Survival Prediction in Critically Ill Patients based on the Serum Molecular Fingerprint

Inês Correia*, Rúben Araújo*, Tiago A. H. Fonseca, Cristiana P. Von Rekowski, Luís Bento and Cecília R.C. Calado

***Abstract* — It is relevant to discover biomarkers enabling to predict critically ill patients' survival. This study focused on 45 patients, from which 22 deceased and 23 were discharged from an Intensive Care Unit (ICU). It was considered the serum molecular fingerprint, as acquired by Fourier Transform Infra-Red (FTIR) spectroscopy, obtained 3 days before the patients discharged or death at the ICU. It was possible to obtain ratios of bands of the sera spectra, statistically different between the two groups of patients. Furthermore, good Naïve Bayes models were developed based on the second derivative spectra enabling an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 0.77. These promising outputs suggest further investigation with a larger cohort.**

***Keywords*—Intensive Care Unit, Survival and Biomarkers**

I. Introduction

Due to the relevance of mortality prediction in critically ill patients, it is common practice at intensive care units (ICU) to use physiological scores, *e.g.*, Acute Physiology and Chronic Health Evaluation (APACHE). However, these types of scorings don't enable the prediction of individual patients' outcomes, being mostly used for comparing groups of patients and ICUs [1].

FTIR spectroscopy, associated to machine learning algorithms can represent an appealing method to discover sensitive and specific biomarkers for medical diagnosis and prognosis [2]. Indeed, this type of platforms, by capturing the whole molecular fingerprint of a defined biofluid, have been evaluated for diagnosis of diverse diseases [3]. In this work, the serum whole molecular fingerprint, as captured by FTIR-spectroscopy was evaluated to discover serum biomarkers enabling survival prediction, 3 days before its occurrence at an ICU.

II. Materials and Methods

A. Population

A total of 45 patients admitted at the ICU of the *Hospital de São José*, in Lisbon, were considered, from which 22 died at the ICU. The present study is inserted in the PREMO project, approved by the Hospital Ethics Committee, *Unidade Local de Saúde São José*, with the informed consent obtained from each patient or their family members for data collection before participation. The patients' clinical information used was anonymized.All patients were critically ill patients with COVID-19 and were under invasive mechanical ventilation.

B. Blood collection

Peripheral blood was collected in a tube with no anticoagulant at the ICU, between 7 and 9 a.m., and maintained at−4ºC, between 2 to 4 hours, till centrifugation (3500 rpm for 10min in a centrifuge Mikro220T, Hettich, Tuttlingen, Germany). Serum samples were maintained at −80ºC until further analysis. Samples were collected 3 days before patients' death or discharge from the ICU.

C. Serum whole molecular fingerprint acquisition

Triplicates of 25µL of serum, pre-diluted at 1/10 in water, from each sample, were pipetted to a 96-well Si plate and subsequently dehydrated for about 3.5h in a desiccator under vacuum (Vacuubrand, ME2, Wertheim, Germany). Spectral data were collected using an FTIR spectrometer (Vertex70, Bruker) equipped with an HTS-XT (Bruker, Billerica, MA, USA) accessory. Each spectrum represented 64 coadded scans, with a $2cm^{-1}$ resolution, and was collected in transmission mode, between 400 and 4000 cm^{-1}. The first well of the 96-well plate did not contain a sample and the corresponding spectra were acquired and used as the background, according to the HTS-XT manufacturer.

D. Spectra pre-processing and processing

Spectra with atmospheric correction were subsequently submitted to baseline correction and unit vector normalization or, in alternative, the second derivative spectra were obtained using a Savitzky–Golay filter, with a second-order polynomial over a 15-point window. The impact of spectra pre-processing was evaluated on a Principal Component analysis (PCA). The following predicting models were developed: PCA-Linear Discriminant Analysis (PCA-LDA), Support Vector Machines (SVM) and Naïve-Bayes models. A cross-validation method with 5 folds (80% training, 20% test size) were applied. Spectra pre-processing, PCA, PCA-LDA and SVM were conducted with the Unscrambler ® X 10.4 software (CAMO software AS, Oslo, Norway). The Naïve-Bayes model was conducted with the Mining Toolbox [4], version 3.36.2 (Bioinformatics Lab, University of Ljubljana, Ljubljana, Slovenia).

