 4** Variable Importance in the projection chart of the most important biomarkers for diagnosing COVID-19. The figures in (A), (B), and (C) represent the biomarkers that were important in predicting the diagnosis of COVID-19 using data from China (dataset 2, GC–MS Data), Spain (dataset 3, LC–MS data), and Italy (dataset 4, GC–MS Data), respectively. The *X-axis* represents all analyzed metabolites; the Y-axis represents the VIP score that reflects the importance of each metabolite in predicting the different classes of samples (COVID-19 represented by blue color, non-COVID-19 by black color, and healthy volunteers by green color). The higher the VIP score, the more important the biomarker is in diagnosing COVID-19. The red dashed line parallel to the X axis represents the VIP score threshold (VIP score threshold = 1). The metabolites contributing significantly to the prediction of the different sample classes are above the threshold (VIP score > 1). In the figures in (**A**), (**B**), and (**C**) only the main biomarkers most important for the diagnosis of COVID-19 are shown (color figure online)

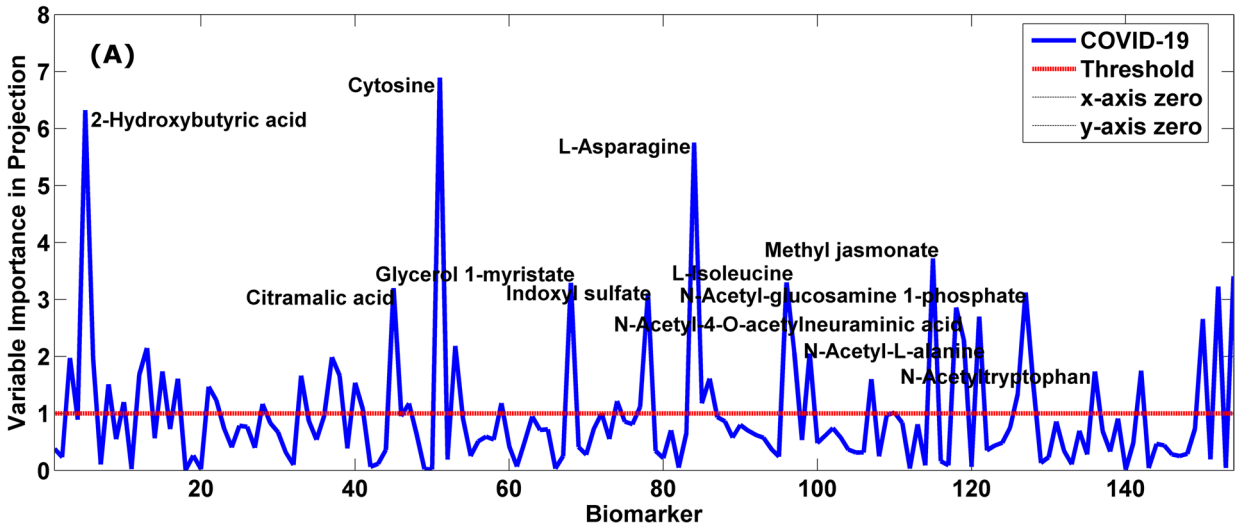
studies by Paéz-Franco (2022) and Reverté (2021), where higher levels of glutamic acid were observed in moderately severe COVID-19 patients than in patients with mild or moderate disease [57, 58]. In immune cells, glutamine is converted to glutamate, aspartate, and alanine by partial oxidation to CO_2, in a process called glutaminolysis, and this conversion plays a key role in the effective functioning of immune system cells, in addition, through the pentose phosphate pathway, a metabolic pathway parallel to the glycolysis pathway, cells can produce ribose-5-phosphate, which is a precursor to the pentose sugars seen in the structure of RNA and DNA, as well as glycerol-3-phosphate for phospholipid synthesis [59, 60]. Another biosynthetic pathway for glutamic acid is via the gamma-glutamyl cycle, which is the synthetic pathway for GSH, a tripeptide with potent antioxidant activity. The COVID-19 infection disrupts this cycle [17, 57, 61], promoting an uncontrolled increase in glutamic and pyroglutamic acid levels, causing metabolic acidosis and liver failure [62, 63]. In our study, citric acid levels were reduced in COVID-19 patients than in patients with other pneumonia or unhealthy volunteers (Fig. 4A). Citrate is an intermediate metabolite of the Krebs cycle (occurring in mitochondria), and due to the low oxygen availability caused by COVID-19, this results in reduced levels of all Krebs cycle metabolites, including citrate, resulting in decreased ATP production. Results like ours were found in the recent study by Shi (2021), proving the importance of this metabolite in the pathogenesis and diagnosis of the disease [15].

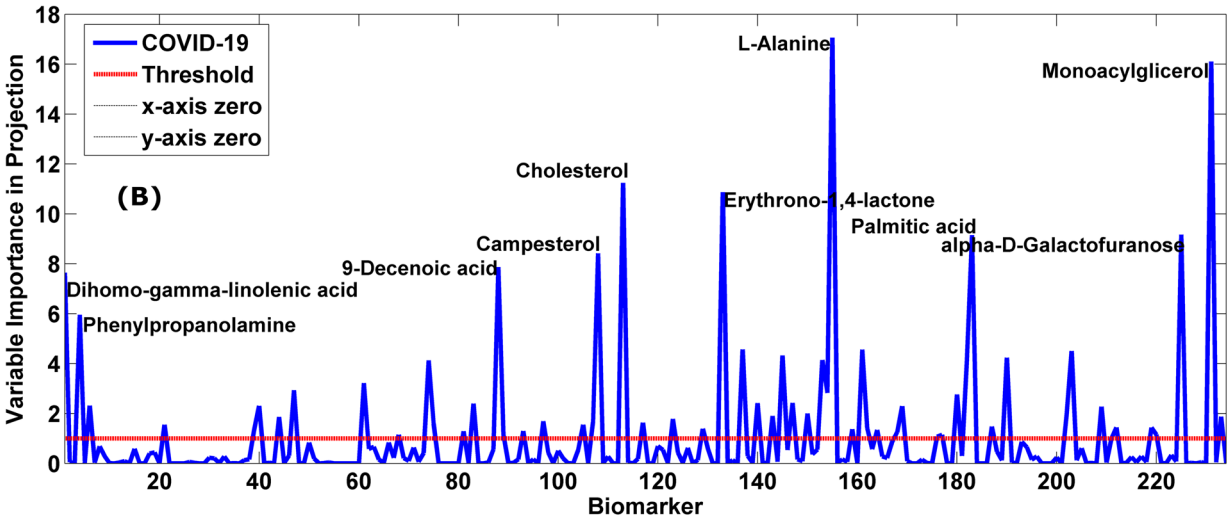
In our study, glycerol 3-phosphate levels were much higher in COVID-19 patients than in healthy volunteers. It is important to highlight that glycerol 3-phosphate is an important metabolite in glycidic and lipid metabolism, and actively regulates the body's acquired immunity in response to viral infections [64]. In COVID-19, changes in glycerol 3-phosphate levels have been used as an indicator of immunological damage caused by the virus and have been used as a biomarker associated with disease severity [65, 66].

