



HAL
open science

Advantages of Metabolomics-Based Multivariate Machine Learning to Predict Disease Severity: Example of COVID

Maryne Lepoittevin, Quentin Blancart Remaury, Nicolas Lévêque, Arnaud Thille, Thomas Brunet, Karine Salaun, Mélanie Catroux, Luc Pellerin, Thierry Hauet, Raphael Thuillier

► To cite this version:

Maryne Lepoittevin, Quentin Blancart Remaury, Nicolas Lévêque, Arnaud Thille, Thomas Brunet, et al.. Advantages of Metabolomics-Based Multivariate Machine Learning to Predict Disease Severity: Example of COVID. *International Journal of Molecular Sciences*, 2024, 25 (22), pp.12199. 10.3390/ijms252212199 . hal-04831902

HAL Id: hal-04831902

<https://hal.science/hal-04831902v1>

Submitted on 11 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Communication

Advantages of Metabolomics-Based Multivariate Machine Learning to Predict Disease Severity: Example of COVID

Maryne Lepoittevin^{1,2}, Quentin Blancart Remaury³, Nicolas Lévêque⁴, Arnaud W. Thille⁵ , Thomas Brunet⁶, Karine Salaun⁵, Mélanie Catroux⁷, Luc Pellerin^{1,2,8} , Thierry Hauet^{1,2,8} and Raphael Thuillier^{1,2,8,*}

- ¹ Inserm Unit Ischémie Reperfusion, Métabolisme et Inflammation Stérile en Transplantation (IRMETIST), UMR U1313, F-86073 Poitiers, France; lepoittevin.maryne@gmail.com (M.L.); luc.pellerin@univ-poitiers.fr (L.P.); thierry.hauet@univ-poitiers.fr (T.H.)
- ² Faculty of Medicine and Pharmacy, University of Poitiers, F-86073 Poitiers, France
- ³ UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), University of Poitiers, 4 rue Michel-Brunet, TSA 51106, F-86073 Poitiers cedex 9, France; quentin.blancart.remaury@univ-poitiers.fr
- ⁴ LITEC, CHU de Poitiers, Laboratoire de Virologie et Mycobactériologie, Université de Poitiers, 2 r Milétrie, F-86000 Poitiers, France; nicolas.leveque@chu-poitiers.fr
- ⁵ Intensive Care Medicine Department, CHU Poitiers, F-86021 Poitiers, France; arnaud.thille@chu-poitiers.fr (A.W.T.); karine.salaun@chu-poitiers.fr (K.S.)
- ⁶ Geriatric Medicine Department, CHU Poitiers, F-86021 Poitiers, France; thomas.brunet@chu-poitiers.fr
- ⁷ Internal Medicine and Infectious Disease Department, CHU Poitiers, F-86021 Poitiers, France; melanie.catroux@chu-poitiers.fr
- ⁸ Biochemistry Department, CHU Poitiers, F-86021 Poitiers, France
- * Correspondence: raphael.thuillier@univ-poitiers.fr

Abstract: The COVID-19 outbreak caused saturations of hospitals, highlighting the importance of early patient triage to optimize resource prioritization. Herein, our objective was to test if high definition metabolomics, combined with ML, can improve prognostication and triage performance over standard clinical parameters using COVID infection as an example. Using high resolution mass spectrometry, we obtained metabolomics profiles of patients and combined them with clinical parameters to design machine learning (ML) algorithms predicting severity (herein determined as the need for mechanical ventilation during patient care). A total of 64 PCR-positive COVID patients at the Poitiers CHU were recruited. Clinical and metabolomics investigations were conducted 8 days after the onset of symptoms. We show that standard clinical parameters could predict severity with good performance (AUC of the ROC curve: 0.85), using SpO₂, first respiratory rate, Horowitz quotient and age as the most important variables. However, the performance of the prediction was substantially improved by the use of metabolomics (AUC = 0.92). Our small-scale study demonstrates that metabolomics can improve the performance of diagnosis and prognosis algorithms, and thus be a key player in the future discovery of new biological signals. This technique is easily deployable in the clinic, and combined with machine learning, it can help design the mathematical models needed to advance towards personalized medicine.

Keywords: predictive algorithm; metabolomics; machine learning; COVID-19



Citation: Lepoittevin, M.; Remaury, Q.B.; Lévêque, N.; Thille, A.W.; Brunet, T.; Salaun, K.; Catroux, M.; Pellerin, L.; Hauet, T.; Thuillier, R. Advantages of Metabolomics-Based Multivariate Machine Learning to Predict Disease Severity: Example of COVID. *Int. J. Mol. Sci.* **2024**, *25*, 12199. <https://doi.org/10.3390/ijms252212199>

Academic Editor: Makrina Karaglani

Received: 11 October 2024

Revised: 7 November 2024

Accepted: 8 November 2024

Published: 13 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Emergency services are often saturated, as was the case during the COVID-19 outbreak, with significant consequences for healthcare providers, from resource allocations to prioritization of patients [1]. A necessary step for such triage is correct patient outcome prediction. While recent studies have demonstrated the benefits of newer techniques such as Forced Oscillation Technique (FOT) [2], as well as novel biomarkers such as plasma KL-6 levels combined with chest radiographic severity grade (RSG) [3], the limited number of tools at the clinician's disposal condemn the healthcare staff to react to the symptoms

rather than anticipate them. It is therefore crucial to develop and validate new tools to help design the care plan for each patient [4].