Manuscript received: October 2024.

I. Correia, ISEL- Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Lisbon, Portugal

R. Araújo, C. Rekowski and T. Fonseca, ISEL; NMS- Nova Medical School, Universidade Nova de Lisboa, Lisbon, Portugal; and CHRC - Comprehensive Health Research Centre, Universidade Nova de Lisboa.

L. Bento, Intensive Care Department, ULSSJ - Unidade Local de Saúde de São José, Lisbon, Portugal and from Integrated Pathophysiological Mechanisms, CHRC.

C. Calado, ISEL and iBB-Institute for Bioengineering and Biosciences, The Associate Laboratory Institute for Health and Bioeconomy (i4HB), Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.

Univariate data analysis of spectral bands among the populations were conducted by the non-parametric Mann–Whitney U test, with IBM SPSS Statistics software, version 27 (IBM Corp., New York, USA).

III. Results and Discussion

The two groups of patients (*i.e.*, discharged or deceased), were not statistically different concerning variables such as age, gender and if under Extracorporeal Membrane Oxygenation (ECMO) (Table I). This increases the probability that the predictive model output isn't due to possible confounding variables.

TABLE I. DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF THE 45 PATIENTS, WITH THE P-VALUE OF THE STATISTICAL ANALYSIS COMPARING THESE TWO GROUPS.

Variable		Discharged (n=23 patients)	Deceased (n=22 patients)	*p*-value
Age (years), (median/IQR)		61 (16)	67(15)	0.309[#]
Gender (n/proportion)	Female	10 (0.45)	8 (0.35)	0.670[*]
	Male	12 (0.55)	15 (0.65)	
ECMO (n/proportion)	No	20 (0.91)	23 (1.00)	0.233[+]
	Yes	2 (0.09)	0 (0.00)	

[#]Mann-Whitney U, *Students t-test, [+] Fishers exact test.

Fig. 1A represents all serum spectra after baseline correction, while Fig. 1B represents the average of the serum spectra of the patients that were deceased or discharged from the ICU. These average spectra are very similar, and consequently the PCA score plot (Fig. 1C) did not enable a data pattern separation between the two groups of patients.

Second derivative, by resolving spectral bands, increases the differences between the two groups of patients (Fig. 2A) improving the separation in the PCA score plot between the patient's group (Fig. 2B). Despite that, a data pattern separation between the two groups of patients is not clear.

The major bands pointed in the normalized baseline-corrected spectra and corresponding PCA loadings, were analyzed. From the 21 bands and 84 ratios of bands analyzed, 17 were statistically different (p<0.05) between the two groups of patients (Table II). The band ratio that was the most significantly different between the two groups, was between 1321 cm^{-1} / 1244 cm^{-1}, associated to lipids and phosphate groups, respectively. As expected, the same analysis based on the second derivative spectra, resulted in a much higher number of bands and ratios between bands (n=41), statistically different, between the two groups (Table III), mostly due to resolution of overlapped bands. The most significant ratios between bands of the second derivative spectra included, the Amide III (1400 cm^{-1}), amide I and II (1657 and 1543 cm^{-1}), esters of phospholipids (1744 cm^{-1}), and the fingerprint region (660 cm^{-1}).