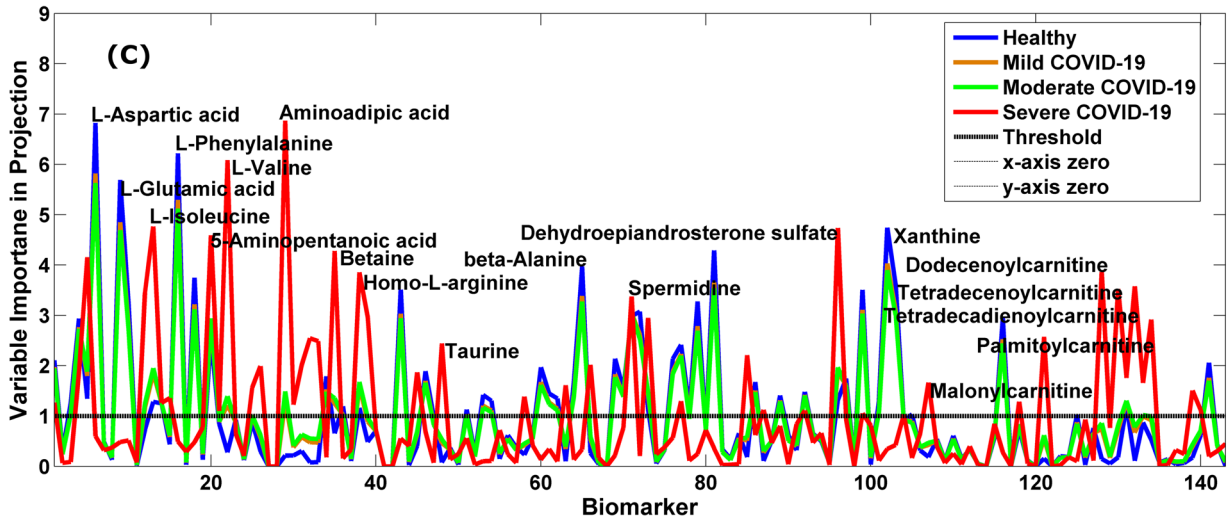
Glycerol 3-phosphate is a crucial precursor in the biosynthesis of glucose and lipids [67, 68]. The disturbances that the SARS-CoV-2 virus causes in this metabolism can result in dysfunctions in the metabolic processes of these sugars and lipids, especially glucose, triglycerides, and LDL cholesterol [69, 70]. The latter are biomarkers widely used in the clinic to diagnose and monitor conditions such as diabetes, dyslipidemia, obesity, and other metabolic diseases. Evidence indicates that patients with diabetes or high blood glucose levels are more susceptible to serious complications from COVID-19 [71]. Hyperglycemia compromises the immune system, making it more challenging for the body to fight viral infection [72, 73]. On the other hand, it is known that cholesterol plays a crucial role in the structure of cell membranes and can impact the functioning of the immune system and inflammation [74]. COVID-19 is known to trigger an intense inflammatory response in some individuals, and there is evidence linking elevated LDL cholesterol levels to a dysregulated or excessive immune response [75]. Individuals with high triglyceride levels often have underlying conditions, such as obesity, type 2 diabetes, and cardiovascular disease, which are also associated with an increased risk of serious complications from COVID-19. These comorbidities can influence the immune system response and the body's ability to fight infection [76].

Another molecule that was at higher levels in COVID-19 patients than in healthy ones was cysteine-S-sulfate. Cysteine-S-sulfate is known to be an endogenous metabolite involved in inflammatory processes [77]. The role of cysteine-S-sulphate in COVID-19 is little known, however there is a recent study citing the participation of cysteine-S-sulphate in the inflammatory processes of COVID-19, and the authors propose this molecule as a possible biomarker that should be considered in the pathogenesis and diagnosis of COVID-19 [78]. Recent studies have demonstrated that ferritin, an essential protein responsible for iron storage in the body, is being used in clinical practice as a biomarker indicative of inflammation associated with COVID-19 [79]. During COVID-19 infection, some patients experience a significant increase in ferritin levels [79]. This elevation is related to the body's inflammatory response to the viral infection, and may be indicative of an exacerbated inflammatory reaction, known as a 'cytokine storm', which may contribute to the severity of the disease [80]. As a result of this exacerbated inflammatory response in severe cases of the disease, problems such as dehydration and kidney dysfunction may arise. These conditions can cause an increase in urea levels in the blood, resulting from a decrease in the volume of water in the body and a reduction in the kidneys' ability to filter and excrete urea properly [81].

We identified several acyl-carnitine molecules (dodecenoylcarnitine, tetradecenoylcarnitine, tetradecadienoylcarnitine, palmitoylcarnitine, and malonylcarnitine) that were


(A)
COVID-19
Threshold
x-axis zero
y-axis zero
Variable Importance in Projection
Biomarker
2-Hydroxybutyric acid
Cytosine
L-Asparagine
Citramalic acid
Glycerol 1-myristate
Indoxyl sulfate
L-Isoleucine
Methyl jasmonate
N-Acetyl-glucosamine 1-phosphate
N-Acetyl-4-O-acetylneuraminic acid
N-Acetyl-L-alanine
N-Acetyltryptophan


(B)
COVID-19
Threshold
x-axis zero
y-axis zero
Variable Importance in Projection
Biomarker
L-Alanine
Monoacylglicerol
Cholesterol
Erythrono-1,4-lactone
Palmitic acid
alpha-D-Galactofuranose
Campesterol
9-Decenoic acid
Dihomo-gamma-linolenic acid
Phenylpropanolamine


(C)
Healthy
Mild COVID-19
Moderate COVID-19
Severe COVID-19
Threshold
x-axis zero
y-axis zero
Variable Importane in Projection
Biomarker
L-Aspartic acid
Aminoadipic acid
L-Phenylalanine
L-Valine
L-Glutamic acid
L-Isoleucine
5-Aminopentanoic acid
Betaine
Homo-L-arginine
Taurine
beta-Alanine
Dehydroepiandrosterone sulfate
Spermidine
Xanthine
Dodecenoylcarnitine
Tetradecenoylcarnitine
Tetradecadienoylcarnitine
Palmitoylcarnitine
Malonylcarnitine