One major hurdle in creating such a tool is the multifaceted nature of the encountered pathologies, the disparity of the symptoms, their apparent severity and the speed at which an apparently mild ailment can progress to severe outcomes.

To apprehend the multifactorial nature of such diseases, researchers have deployed machine learning-based strategies that can easily integrate a variety of clinical parameters. Indeed, such data mining technologies demonstrate an interesting ability of predicting the outcome; for instance, with COVID, measurable benefits for the patients and the healthcare system were shown [5], with several authors demonstrating the prognostic value of blood parameters when risk stratification is optimized by machine learning, combining co-morbidities with blood parameters [6], or only using blood gas parameters which can be obtained rapidly with delocalized biochemistry automata [7]. Usually, ML is capable of producing reliable risk scores for hospitalization or mortality based on typically measured blood parameters [8–10].

On the other hand, ML is sometimes not capable of producing reliable and reproducible scoring algorithms, as shown in a recent systematic review of more than 400 ML models using radiographs and scans in COVID-positive patients [11]. Hence, it is important not to limit the source of data to train the model. One important new source of data is high resolution mass spectrometry to perform metabolomics analysis. This permits the extraction of numerous biological signals from a small volume of sample and thus optimizes valorization of the patient's fluids. Regarding COVID, these studies have identified a wide variety of biomarkers [12], with many studies identifying dysregulation of amino acids, and others focusing on patient lipidomes [13–16]. Indeed, authors have recently shown a good degree of performance from prediction algorithms using metabolomics data from saliva [17,18] or plasma [19,20].

A synthetic literature review is proposed in the Supplementary Materials.

The objective of our study was to test the hypothesis that metabolomics, combined with ML, can improve COVID severity prediction. To this end, we compared the performance of models issued from blood parameters highlighted in the literature with models issued from a dataset combining such parameters with results from a metabolomics screening.

2. Results

2.1. Patients Clinical and Bloodwork Parameters

A total of 69 patients were included in the study, of which 5 withdrew consent (Supplementary Figure S1). No patients were under artificial ventilation at the time of blood sample collection. The severity of COVID infection was determined by the necessity for mechanical ventilation during patient care. Such a parameter is indeed acknowledged in the Institute for Health and Care Excellence guidelines for the use of dexamethasone in COVID-positive patients, a recommendation also formulated by the World Health Organization [21]. The non-severe patients were slightly more numerous than the severe patients (35–29, respectively).

We explored blood parameters highlighted as pertinent for COVID severity prediction by the literature (detailed list and references of the 39 parameters are displayed in Supplementary Table S1). Of these, only 5 were statistically different between the populations, with severe patients surprisingly demonstrating a higher proportion of younger people, a higher respiratory rate and SpO₂. Severe patients also showed significantly lower coagulation time as well as lower serum albumin.

2.2. Predicting COVID Severity with Clinical and Bloodwork Parameters and/or Metabolomics

To test the predictive potential of the 39 clinical parameters, we conducted machine learning-based approaches to generate predictive algorithms. Of all the algorithms tested, Random Forest showed the most relevant performances across all indicators, however with

a large confidence interval. SpO₂, first respiratory rate, Horowitz quotient, age and CRP were the most important variables (Supplementary Table S2, Top, and Figure 1A).