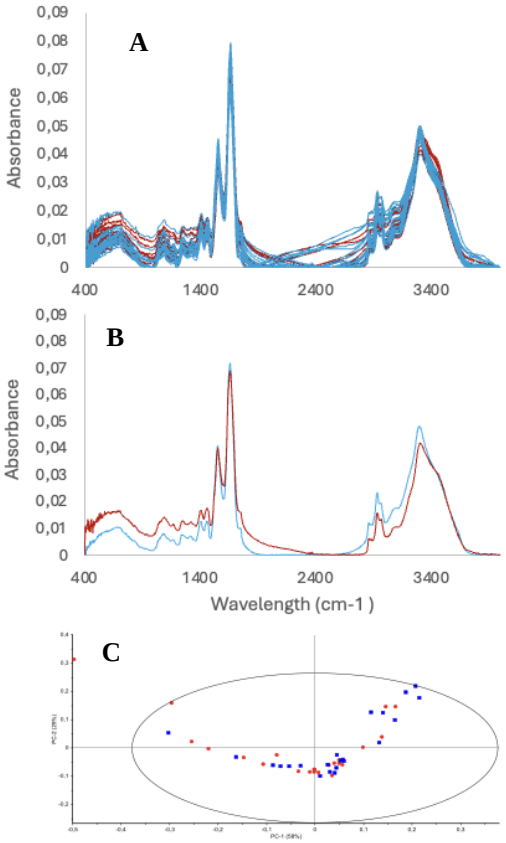

Fig. 1. Serum spectra after baseline correction and normalization, of patients, obtained 3 days before the patients were either discharged (blue) or deceased (red) from the ICU (A), and its corresponding averaged spectra (B) and PCA (C).

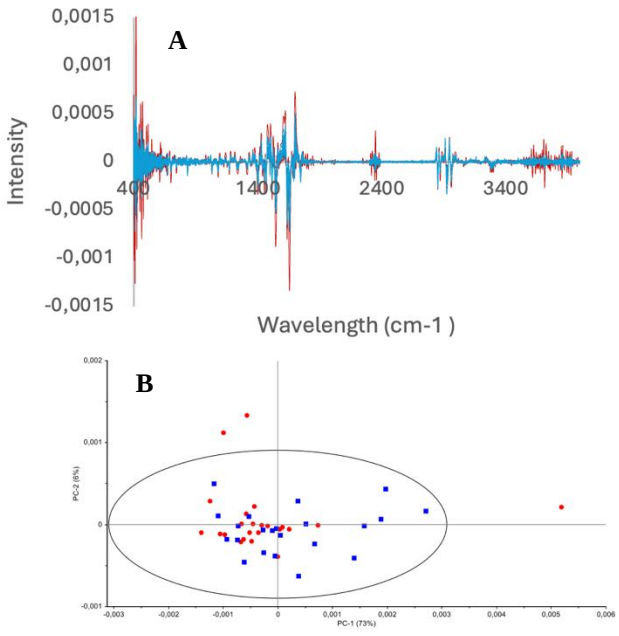

Fig. 2. Second derivative spectra of serum from patients 3 days before the patients were discharged (blue) or deceased (red) (A), and its PCA (B).

TABLE II. MEDIAN AND INTER-QUARTILE RANGE (IQR) OF BANDS OBTAINED FROM NORMALIZED AND BASELINE CORRECTED SPECTRA, AND THE P-VALUE WHEN COMPARING THESE BANDS BETWEEN THE PATIENT'S GROUPS THAT DIED OR WERE DISCHARGED FROM THE ICU. ONLY BANDS THAT PRESENTED P-VALUE <0.05 ARE PRESENTED.