◂**Fig. 5** Variable Importance in the projection chart of the most important biomarkers for the diagnosis of COVID-19. The figures in (**A**) and (**B**) represent the biomarkers that were important in predicting the diagnosis of COVID-19 using data from Italy (dataset 7, LC–MS data) and France (dataset 5, GC–MS Data), respectively. The figure in (C) represents the biomarkers that were important in predicting the prognosis of COVID-19 using data from Italy (LC–MS Data, dataset 6). The X-axis represents all analyzed metabolites; For figures in (**A**) and (**B**), the Y-axis represents the VIP score that reflects the importance of each metabolite in predicting the different classes of samples (COVID-19 represented by blue color, non-COVID-19 by black color, and healthy volunteers by green color). For figure in (**C**), the Y-axis represents the VIP score that reflects the importance of each metabolite in predicting the different classes of samples (severe COVID-19 represented by pink, moderate COVID-19 represented by blue, mild COVID-19 represented by green, and healthy volunteers by red color). The higher the VIP score, the more important the biomarker is in diagnosing COVID-19. The dashed line parallel to the X-axis represents the VIP score threshold (VIP score threshold = 1). The metabolites that contribute significantly to the prediction of the different sample classes are above the threshold (VIP score > 1). In the figures in (**A**), (**B**), and (**C**) only the main biomarkers most important for the diagnosis (figures **A** and **B**) and prognosis (figure **C**) of COVID-19 are shown (color figure online)

at higher levels in severe COVID-19 patients than in mildly ill or healthy volunteers. Similar results were found in the recent study by Lauro [82]. The increase in acyl-carnitine levels may indicate a marked increase in the transport of lipids to the mitochondria for the beta-oxidation process to produce energy for the cells, being a response to the low energy production of the anaerobic processes resulting from the respiratory dysfunction caused by COVID-19 [44]. In addition to the biomarkers mentioned above, we also identified for the first time some biomarkers that are associated with the diagnosis and severity of COVID-19 namely: N-Acetyl-4-O-acetylneuraminic acid, N-Acetyl-L-Alanine, N-Acetyltriptophan, palmitoylcarnitine, and glycerol 1-myristate.

As limitations of our study, we can mention that although the study was carried out involving samples of plasma and serum from patients from several countries (China, Spina, Italy, and France), we recognize that the constant emergence of new variants of the SARS-CoV-2 can alter the metabolome, proteome, and lipome of patients with COVID-19 and consequently the emergence of new biomarkers associated with these new variants. Long-term follow-up longitudinal cohort metabolic studies monitoring these patients are warranted.

Conclusion

In this study, machine learning models (PCA and PLS-DA) were used to analyze several multi-omics datasets of COVID-19 patients from different countries (China, Spina, Italy, and France) to identify new biomarkers associated with rapid diagnosis and prognosis of COVID-19, aiming to prevent long-term symptoms. The PLS-DA model was able to predict the diagnosis and prognosis of COVID-19 around 95%.

A total of 23 biomarkers have been identified as being associated with the diagnosis and prognosis of COVID-19 (e.g., spermidine, taurine, L-aspartic, L-glutamic, L-phenylalanine and xanthine, N-acetyl glucosamine-1-phosphate, ornithine, and ribothymidine). Moreover, we introduce five novel biomarkers linked to COVID-19 for the first time, encompassing N-Acetyl-4-O-acetylneuraminic acid, N-Acetyl-L-Alanine, N-Acetyltriptophan, palmitoylcarnitine, and glycerol 1-myristate. Among these, palmitoylcarnitine emerges as notably significant in predicting the prognosis of COVID-19, given its markedly heightened levels observed in patients experiencing severe symptoms in contrast to those with mild COVID-19 or healthy volunteers. Regarding the other four biomarkers (N-Acetyl-4-O-acetylneuraminic acid, N-Acetyl-L-Alanine, N-Acetyltriptophan, and glycerol 1-myristate), their prominence lies in diagnosing COVID-19 owing to their elevated presence in patients affected by the disease as opposed to the group comprising healthy volunteers or patients unaffected by COVID-19.

In summary, the comprehensive analysis of multi-omics data through advanced machine learning models has yielded a breakthrough in identifying biomarkers pivotal for the swift diagnosis and prognosis of COVID-19. These findings pave the way for enhanced diagnostic accuracy, prognosis prediction, and potential targeted interventions, marking a significant stride toward effectively managing and mitigating the long-term implications of this global health crisis.