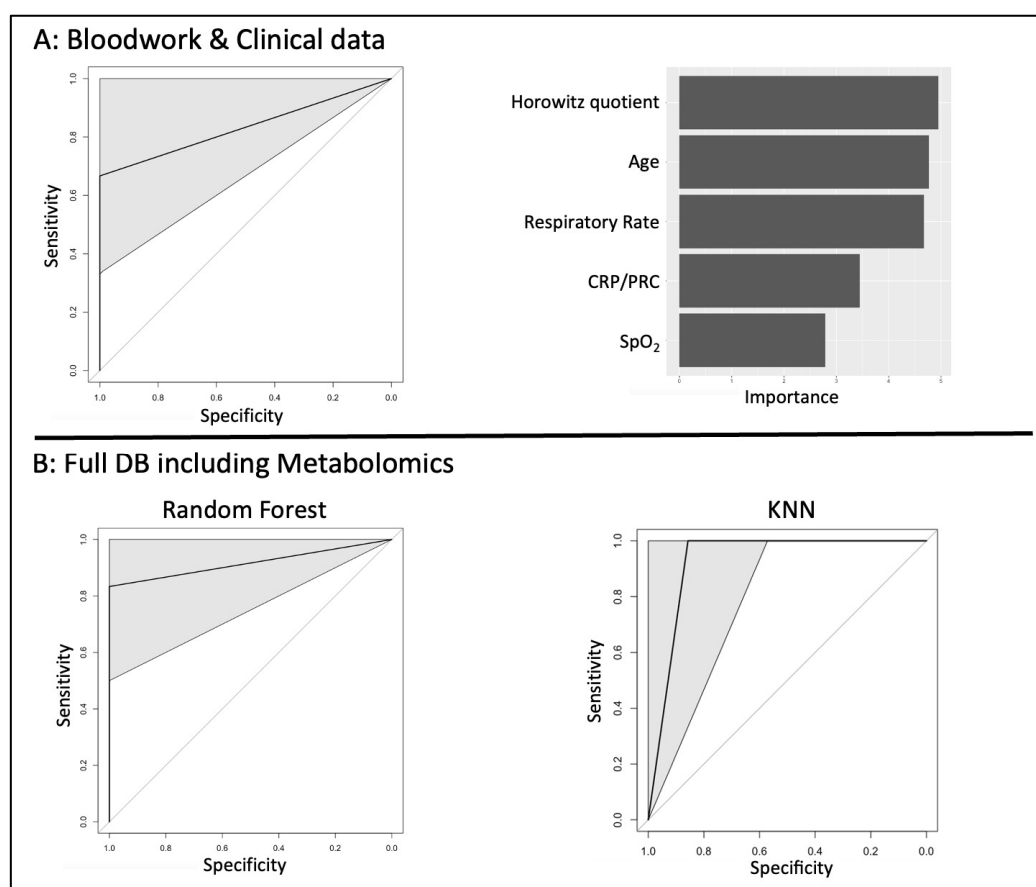


Figure 1. COVID-19 severity predictive algorithm performance. Several machine learning classifiers were used to attempt to build a predictive algorithm for COVID-19 severity 8 days after the onset of symptoms, using two different datasets. Representative ROC curves and main contributing variables are shown. (A) Performance and contributive variables of the Random Forest model built from the database containing bloodwork and clinical data only. (B) Performance and contributive variables of the Random Forest model built from the dataset containing bloodwork, clinical data and metabolomics screening. ROC curves including 95% confidence intervals were drawn using the pROC package and a bootstrapping strategy.

Then, we explored the ability of metabolomics data alone to generate a predictive algorithm, (Table 1, Middle) showing that, on its own, it was unsuccessful in providing a performant discrimination between severe and non-severe patients.

However, combining both clinical and bloodwork parameters with results from the metabolomic study (Table 1, Bottom and Figure 1B), we showed that both Random Forest and KNN had significant performances, with higher AUC of the ROC curves compared to the clinical and bloodwork parameters model, with tighter confidence intervals. Extracting the parameters used for these algorithms revealed that the top ten included only metabolites (Supplementary Table S3).

Hence, metabolomics substantially improved the predictive performance of typical clinical parameters, showing a high potential of complementarity between the techniques.

Table 1. Performance indicators for each of the predictive algorithms built with either the clinical and bloodwork dataset (top) or the full dataset which includes metabolites (middle) or both (bottom).

Clinical and Bloodwork Dataset						
	AUC (95%CI)	Specificity	Sensitivity	Brier Score	Youden's index	[Acc > NIR] p-Value
GLM	0.62 (0.32, 0.86)	0.71	0.5	1.31	0.214	0.3938
RandomForest	0.85 (0.55, 0.98)	1	0.67	1.38	0.67	0.0222
KNN	0.77 (0.46, 0.95)	0.86	0.67	1.46	0.52	0.0798
SVM	0.69 (0.39, 0.91)	0.71	0.67	1.54	0.381	0.2033
C5.0	0.77 (0.46, 0.95)	0.71	0.83	1.77	0.55	0.0798
Metabolomics Alone Dataset						
	AUC (95%CI)	Specificity	Sensitivity	Brier Score	Youden's index	[Acc > NIR] p-Value
GLM	0.7 (0.46, 0.88)	0.81	0.56	1.3	0.374	0.1299
RandomForest	0.75 (0.50, 0.91)	1	0.44	1.05	0.44	0.0553
KNN	0.7 (0.46, 0.88)	0.64	0.78	1.7	0.414	0.1299
SVM	0.8 (0.56, 0.94)	1	0.56	1.2	0.556	0.0189
C5.0	0.7 (0.46, 0.88)	1	0.33	0.9	0.33	0.1299
Full Dataset Which Includes Metabolites						
	AUC (95%CI)	Specificity	Sensitivity	Brier Score	Youden's index	[Acc > NIR] p-Value
GLM	0.77 (0.46, 0.95)	0.86	0.67	1.46	0.52	0.0798
RandomForest	0.92 (0.64, 0.99)	1	0.83	1.62	0.83	0.0039
KNN	0.92 (0.64, 0.99)	0.86	1	1.92	0.857	0.0039
SVM	0.69 (0.39, 0.91)	1	0.64	0.923	0.333	0.9623
C5.0	0.77 (0.46, 0.95)	0.86	0.67	1.46	0.524	0.0798

Performance statistics were calculated with R.

3. Discussion

Herein, we conducted a small-scale clinical study to evaluate the potential of using LC-MS HRMS metabolomics to improve disease prediction. COVID-infection was used as an example of a pathology requiring such prognostication, using the need for mechanical ventilation as severity [21].

Our results show that, indeed, while parameters alone have good predictive performance, combination with metabolomics substantially improves prediction and precision of infection severity in COVID-positive patients.