	Discharged		Deceased		
Bands (cm^{-1})	**Median**	**Interquartile range**	**Median**	**Interquartile range**	***p*-Value**
704	1.06×10^{-2}	2.93×10^{-3}	1.17×10^{-2}	4.51×10^{-3}	0.028
2931	2.01×10^{-2}	4.81×10^{-3}	1.70×10^{-2}	5.23×10^{-3}	0.013
2964	1.73×10^{-2}	$4,39 \times 10^{-3}$	1.55×10^{-2}	3.36×10^{-3}	0.019
1244/1083	9.83×10^{-1}	1.93×10^{-1}	9.02×10^{-1}	1.32×10^{-1}	0.016
1552/1083	5.38	1.74	4.52	2.07	0.043
640/1244	1.30	1.83×10^{-1}	1.43	3.48×10^{-1}	0.023
704/1244	1.39	2.16×10^{-1}	1.56	3.48×10^{-1}	0.043
1013/1244	5.73×10^{-1}	2.47×10^{-1}	$6,85 \times 10^{-1}$	1.91×10^{-1}	0.021
1042/1244	7.68×10^{-1}	1.79×10^{-1}	8.96×10^{-1}	2.14×10^{-1}	0.003
1083/1244	1.04	1.93×10^{-1}	1.11	1.58×10^{-1}	0.016
1321/1244	9.31×10^{-1}	6.96×10^{-2}	1.00	1.80×10^{-1}	**<0.001**
640/1552	2.34×10^{-1}	3.37×10^{-2}	2.79×10^{-1}	1.01×10^{-1}	0.006
704/1552	2.54×10^{-1}	4.48×10^{-2}	2.79×10^{-1}	8.53×10^{-2}	0.004
1321/1552	1.71×10^{-1}	6.53×10^{-2}	1.98×10^{-1}	7.01×10^{-2}	0.026
1407/1552	3.04×10^{-1}	6.68×10^{-2}	3.43×10^{-1}	8.79×10^{-2}	0.005
1460/1552	2.80×10^{-1}	6.79×10^{-2}	3.24×10^{-1}	1.06×10^{-1}	0.017
2931/1658	2.65×10^{-1}	9.85×10^{-2}	2.33×10^{-1}	4.15×10^{-2}	0.028

TABLE III. MEDIAN AND INTER-QUARTILE RANGE (IQR) OF BANDS OBTAINED FROM SECOND DERIVATIVE SERUM SPECTRA, AND THE P-VALUE WHEN COMPARING THESE BANDS BETWEEN THE PATIENT'S GROUPS THAT WERE DISCHARGED OR DECEASED IN THE ICU. ONLY BANDS THAT PRESENTED P-VALUE <0.05 ARE PRESENTED.