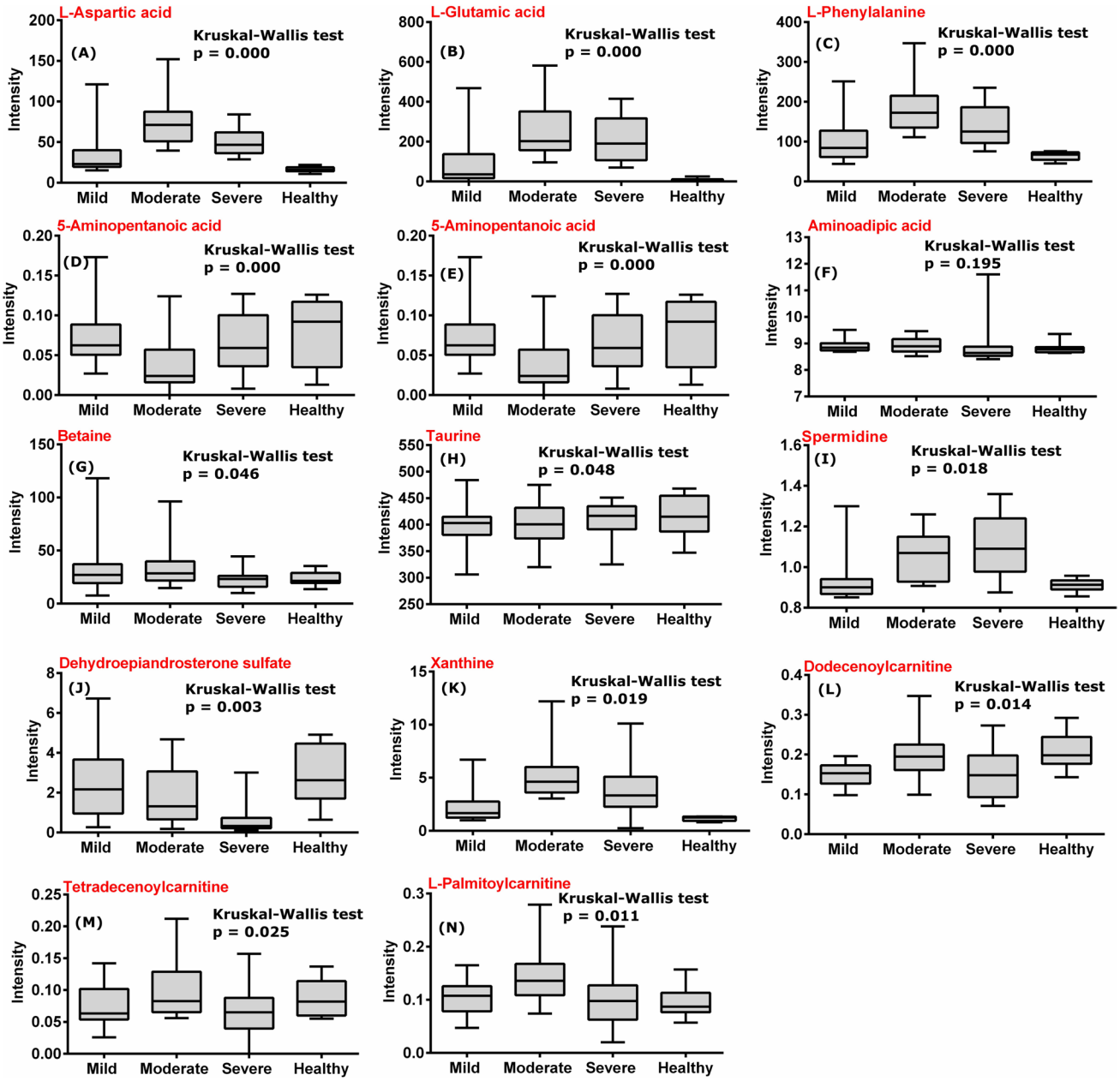


Fig. 6 Profile of the main blood biomarkers associated with the severity of COVID-19 using data from Italy (LC–MS dataset, dataset 6). Results are grouped according to classes: severe COVID-19 patients ($n=16$), moderate COVID-19 ($n=16$), mild COVID-19 ($n=20$), and volunteers healthy subjects ($n=9$). Boxes indicate interquartile ranges (median); horizontal lines indicate minimum and maximum values

Table 3 Important biomarkers in predicting the diagnosis and severity of COVID-19 were common in two or more different databases investigated in the study

Biomarker	Thaizou dataset* (LC–MS, China)	Wuhan dataset* (LC–MS, China)	Dataset 1 (MNR, Spain)	Dataset 2 (GC–MS, China)	Dataset 3 (LC–MS, Spain)	Dataset 4 (GC–MS, Italy)	Dataset 5 (LC–MS, France)	Dataset 6 (LC–MS, Italy)	Dataset 7 (GC–MS, Italy)	Outcome
Alanine	+	–	–	–	–	–	–	+	+	Diagnosis
Ribotimidine	+	–	–	–	+	–	–	–	–	Diagnosis
L-Ornithine	+	+		+	–	–		+	–	Severe
L-Valine	–	–	–	–	–	+		+	+	Severity
L-Glutamic acid				+	–	–	–	+	+	Severity
Isoleucine	+	–	–	–	–	–	–	+	+	Severity
N-acetyl-glucosamine-1 phosphate	+	–	–	–	–	+	–	–	–	Severity
2,6 dihidroxypurine	–	–	–	+	–	–	–	–	–	Diagnosis
4-Hydroxyphenylacetoyl carnitine	–	+	–	–	–	–	–	+	–	Severity

(+) present; (–) absent

*Dataset from our previous study [30]

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11739-024-03547-1.

Acknowledgements The authors express their gratitude to the Brazilian National Council of Technological and Scientific Development (CNPq) and CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil) for research funding – Finance Code 001.

Author contribution AdFC: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. ACA: Formal analysis, Writing – review & editing, Visualization. ARG: Formal analysis, Writing – review & editing, Visualization, Supervision, Visualization. RELL: Conceptualization, Methodology, Validation, Formal analysis, Investigation. KZAD: Formal analysis, Writing – review & editing. LMF: Methodology, formal analysis, Investigation, Writing – review & editing, Visualization. FST: Conceptualization, Writing – review & editing, Supervision. RP: Conceptualization, Writing – review & editing, Visualization, Supervision.

Funding The authors reported there is no funding associated with the work featured in this article.

Data Availability Study data may be made available by the corresponding author when requested by readers.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

References

1. Byeon SK, Madugundu AK, Garapati K, Ramarajan MG, Saraswat M, Kumar P-M, Hughes T, Shah R, Patnaik MM, Chia N, Ashrafzadeh-Kian S, Yao JD, Pritt BS, Cattaneo R, Salama ME, Zenka RM, Kipp BR, Grebe SKG, Singh RJ, Sadighi Akha AA, Algeciras-Schimnich A, Dasari S, Olson JE, Walsh JR, Venkatakrishnan AJ, Jenkinson G, O'Horo JC, Badley AD, Pandey A (2022) Development of a multiomics model for identification of predictive biomarkers for COVID-19 severity: a retrospective cohort study. Lancet Digit Health 4:e632–e645. https://doi.org/10.1016/S2589-7500(22)00112-1
2. Richard VR, Gaither C, Popp R, Chaplygina D, Brzhozovskiy A, Kononikhin A, Mohammed Y, Zahedi RP, Nikolaev EN, Borchers CH (2022) Early prediction of COVID-19 patient survival by targeted plasma multi-omics and machine learning. Mol Cell Proteom. https://doi.org/10.1016/j.mcpro.2022.100277
3. Frampas CF, Longman K, Spick M, Lewis HM, Costa CDS, Stewart A, Dunn-Walters D, Greener D, Evetts G, Skene DJ, Trivedi D, Pitt A, Hollywood K, Barran P, Bailey MJ (2022) Untargeted saliva metabolomics by liquid chromatography–mass spectrometry reveals markers of COVID-19 severity. PLoS ONE 17:e0274967. https://doi.org/10.1371/journal.pone.0274967
4. Ruszkiewicz DM, Sanders D, O'Brien R, Hempel F, Reed MJ, Riepe AC, Bailie K, Brodrick E, Darnley K, Ellerkmann R, Mueller O, Skarysz A, Truss M, Wortelmann T, Yordanov S, Thomas CLP, Schaaf B, Eddleston M (2020) Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry—a feasibility study. EClinicalMedicine 29–30:100609. https://doi.org/10.1016/j.eclinm.2020.100609