Considering the later algorithm was based mainly on metabolites, such improvement in the performance in terms of patient discrimination demonstrates the potential of metabolomics regarding decision making at the bedside, confirming recent conclusions in the literature [22].

We tested several classifiers in our ML strategy and differing results demonstrated the wisdom of this approach, suggesting that any ML-based approach should propose several classifiers each representing a different way to analyze data. The robustness of Random Forest [23] in regard to its capability to extract relevant features from numerous variables and a limited number of cases was again demonstrated here as it performed well with all datasets.

Interestingly, the fact that the model used metabolites issued for all four settings of LC/MS (i.e., HILIC and C18 columns with both positive and negative ionization) showed the importance of multiplying the protocols to optimize the output of a metabolomics investigation. Of note, the dataset built only with metabolomics data did not show high levels of performance, highlighting the complementary nature of both sources of data.

Comparing our results on the clinical dataset alone with published investigations (references in Supplementary Table S1 and the Supplementary Synthetic Literature Review), we notice similarities in the variables demonstrating the highest weight in the model. However, other parameters are not represented in our model, such as immune cell balance, blood gas results, etc.; however, this may be explained by the fact that there is a lot of correlation between these parameters (Supplementary Figure S2). To our knowledge, this is the first study regrouping this number of clinical parameters in a single patient cohort and presenting this type of result, which could justify a reduction in the number of biochemical analyses performed in the future.

The level of performance reached by our algorithm (0.92) is higher than the results of other studies (Supplementary Table S1 and the Supplementary Synthetic Literature Review [18,24]) as well as on par with more recent algorithms such as the one combining plasma KL-6 and RSG [3]. It would thus be very interesting to combine these approaches in a larger-scale study.

The important variables of the full dataset models revealed a majority of metabolites, and preliminary attempts at identification showed sensible data in regard to the pathology (Supplementary Table S3). Interestingly, our findings are in line with a recent review which highlights the involvement of ceramides and other lipid mediators in COVID infection [12], although not in totality. This may be due to the differing aims between previous studies and ours: while others have been describing the metabolic profiles in the context of COVID infection [25], we aimed at predicting severity.

Of note, measuring the potential of the top 10 metabolites by PCA to discriminate between the groups (Supplementary Figure S3) demonstrated a good level of performance and the importance of interaction between the metabolites.

Our study remains limited by the sample size, rather small for a machine learning approach. However, in accordance with the TRIPOD guidelines, we showed that the risk of overfitting is low with a decoy strategy (Supplementary Figure S4), highlighting the specific nature of our models towards COVID severity in our patients. Our results are thus preliminary, but our goal was to demonstrate the added value of metabolomics and not necessarily to publish a defined predictive algorithm. An interesting perspective would be to test the complementarity of metabolomics with newer techniques such as FOT [2] or plasma KL-6 levels and RS) [3], which may be the object of a future study on a larger scale. A final limitation of the paper is the use of standard ML algorithms, and indeed, further studies would profit from the use of novel and innovative approaches [26,27]

While an important limitation in metabolomics study is the lack of inter-laboratory unification in terms of techniques and controls, recent efforts have been made towards standardization of the technique [28,29]. Hence, high throughput metabolomics has a very high potential for deployment in the clinic.

4. Materials and Methods

This work was performed following the TRIPOD guidelines (Supplementary Data) [30].

4.1. Patient Population

Patients were included at the Poitiers CHU (France) from 12-2020 to 4-2021 if they were hospitalized, their SARS-CoV-2-positive status was confirmed by RT-PCR, they were adults and free of any tutelage, they were affiliated to social security and gave informed consent. The patients included in the cohort were not vaccinated.

4.2. Judgment Criteria

Patients requiring mechanical ventilation were deemed to have a severe infection (45.31% of included patients, 29/64).

4.3. Clinical Parameters

Clinical parameters were chosen according to the literature (Supplementary Table S1). Biochemical analyses were performed on COBAS Pro automatons (Roche, Meylan, France). Samples were stored at the Poitiers CHU Biological Resources Center (BB-0033-00068).

4.4. Metabolomics Analysis

Procedures were performed in a blinded manner according to previously optimized protocol in which all the technical details are highlighted and the choices in extraction and liquid chromatography parameters are optimized [31]. The data were then pre-processed by the Compound Discoverer 3.3 software. Contributing variable names were imputed from interrogating MetaboLights, Metabolomics Workbench and Human Metabolome datasets [32] with the LC-HRMS signals.

4.5. Statistical Analysis and Machine Learning

To reduce interference from the LC-HRMS data, feature pre-selection was applied only for metabolites appearing to be different between the severe/non-severe groups (evaluated by *t*-test after parameter checks with F-test and Shapiro–Wilk test, *p* threshold set at 0.2) [33]. We used the R software 4.3.3 [34] with the tidyverse environment [35] to analyze the dataset created from clinical parameters and metabolomics screening. Missing data were minimal (1.2%) and were imputed using missRanger.