	Discharged		Deceased		
Bands (cm^{-1})	**Median**	**Interquartile range**	**Median**	**Interquartile range**	***p*-Value**
662	1.45×10^{-5}	8.86×10^{-6}	-5.32×10^{-6}	1.95×10^{-5}	**<0.001**
675	1.83×10^{-5}	1.92×10^{-5}	5.06×10^{-6}	2.26×10^{-5}	0.048
839	-1.82×10^{-5}	7.23×10^{-6}	-3.60×10^{-5}	3.40×10^{-5}	0.016
1131	-2.32×10^{-6}	4.40×10^{-6}	-1.25×10^{-5}	1.64×10^{-5}	0.019
1357	1.80×10^{-5}	1.15×10^{-5}	9.21×10^{-6}	8.86×10^{-6}	0.004
1368	5.35×10^{-6}	6.62×10^{-6}	-9.05×10^{-7}	1.19×10^{-5}	0.011
2784	5.76×10^{-7}	1.99×10^{-6}	-3.05×10^{-6}	9.54×10^{-6}	0.014
2855	-1.11×10^{-4}	2.62×10^{-5}	-9.02×10^{-5}	6.86×10^{-5}	0.022
2927	-1.15×10^{-4}	1.62×10^{-5}	-8.73×10^{-5}	5.39×10^{-5}	0.014
3158	4.47×10^{-6}	3.36×10^{-6}	4.91×10^{-7}	5.93×10^{-6}	0.028
1412/675	5.76×10^{-2}	6.74×10^{-1}	8.37×10^{-1}	3.62	0.010
662/1402	-1.56×10^{-1}	1.15×10^{-1}	6.24×10^{-2}	2.38×10^{-1}	**<0.001**
839/1402	1.86×10^{-1}	1.35×10^{-1}	4.18×10^{-1}	3.61×10^{-1}	0.007
1131/1402	2.07×10^{-2}	5.95×10^{-2}	9.57×10^{-2}	1.47×10^{-1}	0.013
1317/1402	4.07×10^{-1}	7.59×10^{-2}	3.61×10^{-1}	1.15×10^{-5}	0.041
1357/1402	-1.82×10^{-1}	$5,09 \times 10^{-2}$	-1.00×10^{-1}	9.99×10^{-2}	0.008
1368/1402	-8.05×10^{-2}	9.73×10^{-2}	9.85×10^{-3}	9.99×10^{-2}	0.008
662/1469	-1.40×10^{-1}	1.08×10^{-1}	6.09×10^{-2}	2.03×10^{-1}	**<0.001**
839/1469	1.55×10^{-1}	1.51×10^{-1}	3.70×10^{-1}	3.42×10^{-1}	0.002
1131/1469	1.85×10^{-2}	4.83×10^{-2}	8.47×10^{-2}	1.51×10^{-1}	0.013
1368/1469	-6.59×10^{-2}	8.17×10^{-2}	1.36×10^{-2}	1.37×10^{-1}	0.013
662/1543	-7.64×10^{-2}	5.70×10^{-2}	3.07×10^{-2}	1.25×10^{-1}	**<0.001**
703/1543	2.53×10^{-1}	3.52×10^{-2}	3.29×10^{-1}	1.76×10^{-1}	0.031
839/1543	8.09×10^{-2}	9.55×10^{-2}	2.32×10^{-1}	2.30×10^{-1}	0.004
1131/1543	1.07×10^{-2}	3.71×10^{-2}	4.39×10^{-2}	9.02×10^{-2}	0.013
1368/1543	-4.17×10^{-2}	5.16×10^{-2}	6.98×10^{-3}	6.79×10^{-2}	0.012
662/1657	-4.51×10^{-2}	3.57×10^{-2}	1.71×10^{-2}	7.19×10^{-2}	**<0.001**
703/1657	1.29×10^{-1}	1.96×10^{-2}	1.65×10^{-1}	7.49×10^{-2}	0.016
839/1657	4.33×10^{-2}	3.76×10^{-2}	1.23×10^{-1}	1.28×10^{-1}	0.004
1131/1657	5.81×10^{-3}	1.73×10^{-2}	2.57×10^{-2}	5.11×10^{-2}	0.017
1368/1657	-2.19×10^{-2}	2.77×10^{-2}	4.74×10^{-3}	3.51×10^{-2}	0.015
662/1744	-2.13×10^{-1}	1.33×10^{-1}	1.78×10^{-1}	5.19×10^{-1}	**<0.001**
1368/1744	-1.11×10^{-1}	1.34×10^{-1}	5.49×10^{-2}	3.83×10^{-1}	0.006
2784/2855	-5.01×10^{-3}	2.31×10^{-2}	3.57×10^{-2}	3.83×10^{-1}	0.012
3158/2855	-4.04×10^{-2}	4.75×10^{-2}	-4.12×10^{-3}	8.96×10^{-2}	0.041
2784/2927	-4.55×10^{-3}	2.01×10^{-2}	3.59×10^{-2}	1.02×10^{-1}	0.009
2962/2927	6.77×10^{-1}	6.99×10^{-2}	7.39×10^{-1}	2.64×10^{-1}	0.037
3158/2927	-3.85×10^{-2}	4.49×10^{-2}	-4.25×10^{-3}	7.12×10^{-2}	0.037
2784/2962	-7.04×10^{-3}	2.88×10^{-2}	3.92×10^{-2}	1.25×10^{-1}	0.011
2927/2962	1.48	2.21×10^{-1}	1.35	4.36×10^{-1}	0.037
3158/2962	-5.83×10^{-2}	6.64×10^{-2}	-9.27×10^{-3}	1.02×10^{-1}	0.018

Due to these promising results, it was also developed SVM, PCA-LDA and Naïve Bayes predicting models of the patients survival (Table IV). As expected, due to the modest-sized dataset, the Naïve Bayes (Table V) generated the highest accuracy (0.66), with an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 0.77 for the second derivative spectra (Fig. 3).