5. Correia BSB, Ferreira VG, Piagge PMFD, Almeida MB, Assunção NA, Raimundo JRS, Fonseca FLA, Carrilho E, Cardoso DR (2022) 1H qNMR-based metabolomics discrimination of Covid-19 severity. J Proteome Res 21:1640–1653. https://doi.org/10.1021/acs.jproteome.1c00977
6. Mendez KM, Broadhurst DI, Reinke SN (2019) The application of artificial neural networks in metabolomics: a historical perspective. Metabolomics. https://doi.org/10.1007/s11306-019-1608-0
7. Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, Goodacre R (2015) A tutorial review: metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. Anal Chim Acta 879:10–23. https://doi.org/10.1016/j.aca.2015.02.012
8. Mendez KM, Reinke SN, Broadhurst DI (2019) A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. Metabolomics 15:150. https://doi.org/10.1007/s11306-019-1612-4
9. Alwosheel A, van Cranenburgh S, Chorus CG (2018) Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. J Choice Modell 28:167–182. https://doi.org/10.1016/j.jocm.2018.07.002
10. Albóniga OE, Moreno E, Martínez-Sanz J, Vizcarra P, Ron R, Díaz-Álvarez J, Rosas M, Sánchez-Conde M, Galán JC, Angulo S, Moreno S, Barbas C, Serrano-Villar S (2023) Differential abundance of lipids and metabolites related to SARS-CoV-2 infection and susceptibility. Sci Rep. https://doi.org/10.1038/s41598-023-40999-5
11. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P-É, Li S, Xia J (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49:W388–W396. https://doi.org/10.1093/nar/gkab382
12. Saheb Sharif-Askari N, Soares NC, Mohamed HA, Saheb Sharif-Askari F, Alsayed HAH, Al-Hroub H, Salameh L, Osman RS, Mahboub B, Hamid Q, Semreen MH, Halwani R (2022) Saliva metabolomic profile of COVID-19 patients associates with disease severity. Metabolomics. https://doi.org/10.1007/s11306-022-01936-1
13. Mahmud I, Garrett TJ (2020) Mass spectrometry techniques in emerging pathogens studies: COVID-19 perspectives. J Am Soc Mass Spectrom 31:2013–2024. https://doi.org/10.1021/jasms.0c00238
14. Bruzzone C, Bizkarguenaga M, Gil-Redondo R, Diercks T, Arana E, García de Vicuña A, Seco M, Bosch A, Palazón A, San Juan I, Laín A, Gil-Martínez J, Bernardo-Seisdedos G, Fernández-Ramos D, Lopitz-Otsoa F, Embade N, Lu S, Mato JM, Millet O (2020) SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. IScience 23:101645. https://doi.org/10.1016/j.isci.2020.101645
15. Shi D, Yan R, Lv L, Jiang H, Lu Y, Sheng J, Xie J, Wu W, Xia J, Xu K, Gu S, Chen Y, Huang C, Guo J, Du Y, Li L (2021) The serum metabolome of COVID-19 patients is distinctive and predictive. Metabolism 118:154739. https://doi.org/10.1016/j.metabol.2021.154739
16. Albóniga OE, Jiménez D, Sánchez-Conde M, Vizcarra P, Ron R, Herrera S, Martínez-Sanz J, Moreno E, Moreno S, Barbas C, Serrano-Villar S (2022) Metabolic snapshot of plasma samples reveals new pathways implicated in SARS-CoV-2 pathogenesis. J Proteome Res 21:623–634. https://doi.org/10.1021/acs.jproteome.1c00786
17. Barberis E, Timo S, Amede E, Vanella VV, Puricelli C, Cappellano G, Raineri D, Cittone MG, Rizzi E, Pedrinelli AR, Vassia V, Casciaro FG, Priora S, Nerici I, Galbiati A, Hayden E, Falasca M, Vaschetto R, Sainaghi PP, Dianzani U, Rolla R, Chiocchetti A, Baldanzi G, Marengo E, Manfredi M (2020) Large-scale plasma analysis revealed new mechanisms and molecules associated with the host response to SARS-CoV-2. Int J Mol Sci 21:8623. https://doi.org/10.3390/ijms21228623
18. Blasco H, Bessy C, Plantier L, Lefevre A, Piver E, Bernard L, Marlet J, Stefic K, Benz-de Bretagne I, Cannet P, Lumbu H, Morel T, Boulard P, Andres CR, Vourc'h P, Hérault O, Guillon A, Emond P (2020) The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism. Sci Rep 10:16824. https://doi.org/10.1038/s41598-020-73966-5
19. Caterino M, Costanzo M, Fedele R, Cevenini A, Gelzo M, Di Minno A, Andolfo I, Capasso M, Russo R, Annunziata A, Calabrese C, Fiorentino G, D'Abbraccio M, Dell'Isola C, Fusco FM, Parrella R, Fabbrocini G, Gentile I, Castaldo G, Ruoppolo M (2021) The serum metabolome of moderate and severe COVID-19 patients reflects possible liver alterations involving carbon and nitrogen metabolism. Int J Mol Sci 22:9548. https://doi.org/10.3390/ijms22179548
20. Barberis E, Amede E, Tavecchia M, Marengo E, Cittone MG, Rizzi E, Pedrinelli AR, Tonello S, Minisini R, Pirisi M, Manfredi M, Sainaghi PP (2021) Understanding protection from SARS-CoV-2 using metabolomics. Sci Rep 11:13796. https://doi.org/10.1038/s41598-021-93260-2
21. Zhang T-L, Wu S, Tang H-S, Wang K, Duan Y-X, Li H (2015) Progress of chemometrics in laser-induced breakdown spectroscopy analysis. Chin J Anal Chem 43:939–948. https://doi.org/10.1016/S1872-2040(15)60832-5
22. El Haddad J, Canioni L, Bousquet B (2014) Good practices in LIBS analysis: review and advices. Spectrochim Acta Part B At Spectrosc 101:171–182. https://doi.org/10.1016/j.sab.2014.08.039
23. Galbács G (2015) A critical review of recent progress in analytical laser-induced breakdown spectroscopy. Anal Bioanal Chem 407:7537–7562. https://doi.org/10.1007/s00216-015-8855-3
24. Castro JP, Pereira-Filho ER (2016) Twelve different types of data normalization for the proposition of classification{,} univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS). J Anal At Spectrom 31:2005–2014. https://doi.org/10.1039/C6JA00224B
25. Zorov NB, Gorbatenko AA, Labutin TA, Popov AM (2010) A review of normalization techniques in analytical atomic spectrometry with laser sampling: from single to multivariate correction. Spectrochim Acta Part B At Spectrosc 65:642–657. https://doi.org/10.1016/j.sab.2010.04.009
26. Pořízka P, Klus J, Hrdlička A, Vrábel J, Škarková P, Prochazka D, Novotný J, Novotný K, Kaiser J (2017) Impact of laser-induced breakdown spectroscopy data normalization on multivariate classification accuracy. J Anal At Spectrom 32:277–288. https://doi.org/10.1039/C6JA00322B
27. Pořízka P, Klus J, Prochazka D, Képeš E, Hrdlička A, Novotný J, Novotný K, Kaiser J (2016) Laser-induced breakdown spectroscopy coupled with chemometrics for the analysis of steel: the issue of spectral outliers filtering. Spectrochim Acta Part B At Spectrosc 123:114–120. https://doi.org/10.1016/j.sab.2016.08.008
28. Yaroshchyk P, Eberhardt JE (2014) Automatic correction of continuum background in laser-induced breakdown spectroscopy using a model-free algorithm. Spectrochim Acta Part B At Spectrosc 99:138–149. https://doi.org/10.1016/j.sab.2014.06.020
29. Brindle JT, Nicholson JK, Schofield PM, Grainger DJ, Holmes E (2003) Application of chemometrics to 1H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. Analyst 128:32–36. https://doi.org/10.1039/b209155k
30. de Fátima Cobre A, Surek M, Stremel DP, Fachi MM, Lobo Borba HH, Tonin FS, Pontarolo R (2022) Diagnosis and prognosis of COVID-19 employing analysis of patients' plasma and serum via