A standard training/testing set approach was adopted (80/20 split). The artificial intelligence classifiers were chosen according to the literature [36,37], as well as their technology, to test several ways of apprehending the data in order to extract the relevant information. We thus compared the performance of general logarithmic regression (GLM), K-nearest neighbor (KNN), support vector machine (SVM), Random Forest and boosted trees (C5.0). Candidates with the best performance (estimated by shape of the ROC curve, its AUC, the specificity, sensitivity, Brier Score and Youden's Index) on the testing set of the data (i.e., not used for classifier training) were selected. Hyperparameter tuning was performed by 0.632 Bootstrapping. The risk of overfitting was estimated by performing the decoy strategy, in which the outcome variable was scrambled and the whole ML strategy was reperformed. The significance level was set at 0.05.

5. Conclusions

We demonstrate the highly relevant potential of metabolomics in studies aimed at the discovery of new biological signals and their inclusion into novel predictive algorithms. While our study was limited by the number of patients and was thus preliminary, we provided good evidence that combined with the power of machine learning, metabolomics can help design the mathematical models needed to advance towards personalized medicine. Moreover, unlike biological signals issued from other omics [38], the transfer of a metabolomics-based prediction algorithm to the healthcare professional does not present many hurdles and could rapidly improve patient care. Further development of metabolomics should include larger cohorts and more modern ML approaches to fully take advantages of the richness of the data produced.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms252212199/s1>. References [39–53] are cited in the Supplementary Materials.

Author Contributions: The authors declare that all data were generated in-house and that no paper mill was used. M.L.: acquisition, analysis, interpretation of data, drafting the work, final approval; Q.B.R.: analysis of data, revising the work, final approval; N.L.: conception or design of the work, revising the work, final approval; A.W.T.: patient recruitment, sample collection, data acquisition, revising the work, final approval; T.B.: patient recruitment, sample collection, data acquisition, revising the work, final approval; K.S.: patient recruitment, sample collection, data acquisition, revising the work, final approval; M.C.: patient recruitment, sample collection, data acquisition,

revising the work, final approval; L.P.: conception or design of the work, revising the work, final approval; T.H.: conception or design of the work, revising the work, final approval; R.T.: conception of the work, analysis, interpretation of data, drafting and revising the work, final approval. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Région Nouvelle Aquitaine (AMI COVID grant), CHU de Poitiers and University of Poitiers.

Institutional Review Board Statement: This monocentric prospective study included 64 infected patients. Ethics approval was granted by Comité de Protection des Personnes (CPP) Nord Ouest IV on 26 November 2020 (approval number 20.10.05.42632). All participants gave written informed consent before enrolment.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are attached as Supplementary Materials.