TABLE IV. ACCURACY OF THE DEVELOPED PREDICTING MODELS, ACCORDING TO THE SPECTRA PRE-PROCESSING METHOD.

Preprocessing	**Method**	**Others**	**Accuracy**
BC+UVN	SVM		0.46
	PCA+LDA	3 PCA's	0.66
	NB		0.59
2D	SVM		0.46
	PCA+LDA	3 PCA's	0.60
	NB		0.66

TABLE V. OUTPUTS OF THE NAÏVE BAYES PREDICTING MODEL, ACCORDING TO THE SPECTRA PRE-PROCESSING METHOD.

Pre processing	**AUC**	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
BC+UN	0.61	0.59	0.60	0.65	0.52
2D	0.77	0.66	0.70	0.61	0.71

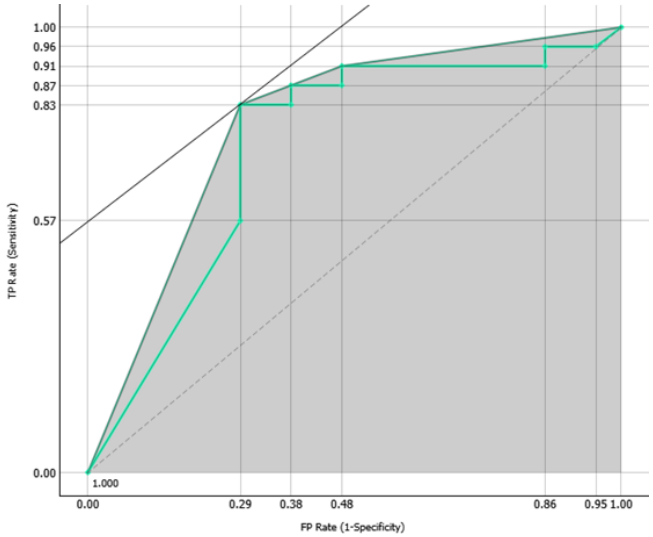

Fig. 3. AUC-ROC for the Naïve Bayes model, based on the full dataset of 45 patients, for death prognosis, 3 days before its occurrence at an ICU.

The present work points, therefore, that the serum molecular profile, captured the metabolic fingerprint associated to the patients' pathophysiological status, including the survival prediction. Since the FTIR spectra of serum is acquired in a simple, economic, and rapid mode, the method presents the potential to be a cost-effective methodology to predict critically ill patients survival.

ACKNOWLEDGMENT

This work was supported by the R-DICIP project (reference IPL/IDI&CA2024/R-DICIP_ISEL) granted by Instituto Politécnico de Lisboa (Portugal) and PREMO (reference DSAIPA/DS/0117/2020) granted by the *Fundação para a Ciência e a Tecnologia* (Portugal) (FCT). C. Rekowski, T. Fonseca, and R. Araújo, acknowledge the PhD grants from FCT, 2023.01951/BD, 2024.02043.BD and 2021.05553.BD, respectively.

REFERENCES

[1] W. A. Knaus, J. E. Zimmerman, D. P. Wagner, E. A. Draper, and D. E. Lawrence, "APACHE-acute physiology and chronic health evaluation: a physiologically based classification system.", Crit. Care Med., vol. 9, no. 8, pp. 591–597, 1981, doi: 10.1097/00003246-198108000-00008. https://doi.org/10.1097/00003246-198108000-00008

[2] R. Araújo, L. Ramalhete, E. Ribeiro, C.R.C. Calado, "Plasma versus Serum Analysis by FTIR Spectroscopy to Capture the Human Physiological State", BioTech, vol.11, no.4, p.56, Dec. 2022, doi:10.3390/biotech11040056. https://doi.org/10.3390/biotech11040056