LC–MS and machine learning. Comput Biol Med 146:105659. https://doi.org/10.1016/j.compbiomed.2022.105659
31. Pang Z, Zhou G, Chong J, Xia J (2021) Comprehensive meta-analysis of COVID-19 global metabolomics datasets. Metabolites 11:44. https://doi.org/10.3390/metabo11010044
32. Folch-Fortuny A, Arteaga F, Ferrer A (2015) PCA model building with missing data: new proposals and a comparative study. Chemometr Intell Lab Syst 146:77–88. https://doi.org/10.1016/j.chemolab.2015.05.006
33. de Fátima Cobre A, Stremel DP, Noleto GR, Fachi MM, Surek M, Wiens A, Tonin FS, Pontarolo R (2021) Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? Comput Biol Med 134:104531. https://doi.org/10.1016/j.compbiomed.2021.104531
34. Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11:137–148. https://doi.org/10.1080/00401706.1969.10490666
35. Wienold J, Iwata T, Sarey Khanie M, Erell E, Kaftan E, Rodriguez RG, Yamin Garreton JA, Tzempelikos T, Konstantzos I, Christoffersen J, Kuhn TE, Pierson C, Andersen M (2019) Cross-validation and robustness of daylight glare metrics. Light Res Technol 51:983–1013. https://doi.org/10.1177/1477153519826003
36. Ruiz-Perez D, Guan H, Madhivanan P, Mathee K, Narasimhan G (2020) So you think you can PLS-DA? BMC Bioinform 21:2. https://doi.org/10.1186/s12859-019-3310-7
37. Ballabio D, Consonni V (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal Methods 5:3790–3798. https://doi.org/10.1039/c3ay40582f
38. Favilla S, Durante C, Vigni ML, Cocchi M (2013) Assessing feature relevance in NPLS models by VIP. Chemom Intell Lab Syst 129:76–86. https://doi.org/10.1016/j.chemolab.2013.05.013
39. Cocchi M, Biancolillo A, Marini F (2018) Chapter ten - chemometric methods for classification and feature selection. In: Jaumot J, Bedia C, Tauler R (eds) Data analysis for omic sciences: methods and applications. Elsevier, pp 265–299
40. Wang L, Zhang Y, Wang D, Tong X, Liu T, Zhang S, Huang J, Zhang L, Chen L, Fan H, Clarke M (2021) Artificial intelligence for COVID-19: a systematic review. Front Med (Lausanne) 8:704256. https://doi.org/10.3389/fmed.2021.704256
41. Simón-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, Mallard WG, Bearden DW, Schock TB, Tchekhovskoi DV, Blonder N, Yan X, Liang Y, Zheng Y, Wallace WE, Neta P, Phinney KW, Remaley AT, Stein SE (2013) Metabolite profiling of a NIST standard reference material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. Anal Chem 85:11725–11731. https://doi.org/10.1021/ac402503m
42. Zhang J, Bowers J, Liu L, Wei S, Gowda GAN, Hammoud Z, Raftery D (2012) Esophageal cancer metabolite biomarkers detected by LC-MS and NMR methods. PLoS ONE 7:e30181. https://doi.org/10.1371/journal.pone.0030181
43. Rees CA, Rostad CA, Mantus G, Anderson EJ, Chahroudi A, Jaggi P, Wrammert J, Ochoa JB, Ochoa A, Basu RK, Heilman S, Harris F, Lapp SA, Hussaini L, Vos MB, Brown LA, Morris CR (2021) Altered amino acid profile in patients with SARS-CoV-2 infection. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.2101708118
44. Luporini RL, Pott-Junior H, Di Medeiros Leal MCB, Castro A, Ferreira AG, Cominetti MR, de Freitas Anibal F (2021) Phenylalanine and COVID-19: tracking disease severity markers. Int Immunopharmacol 101:108313. https://doi.org/10.1016/j.intimp.2021.108313
45. Kamel KS, Oh MS, Halperin ML (2020) L-lactic acidosis: pathophysiology, classification, and causes; emphasis on biochemical and metabolic basis. Kidney Int 97:75–88. https://doi.org/10.1016/j.kint.2019.08.023
46. Nechipurenko YD, Semyonov DA, Lavrinenko IA, Lagutkin DA, Generalov EA, Zaitceva AY, Matveeva OV, Yegorov YE (2021) The role of acidosis in the pathogenesis of severe forms of COVID-19. Biology (Basel) 10:852. https://doi.org/10.3390/biology10090852
47. De Backer D (2003) Lactic acidosis. Intensive Care Med 29:699–702. https://doi.org/10.1007/s00134-003-1746-7
48. Li J, Wang X, Chen J, Zuo X, Zhang H, Deng A (2020) COVID-19 infection may cause ketosis and ketoacidosis. Diabetes Obes Metab 22:1935–1941. https://doi.org/10.1111/dom.14057
49. Henry BM, Aggarwal G, Wong J, Benoit S, Vikse J, Plebani M, Lippi G (2020) Lactate dehydrogenase levels predict coronavirus disease 2019 (COVID-19) severity and mortality: a pooled analysis. Am J Emerg Med 38:1722–1726. https://doi.org/10.1016/j.ajem.2020.05.0734
50. Li X, Yang Y, Zhang B, Lin X, Fu X, An Y, Zou Y, Wang JX, Wang Z, Yu T (2022) Lactate metabolism in human health and disease. Signal Transduct Target Ther. https://doi.org/10.1038/s41392-022-01151-3
51. Adeva-Andany M, López-Ojén M, Funcasta-Calderón R, Ameneiros-Rodríguez E, Donapetry-García C, Vila-Altesor M, Rodríguez-Seijas J (2014) Comprehensive review on lactate metabolism in human health. Mitochondrion 17:76–100. https://doi.org/10.1016/j.mito.2014.05.007
52. Martha JW, Wibowo A, Pranata R (2021) Prognostic value of elevated lactate dehydrogenase in patients with COVID-19: a systematic review and meta-analysis. Postgrad Med J. https://doi.org/10.1136/postgradmedj-2020-139542
53. Hariyanto TI, Japar KV, Kwenandar F, Damay V, Siregar JI, Lugito NPH, Tjiang MM, Kurniawan A (2021) Inflammatory and hematologic markers as predictors of severe outcomes in COVID-19 infection: a systematic review and meta-analysis. Am J Emerg Med 41:110–119. https://doi.org/10.1016/j.ajem.2020.12.076
54. Mehta AA, Haridas N, Belgundi P, Jose WM (2021) A systematic review of clinical and laboratory parameters associated with increased severity among COVID-19 patients. Diabetes Metab Syndr 15:535–541. https://doi.org/10.1016/j.dsx.2021.02.020
55. Carpenè G, Onorato D, Nocini R, Fortunato G, Rizk JG, Henry BM, Lippi G (2022) Blood lactate concentration in COVID-19: a systematic literature review. Clin Chem Lab Med 60:332–337. https://doi.org/10.1515/cclm-2021-1115
56. Li Z, Liu G, Wang L, Liang Y, Zhou Q, Wu F, Yao J, Chen B (2020) From the insight of glucose metabolism disorder: oxygen therapy and blood glucose monitoring are crucial for quarantined COVID-19 patients. Ecotoxicol Environ Saf 197:110614. https://doi.org/10.1016/j.ecoenv.2020.110614
57. Páez-Franco JC, Maravillas-Montero JL, Mejía-Domínguez NR, Torres-Ruiz J, Tamez-Torres KM, Pérez-Fragoso A, Germán-Acacio JM, Ponce-de-León A, Gómez-Martín D, Ulloa-Aguirre A (2022) Metabolomics analysis identifies glutamic acid and cystine imbalances in COVID-19 patients without comorbid conditions. Implications on redox homeostasis and COVID-19 pathophysiology. PLoS ONE 17:e0274910. https://doi.org/10.1371/journal.pone.0274910
58. Reverté L, Yeregui E, Olona M, Gutiérrez-Valencia A, Buzón MJ, Martí A, Gómez-Bertomeu F, Auguet T, López-Cortés LF, Burgos J, Benavent-Bofill C, Boqué C, García-Pardo G, Ruiz-Mateos E, Mestre MT, Vidal F, Viladés C, Peraire J, Rull A (2022) Fetuin-A, inter-α-trypsin inhibitor, glutamic acid and ChoE (18:0) are key biomarkers in a panel distinguishing mild from critical coronavirus disease 2019 outcomes. Clin Transl Med 12:e704. https://doi.org/10.1002/ctm2.704
59. Cruzat V, Rogero MM, Keane KN, Curi R, Newsholme P (2018) Glutamine: metabolism and immune function. Suppl Clin Transl 10:1–31. https://doi.org/10.3390/nu10111564
60. Leite JSM, Cruzat VF, Krause M, Homem de Bittencourt PI (2016) Physiological regulation of the heat shock response by