Acknowledgments: The authors wish to acknowledge the invaluable technical help from Estelle Lemarie, Virigine Ameteau, Maïté Jacquard, Sonia Brishoual, Nadège Boildieu and Amine Zkim. We also wish to acknowledge the Poitiers CHU CRB, Michelle Grosdenier as well as Stéphanie Ragot and Pierre-Jean Saulnier for their help in the sample logistics as well as Corinne Lorrain of the CHU de Poitiers Research department and the clinical research assistants Carole Grand, Nacira Benhamouche, Aline Murhula, Emilie Bedue and Carole Guignon. Finally we wish to acknowledge the invaluable editorial help of Marie-Christine Duperrier-Thuillier. This research was supported by Inserm, Université de Poitiers, CHU de Poitiers and the Region Nouvelle Aquitaine.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jotwani, R.; Cheung, C.A.; Hoyler, M.M.; Lin, J.Y.; Perlstein, M.D.; Rubin, J.E.; Chan, J.M.; Pryor, K.O.; Brumberger, E.D. Trial under Fire: One New York City Anaesthesiology Residency Programme's Redesign for the COVID-19 Surge. *Br. J. Anaesth.* **2020**, *125*, e386–e388. [[CrossRef](#)] [[PubMed](#)]
2. Pelosi, P.; Tonelli, R.; Torregiani, C.; Baratella, E.; Confalonieri, M.; Battaglini, D.; Marchioni, A.; Confalonieri, P.; Clini, E.; Salton, F.; et al. Different Methods to Improve the Monitoring of Noninvasive Respiratory Support of Patients with Severe Pneumonia/ARDS Due to COVID-19: An Update. *J. Clin. Med.* **2022**, *11*, 1704. [[CrossRef](#)] [[PubMed](#)]
3. Zou, J.; Shi, Y.; Xue, S.; Jiang, H. Use of Serum KL-6 and Chest Radiographic Severity Grade to Predict 28-Day Mortality in COVID-19 Patients with Pneumonia: A Retrospective Cohort Study. *BMC Pulm. Med.* **2024**, *24*, 187. [[CrossRef](#)] [[PubMed](#)]
4. Ji, Y.; Ma, Z.; Peppelenbosch, M.P.; Pan, Q. Potential Association between COVID-19 Mortality and Health-Care Resource Availability. *Lancet Glob. Health* **2020**, *8*, e480. [[CrossRef](#)]
5. Asri, H.; Mousannif, H.; Moatassime, H.A.; Noel, T. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Procedia Comput. Sci.* **2016**, *83*, 1064–1069. [[CrossRef](#)]
6. Alle, S.; Kanakan, A.; Siddiqui, S.; Garg, A.; Karthikeyan, A.; Mehta, P.; Mishra, N.; Chattopadhyay, P.; Devi, P.; Waghdhare, S.; et al. COVID-19 Risk Stratification and Mortality Prediction in Hospitalized Indian Patients: Harnessing Clinical Data for Public Health Benefits. *PLoS ONE* **2022**, *17*, e0264785. [[CrossRef](#)]
7. Huyut, M.T.; Üstündağ, H. Prediction of Diagnosis and Prognosis of COVID-19 Disease by Blood Gas Parameters Using Decision Trees Machine Learning Model: A Retrospective Observational Study. *Med. Gas. Res.* **2022**, *12*, 60–66. [[CrossRef](#)]
8. Khedar, R.S.; Gupta, R.; Sharma, K.; Mittal, K.; Ambaliya, H.C.; Gupta, J.B.; Singh, S.; Sharma, S.; Singh, Y.; Mathur, A. Biomarkers and Outcomes in Hospitalised Patients with COVID-19: A Prospective Registry. *BMJ Open* **2022**, *12*, e067430. [[CrossRef](#)]
9. Willette, A.A.; Willette, S.A.; Wang, Q.; Pappas, C.; Klinedinst, B.S.; Le, S.; Larsen, B.; Pollpeter, A.; Li, T.; Mochel, J.P.; et al. Using Machine Learning to Predict COVID-19 Infection and Severity Risk among 4510 Aged Adults: A UK Biobank Cohort Study. *Sci. Rep.* **2022**, *12*, 7736. [[CrossRef](#)]
10. Brinati, D.; Campagner, A.; Ferrari, D.; Locatelli, M.; Banfi, G.; Cabitza, F. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J. Med. Syst.* **2020**, *44*, 135. [[CrossRef](#)]
11. Roberts, M.; Driggs, D.; Thorpe, M.; Gilbey, J.; Yeung, M.; Ursprung, S.; Aviles-Rivero, A.I.; Etmann, C.; McCague, C.; Beer, L.; et al. Common Pitfalls and Recommendations for Using Machine Learning to Detect and Prognosticate for COVID-19 Using Chest Radiographs and CT Scans. *Nat. Mach. Intell.* **2021**, *3*, 199–217. [[CrossRef](#)]
12. Bruzzzone, C.; Conde, R.; Embade, N.; Mato, J.M.; Millet, O. Metabolomics as a Powerful Tool for Diagnostic, Pronostic and Drug Intervention Analysis in COVID-19. *Front. Mol. Biosci.* **2023**, *10*, 1111482. [[CrossRef](#)] [[PubMed](#)]
13. Hasan, M.R.; Suleiman, M.; Pérez-López, A. Metabolomics in the Diagnosis and Prognosis of COVID-19. *Front. Genet.* **2021**, *12*, 721556. [[CrossRef](#)]