[3] R. Araújo, L. Ramalhete, H. Paz, C. Ladeira, C.R.C. Calado, "A new method to predict genotoxic effects based on serum molecular profile", Spectrochim. Acta A Mol. Biomol. Spectrosc., 2021Jul 5; 255, pp 119680. doi:10.1016/j.saa.2021.119680. https://doi.org/10.1016/j.saa.2021.119680

[4] J. Demšar, A. Erjavec, T. Hoˇcevar, M. Milutinoviˇc, M. Možina, M. Toplak, L. Umek, J. Zbontar, B. Zupan, "Orange: Data Mining Toolbox in Python", J. Mach. Learn. Res. 14, pp. 2349-53, 2013.

[5] Bellisola, G., & Sorio, C. "Infrared spectroscopy and microscopy in cancer research and diagnosis." *American Journal of Cancer Research*, vol. 2, no. 1, pp. 1-21, 2012.

I Correia holds a bachelor's degree in chemical and biological engineering (2023) from ISEL. She is currently a master's student in biomedical engineering and working on the FCT-approved PREMO project, with the ambition to pursue a PhD in biomedical engineering. https://orcid.org/0009-0007-5085-2542

R Araújo holds several engineering degrees (Mechanical, Biomedical). A researcher at the Laboratory of Engineering and Health (ISEL) since 2017, focusing on biomarkers through metabolomics and proteomics for applications in toxicology, pharmacology, and more. His current PhD research, part of an FCT-approved project, studies COVID-19 biomarkers in critically ill patients using machine learning. https://orcid.org/0000-0002-9369-6486

T Fonseca holds a bachelor's degree in health sciences (2019) and a master's in biomedical engineering (2022) from ISEL. He is currently pursuing a Ph.D. in Biomedicine at Nova Medical School. Since 2020, he has been a research fellow at ISEL, working on the FCT-approved PREMO project. His research focuses on metabolomics and FTIR analysis, using blood serum samples from ICU patients. His primary goal is to identify biomarkers that can predict patient outcomes in critical care units, hence improving predictive models for morbidity and mortality. https://orcid.org/0000-0003-0741-2211

C Von Rekowski holds a degree in Clinical Physiology (2019) from Escola Superior de Tecnologia da Saúde de Lisboa and a master's in biomedical engineering (2022) from ISEL. She has been a research fellow at ISEL since 2020, working on an FCT-approved project focusing on biomarker identification, using metabolomics, proteomics, and machine learning. She is currently pursuing a Ph.D. in Biomedicine at Nova Medical School, while also supporting the work of various master's students. Her work focuses on building databases, performing statistical analysis of demographic, clinical, and laboratory data, and using this data to develop predictive models for morbidity and mortality in critically ill patients. https://orcid.org/0009-0009-6843-1935

L Bento completed his PhD in Clinical Research in 2018, at *Universidade Nova de Lisboa | Faculdade de Ciências Médicas*, and his Medical degree in 1994, at the same institution. He is the Coordinator of the Medical Emergency Unit at *Centro Hospitalar Universitário Lisboa Central EPE*, Head of the Intensive Care Medicine Department, and a Consultant in Internal Medicine. He also serves as an Assistant Professor at *Universidade Nova de Lisboa* and is a specialist in Internal Medicine and Intensive Care Medicine. He has worked in areas such as acute-on-chronic liver failure, acute kidney injury, multi-organ failure, hepatic transplant, and hemodynamic instability, with a focus on intensive care and emergency medicine. https://orcid.org/0000-0002-0260-003X

CRC Calado is Coordinating professor with Aggregation at ISEL. She has a BSc in Biochemistry, a MSc and a PhD in Biotechnology and an Aggregation in Biochemistry. She is the coordinator of the BSC in Biomedical Engineering and the R&D Laboratory in Health & Engineering. Her research focuses in Bioprocess Engineering and Development of Platforms for Drugs and Biomarkers Discovery. https://orcid.org/0000-0002-5264-975