glutamine: implications for chronic low-grade inflammatory diseases in age-related conditions. Nutrire 41:1–34. https://doi.org/10.1186/s41110-016-0021-y
61. Doğan HO, Şenol O, Bolat S, Yıldız ŞN, Büyüktuna SA, Sarıismailoğlu R, Doğan K, Hasbek M, Hekim SN (2021) Understanding the pathophysiological changes via untargeted metabolomics in COVID-19 patients. J Med Virol 93:2340–2349. https://doi.org/10.1002/jmv.26716
62. Hložek T, Křížek T, Tůma P, Bursová M, Coufal P, Čabala R (2017) Quantification of paracetamol and 5-oxoproline in serum by capillary electrophoresis: Implication for clinical toxicology. J Pharm Biomed Anal 145:616–620. https://doi.org/10.1016/j.jpba.2017.07.024
63. Al-Jishi E, Meyer BF, Rashed MS, Al-Essa M, Al-Hamed MH, Sakati N, Sanjad S, Ozand PT, Kambouris M (1999) Clinical, biochemical, and molecular characterization of patients with glutathione synthetase deficiency. Clin Genet 55:444–449. https://doi.org/10.1034/j.1399-0004.1999.550608.x
64. Collison LW, Murphy EJ, Jolly CA (2008) Glycerol-3-phosphate acyltransferase-1 regulates murine T-lymphocyte proliferation and cytokine production. Am J Physiol Cell Physiol 295:C1543–C1549. https://doi.org/10.1152/ajpcell.00371.2007
65. Wu D, Shu T, Yang X, Song J-X, Zhang M, Yao C, Liu W, Huang M, Yu Y, Yang Q, Zhu T, Xu J, Mu J, Wang Y, Wang H, Tang T, Ren Y, Wu Y, Lin S-H, Qiu Y, Zhang D-Y, Shang Y, Zhou X (2020) Plasma metabolomic and lipidomic alterations associated with COVID-19. Natl Sci Rev 7:1157–1168. https://doi.org/10.1093/nsr/nwaa086
66. Chanda B, Xia Y, Mandal MK, Yu K, Sekine K-T, Gao Q, Selote D, Hu Y, Stromberg A, Navarre D, Kachroo A, Kachroo P (2011) Glycerol-3-phosphate is a critical mobile inducer of systemic immunity in plants. Nat Genet 43:421–427. https://doi.org/10.1038/ng.798
67. Jensen MD, Ekberg K, Landau BR, Landau Lipid BR (2001) Lipid metabolism during fasting. http://www.ajpendo.org
68. Xue LL, Chen HH, Jiang JG (2017) Implications of glycerol metabolism for lipid production. Prog Lipid Res 68:12–25. https://doi.org/10.1016/j.plipres.2017.07.002
69. Abu-Farha M, Thanaraj TA, Qaddoumi MG, Hashem A, Abubaker J, Al-Mulla F (2020) The role of lipid metabolism in COVID-19 virus infection and as a drug target. Int J Mol Sci 21:3544. https://doi.org/10.3390/ijms21103544
70. Mahrooz A, Muscogiuri G, Buzzetti R, Maddaloni E (2021) The complex combination of COVID-19 and diabetes: pleiotropic changes in glucose metabolism. Endocrine 72:317–325. https://doi.org/10.1007/s12020-021-02729-7
71. Huang I, Lim MA, Pranata R (2020) Diabetes mellitus is associated with increased mortality and severity of disease in COVID-19 pneumonia—a systematic review, meta-analysis, and meta-regression: diabetes and COVID-19. Diabetes Metab Syndr 14:395–403. https://doi.org/10.1016/j.dsx.2020.04.018
72. Han M, Ma K, Wang X, Yan W, Wang H, You J, Wang Q, Chen H, Guo W, Chen T, Ning Q, Luo X (2021) Immunological characteristics in type 2 diabetes mellitus among COVID-19 patients. Front Endocrinol (Lausanne). https://doi.org/10.3389/fendo.2021.596518
73. Cheng Y, Yue L, Wang Z, Zhang J, Xiang G (2021) Hyperglycemia associated with lymphopenia and disease severity of COVID-19 in type 2 diabetes mellitus. J Diabetes Compl 35:107809. https://doi.org/10.1016/j.jdiacomp.2020.107809