14. Battaglini, D.; Lopes-Pacheco, M.; Castro-Faria-Neto, H.C.; Pelosi, P.; Rocco, P.R.M. Laboratory Biomarkers for Diagnosis and Prognosis in COVID-19. *Front. Immunol.* **2022**, *13*, 857573. [[CrossRef](#)] [[PubMed](#)]
15. Spick, M.; Lewis, H.M.; Wilde, M.J.; Hopley, C.; Huggett, J.; Bailey, M.J. Systematic Review with Meta-Analysis of Diagnostic Test Accuracy for COVID-19 by Mass Spectrometry. *Metabolism* **2022**, *126*, 154922. [[CrossRef](#)]
16. Bourgin, M.; Durand, S.; Kroemer, G. Diagnostic, Prognostic and Mechanistic Biomarkers of COVID-19 Identified by Mass Spectrometric Metabolomics. *Metabolites* **2023**, *13*, 342. [[CrossRef](#)] [[PubMed](#)]
17. Saheb Sharif-Askari, N.; Soares, N.C.; Mohamed, H.A.; Saheb Sharif-Askari, F.; Alsayed, H.A.H.; Al-Hroub, H.; Salameh, L.; Osman, R.S.; Mahboub, B.; Hamid, Q.; et al. Saliva Metabolomic Profile of COVID-19 Patients Associates with Disease Severity. *Metabolomics* **2022**, *18*, 81. [[CrossRef](#)]
18. Frampas, C.F.; Longman, K.; Spick, M.; Lewis, H.-M.; Costa, C.D.S.; Stewart, A.; Dunn-Walters, D.; Greener, D.; Evetts, G.; Skene, D.J.; et al. Untargeted Saliva Metabolomics by Liquid Chromatography—Mass Spectrometry Reveals Markers of COVID-19 Severity. *PLoS ONE* **2022**, *17*, e0274967. [[CrossRef](#)]
19. Ceperuelo-Mallafré, V.; Reverté, L.; Peraire, J.; Madeira, A.; Maymó-Masip, E.; López-Dupla, M.; Gutierrez-Valencia, A.; Ruiz-Mateos, E.; Buzón, M.J.; Jorba, R.; et al. Circulating Pyruvate Is a Potent Prognostic Marker for Critical COVID-19 Outcomes. *Front. Immunol.* **2022**, *13*, 912579. [[CrossRef](#)]
20. Occelli, C.; Guignon, J.-M.; Lindenthal, S.; Cagnard, A.; Graslin, F.; Brglez, V.; Seitz-Polski, B.; Dellamonica, J.; Levraut, J.; Pourcher, T. Untargeted Plasma Metabolomic Fingerprinting Highlights Several Biomarkers for the Diagnosis and Prognosis of Coronavirus Disease 19. *Front. Med.* **2022**, *9*, 995069. [[CrossRef](#)]
21. Coronavirus Disease (COVID-19): Dexamethasone. Available online: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-dexamethasone> (accessed on 16 January 2023).
22. Di Minno, A.; Gelzo, M.; Caterino, M.; Costanzo, M.; Ruoppolo, M.; Castaldo, G. Challenges in Metabolomics-Based Tests, Biomarkers Revealed by Metabolomic Analysis, and the Promise of the Application of Metabolomics in Precision Medicine. *Int. J. Mol. Sci.* **2022**, *23*, 5213. [[CrossRef](#)] [[PubMed](#)]
23. Sage, A.J. Random Forest Robustness, Variable Importance, and Tree Aggregation. Ph.D. Thesis, Iowa State University, Ames, IO, USA, 2018.
24. Bordbar, M.M.; Samadinia, H.; Sheini, A.; Aboonajmi, J.; Hashemi, P.; Khoshafar, H.; Halabian, R.; Khanmohammadi, A.; Gh, B.F.N.M.; Sharghi, H.; et al. Visual Diagnosis of COVID-19 Disease Based on Serum Metabolites Using a Paper-Based Electronic Tongue. *Anal. Chim. Acta* **2022**, *1226*, 340286. [[CrossRef](#)] [[PubMed](#)]
25. Ghini, V.; Meoni, G.; Pelagatti, L.; Celli, T.; Veneziani, F.; Petrucci, F.; Vannucchi, V.; Bertini, L.; Luchinat, C.; Landini, G.; et al. Profiling Metabolites and Lipoproteins in COMETA, an Italian Cohort of COVID-19 Patients. *PLoS Pathog.* **2022**, *18*, e1010443. [[CrossRef](#)] [[PubMed](#)]
26. Mirmozaffari, M.; Yazdani, M.; Boskabadi, A.; Ahady Dolatsara, H.; Kabirifar, K.; Amiri Golilarz, N. A Novel Machine Learning Approach Combined with Optimization Models for Eco-Efficiency Evaluation. *Appl. Sci.* **2020**, *10*, 5210. [[CrossRef](#)]
27. Mirmozaffari, M.; Yazdani, R.; Shadkam, E.; Khalili, S.M.; Tavassoli, L.S.; Boskabadi, A. A Novel Hybrid Parametric and Non-Parametric Optimisation Model for Average Technical Efficiency Assessment in Public Hospitals during and Post-COVID-19 Pandemic. *Bioengineering* **2021**, *9*, 7. [[CrossRef](#)]
28. Thompson, J.W.; Adams, K.J.; Adamski, J.; Asad, Y.; Borts, D.; Bowden, J.A.; Byram, G.; Dang, V.; Dunn, W.B.; Fernandez, F.; et al. International Ring Trial of a High Resolution Targeted Metabolomics and Lipidomics Platform for Serum and Plasma Analysis. *Anal. Chem.* **2019**, *91*, 14407–14416. [[CrossRef](#)]
29. Siskos, A.P.; Jain, P.; Römisch-Margl, W.; Bennett, M.; Achaintre, D.; Asad, Y.; Marney, L.; Richardson, L.; Koulman, A.; Griffin, J.L.; et al. Interlaboratory Reproducibility of a Targeted Metabolomics Platform for Analysis of Human Serum and Plasma. *Anal. Chem.* **2017**, *89*, 656–665. [[CrossRef](#)]
30. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med.* **2015**, *13*, 1. [[CrossRef](#)]
31. Lepoittevin, M.; Blancart-Remaury, Q.; Kerforne, T.; Pellerin, L.; Hauet, T.; Thuillier, R. Comparison between 5 Extractions Methods in Either Plasma or Serum to Determine the Optimal Extraction and Matrix Combination for Human Metabolomics. *Cell Mol. Biol. Lett.* **2023**, *28*, 43. [[CrossRef](#)]
32. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [[CrossRef](#)]
33. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
34. R Core Team. *R: A Language and Environment for Statistical Computing*; MSOR Connections; R Foundation for Statistical Computing: Vienna, Austria, 2014.
35. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [[CrossRef](#)]
36. Wolpert, D.H.; Macready, W.G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [[CrossRef](#)]
37. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
38. Lepoittevin, M.; Kerforne, T.; Pellerin, L.; Hauet, T.; Thuillier, R. Molecular Markers of Kidney Transplantation Outcome: Current Omics Tools and Future Developments. *Int. J. Mol. Sci.* **2022**, *23*, 6318. [[CrossRef](#)]

39. Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Al-Madeed, S.; Zughair, S.M.; Doi, S.A.R.; Hassen, H.; Islam, M.T. An Early Warning Tool for Predicting Mortality Risk of COVID-19 Patients Using Machine Learning. *Cogn. Comput.* **2021**, *16*, 1778–1793. [[CrossRef](#)]
40. Weng, Z.; Chen, Q.; Li, S.; Li, H.; Zhang, Q.; Lu, S.; Wu, L.; Xiong, L.; Mi, B.; Liu, D.; et al. ANDC: An early warning score to predict mortality risk for patients with Coronavirus Disease. *J. Transl. Med.* **2020**, *18*, 1–10. [[CrossRef](#)]
41. Chen, R.; Liang, W.; Jiang, M.; Guan, W.; Zhan, C.; Wang, T.; Tang, C.; Sang, L.; Liu, J.; Ni, Z.; et al. Risk Factors of Fatal Outcome in Hospitalized Subjects With Coronavirus Disease 2019 From a Nationwide Analysis in China. *Chest* **2020**, *158*, 97–105. [[CrossRef](#)]
42. Kar, S.; Chawla, R.; Haranath, S.P.; Ramasubban, S.; Ramakrishnan, N.; Vaishya, R.; Sibal, A.; Reddy, S. Multivariable mortality risk prediction using machine learning for COVID-19 patients at admission (AICOVID). *Sci. Rep.* **2021**, *11*, 1–11. [[CrossRef](#)]
43. Subudhi, S.; Verma, A.; Patel, A.B.; Hardin, C.C.; Khandekar, M.J.; Lee, H.; McEvoy, D.; Stylianopoulos, T.; Munn, L.L.; Dutta, S.; et al. Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. *npj Digit. Med.* **2021**, *4*, 1–7. [[CrossRef](#)]
44. Mahdavi, M.; Choubdar, H.; Zabeh, E.; Rieder, M.; Safavi-Naeini, S.; Jobbagy, Z.; Ghorbani, A.; Abedini, A.; Kiani, A.; Khanlarzadeh, V.; et al. A machine learning based exploration of COVID-19 mortality risk. *PLOS ONE* **2021**, *16*, e0252384. [[CrossRef](#)] [[PubMed](#)]
45. Hao, B.; Sotudian, S.; Wang, T.; Xu, T.; Hu, Y.; Gaitanidis, A.; Breen, K.; Velmahos, G.C.; Paschalidis, I.C.; Information, C.F.; et al. Early prediction of level-of-care requirements in patients with COVID-19. *eLife* **2020**, *9*. [[CrossRef](#)]
46. Ji, D.; Zhang, D.; Xu, J.; Chen, Z.; Yang, T.; Zhao, P.; Chen, G.; Cheng, G.; Wang, Y.; Bi, J.; et al. Prediction for Progression Risk in Patients With COVID-19 Pneumonia: The CALL Score. *Clin. Infect. Dis.* **2020**, *71*, 1393–1399. [[CrossRef](#)]
47. Magunia, H.; Lederer, S.; Verbuecheln, R.; Gilot, B.J.; Koeppen, M.; Haerberle, H.A.; Mirakaj, V.; Hofmann, P.; Marx, G.; Bickenbach, J.; et al. Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. *Crit. Care* **2021**, *25*, 1–14. [[CrossRef](#)]
48. Yan, L.; Zhang, H.-T.; Goncalves, J.; Xiao, Y.; Wang, M.; Guo, Y.; Sun, C.; Tang, X.; Jing, L.; Zhang, M.; et al. An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.* **2020**, *2*, 283–288. [[CrossRef](#)]
49. Liu, J.; Liu, Y.; Xiang, P.; Pu, L.; Xiong, H.; Li, C.; Zhang, M.; Tan, J.; Xu, Y.; Song, R.; et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. *J. Transl. Med.* **2020**, *18*, 206. [[CrossRef](#)] [[PubMed](#)]
50. Liang, W.; Liang, H.; Ou, L.; Chen, B.; Chen, A.; Li, C.; Li, Y.; Guan, W.; Sang, L.; Lu, J.; et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID. *JAMA Intern. Med.* **2020**, *180*, 1081–1089. [[CrossRef](#)]
51. Kishaba, T.; Tamaki, H.; Shimaoka, Y.; Fukuyama, H.; Yamashiro, S. Staging of Acute Exacerbation in Patients with Idiopathic Pulmonary Fibrosis. *Lung* **2013**, *192*, 141–149. [[CrossRef](#)]
52. Lu, J.; Hu, S.; Fan, R.; Liu, Z.; Yin, X.; Wang, Q.; Lv, Q.; Cai, Z.; Li, H.; Hu, Y.; et al. ACP Risk Grade: A Simple Mortality Index for Patients with Confirmed or Suspected Severe Acute Respiratory Syndrome Coronavirus 2 Disease (COVID-19) During the Early Stage of Outbreak in Wuhan, China. Available online: <https://www.medrxiv.org/content/10.1101/2020.02.20.20025510v1> (accessed on 16 January 2023). [[CrossRef](#)]
53. Liu, Y.; Mao, B.; Liang, S.; Yang, J.-W.; Lu, H.-W.; Chai, Y.-H.; Wang, L.; Zhang, L.; Li, Q.-H.; Zhao, L.; et al. Association between age and clinical characteristics and outcomes of COVID. *Eur. Respir. J.* **2020**, *55*, 2001112. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.