74. Tall AR, Yvan-Charvet L (2015) Cholesterol, inflammation and innate immunity. Nat Rev Immunol 15:104–116. https://doi.org/10.1038/nri3793
75. Kočar E, Režen T, Rozman D (2021) Cholesterol, lipoproteins, and COVID-19: basic concepts and clinical applications. Biochim Biophys Acta Mol Cell Biol Lipids 1866:158849. https://doi.org/10.1016/j.bbalip.2020.158849
76. Masana L, Correig E, Ibarretxe D, Anoro E, Arroyo JA, Jericó C, Guerrero C, Miret ML, Näf S, Pardo A, Perea V, Pérez-Bernalte R, Plana N, Ramírez-Montesinos R, Royuela M, Soler C, Urquizu-Padilla M, Zamora A, Pedro-Botet J, Rodríguez-Borjabad C, Andreychuk N, Malo A, Matas L, del Señor Cortes-Fernandez M, Mauri M, Borrallo RM, Pedragosa À, Gil-Lluís P, Lacal-Martínez A, Barragan-Galló P, Vives-Masdeu G, Arto-Fernández C, El Boutrouki O, Vázquez-Escobales A, Antón-Alonso MC, Rivero-Santana S, Gómez A, García S, Rial-Lorenzo N, Ruiz-Ortega L, Alonso-Gisbert O, Méndez-Martínez AI, Iglesias-López H, Climent E, Güerri R, Soldado J, Fanlo M, Taboada A, Gutierrez L (2021) Low HDL and high triglycerides predict COVID-19 severity. Sci Rep. https://doi.org/10.1038/s41598-021-86747-5
77. Abbas A-K, Xia W, Tranberg M, Wigström H, Weber SG, Sandberg M (2008) S-sulfo-cysteine is an endogenous amino acid in neonatal rat brain but an unlikely mediator of cysteine neurotoxicity. Neurochem Res 33:301–307. https://doi.org/10.1007/s11064-007-9441-7
78. Cai Y, Kim DJ, Takahashi T, Broadhurst DI, Yan H, Ma S, Rattray NJW, Casanovas-Massana A, Israelow B, Klein J, Lucas C, Mao T, Moore AJ, Muenker MC, Oh JE, Silva J, Wong P, Ko AI, Khan SA, Iwasaki A, Johnson CH (2021) Kynurenic acid may underlie sex-specific immune responses to COVID-19. Sci Signal. https://doi.org/10.1126/scisignal.abf8483
79. Cheng L, Li H, Li L, Liu C, Yan S, Chen H, Li Y (2020) Ferritin in the coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. J Clin Lab Anal. https://doi.org/10.1002/jcla.23618
80. Melo AKG, Milby KM, Caparroz ALMA, Pinto ACPN, Santos RRP, Rocha AP, Ferreira GA, Souza VA, Valadares LDA, Vieira RMRA, Pileggi GS, Trevisani VFM (2021) Biomarkers of cytokine storm as red flags for severe and fatal COVID-19 cases: a living systematic review and meta-analysis. PLoS ONE 16:e0253894. https://doi.org/10.1371/journal.pone.0253894
81. Dewulf JP, Martin M, Marie S, Oguz F, Belkhir L, De Greef J, Yombi JC, Wittebole X, Laterre PF, Jadoul M, Gatto L, Bommer GT, Morelle J (2022) Urine metabolomics links dysregulation of the tryptophan-kynurenine pathway to inflammation and severity of COVID-19. Sci Rep. https://doi.org/10.1038/s41598-022-14292-w
82. Martínez-Gómez LE, Ibarra-González I, Fernández-Lainez C, Tusie T, Moreno-Macías H, Martinez-Armenta C, Jimenez-Gutierrez GE, Vázquez-Cárdenas P, Vidal-Vázquez P, Ramírez-Hinojosa JP, Rodríguez-Zulueta AP, Vargas-Alarcón G, Rojas-Velasco G, Sánchez-Muñoz F, Posadas-Sanchez R, de Martínez-Ruiz FJ, Zayago-Angeles DM, Moreno ML, Barajas-Galicia E, Lopez-Cisneros G, Gonzalez-Fernández NC, Ortega-Peña S, Herrera-López B, Olea-Torres J, Juárez-Arias M, Rosas-Vásquez M, Cabrera-Nieto SA, Magaña JJ, Camacho-Rea MDC, Suarez-Ahedo C, Coronado-Zarco I, Valdespino-Vázquez MY, Martínez-Nava GA, Pineda C, Vela-Amieva M, López-Reyes A (2022) Metabolic reprogramming in SARS-CoV-2 infection impacts the outcome of COVID-19 patients. Front Immunol 13:936106. https://doi.org/10.3389/fimmu.2022.936106