Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(4): 426-436

Original Article

CRP, PCT, and D-dimer as Biomarkers for Disease Severity in COVID-19 Patients: A Retrospective Study in Kinshasa, Democratic Republic of Congo

Tasnime Hamdeni[1*], Frederick Tshibasu[2], Asma Kerkeni[3], Soufiane Gasmi[3]

[1]Statistics' Departement, Higher School of Statistics and Information Analysis (ESSAI), University of Carthage, Tunis, Tunisia.
[2]Division of Diagnostic Imaging, University Hospital of Kinshasa, School of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo.
[3]Department of Mathematics, National Higher Engineering School of Tunis (ENSIT), University of Tunis, Tunis, Tunisia.

ARTICLE INFO

Received 12.04.2023
Revised 28.05.2023
Accepted 04.06.2023
Published 15.12.2023

Key words:
COVID-19;
Disease severity;
biomarkers;
Prediction;
Machine learning

ABSTRACT

Introduction: The COVID-19 pandemic has had a significant impact on global health, resulting in more than 6 million reported deaths worldwide as of April 2023. This study aimed to investigate the potential of C-reactive protein (CRP), procalcitonin (PCT), and D-dimer as biomarkers for assessing disease severity in COVID-19 patients in Kinshasa, Democratic Republic of Congo.
Methods: A retrospective examination was conducted involving 339 COVID-19 patients admitted to Kinshasa hospitals between January 2021 and March 2022. CRP, PCT, and D-dimer levels were measured in all patients and compared between those with severe and non-severe illnesses.
Results: Our findings revealed significantly higher CRP, PCT, and D-dimer levels in severe cases compared to non-severe cases. Specifically, the median CRP level was 120.6 mg/L in severe cases, 47.3 mg/L in mild cases, and 13.5 mg/L in moderate cases. The median PCT levels were 0.26 ng/mL in severe cases, 0.08 ng/mL in mild cases, and 0.07 ng/L in moderate cases. Additionally, the median D-dimer level was 1836.9 µg/L in severe cases and 597.6 µg/L in mild cases, with a value of 481.1 µg/L in moderate cases. System learning techniques were also employed to predict disease severity based on these biomarkers, achieving high accuracy.
Conclusion: Our findings suggest that CRP, PCT, and D-dimer serve as valuable biomarkers for identifying severe COVID-19 cases in Kinshasa. Furthermore, the application of machine learning methods can yield accurate predictions of disease severity based on these biomarkers. These biomarkers hold the potential to assist clinicians in informed decision-making regarding patient management and contribute to improved clinical outcomes for COVID-19 patients.

Introduction

The COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become a major global health crisis, affecting millions of people worldwide. The clinical manifestations of COVID-19 vary from asymptomatic or mild respiratory symptoms to severe acute respiratory distress syndrome (ARDS) and multiple organ

*.Corresponding Author: hamdeni.tasnime@gmail.com

failure, which can result in mortality.[1] Early identification of disease severity is essential for appropriate patient management and to prevent adverse outcomes.[2,3]

Biomarkers such as C-reactive protein (CRP), procalcitonin (PCT), and D-dimer have been investigated as potential indicators of disease severity in COVID-19 patients.[4-6] CRP is an acute-phase reactant that increases rapidly in response to inflammation, while PCT is a marker of bacterial infection and inflammation. D-dimer is a fibrin degradation product that reflects the presence of hypercoagulability and thrombosis. Several studies have reported the potential of these biomarkers in predicting disease severity and clinical outcomes in COVID-19 patients.[7-9]

The threshold for a positive D-dimer result can vary depending on the laboratory and the specific assay used for testing. However, in general, a D-dimer level greater than 500 ng/mL is considered to be significantly elevated.[10]

The threshold for PCT can vary depending on the clinical context and the laboratory where the test is performed. Commonly, PCT is used as a biomarker to help diagnose bacterial infections and guide antibiotic therapy. In a general context, a PCT level below 0.1 ng/mL is considered to be a low risk for bacterial infection, while levels above 0.5 ng/mL are more suggestive of bacterial infection.[11]

The threshold for a positive CRP result can vary depending on the laboratory and the specific assay used for testing. However, in general, a CRP level greater than 10 mg/L is significantly elevated and may indicate a serious underlying condition. Lower levels, such as between 3-10 mg/L, may also be considered positive in certain clinical contexts, but the interpretation of CRP results should always be done while taking into account the patient's medical history.[12-14] It should be noted that the study population exhibits relatively high CRP levels, which may be attributed to the fact that 55% of the subjects have comorbidities.

Recent studies have suggested that elevated levels of CRP, PCT, and D-dimer are associated with increased disease severity, need for hospitalization, and mortality in COVID-19 patients. For example, Liu et al. (2020) reported that PCT levels were significantly higher in severe cases compared to mild cases.[15] Moreover, Zhou et al. (2020) found that D-dimer levels were significantly elevated in non-survivors compared to survivors of COVID-19.[16,17]

In this retrospective study, we aim to investigate the utility of CRP, PCT, and D-dimer as biomarkers for disease severity in COVID-19 patients. We will examine the levels of these biomarkers in patients with mild (stage 1), moderate (stage 2), and severe disease (stage 3 and 4). The utilized data set has been collected by the Department of Radiodiagnosis and Medical Imaging, University Clinics of Kinshasa, Kinshasa, Democratic Republic of Congo.

Overall, this study will contribute to the growing body of evidence on the potential of CRP, PCT, and D-dimer as biomarkers for disease severity in COVID-19 patients and may inform clinical decision-making and patient management strategies.

The paper is organized as follows: Section II, reserved for Materials & Methods, includes descriptions of the patient data, the imaging technology used, the data collection protocol, and the initial handling of the data. Section III focuses on the analytical methods, machine learning model configurations, and statistical

evaluations performed on the collected data ensuring transparency and facilitating easier comprehension of the Results in section IV. We discuss the implications and limitations of our findings in the subsequent Section V. Finally, we conclude the paper with a summary of our main contributions and directions for future research in section VI.

Methods

The utilized primary database provides information about COVID-19 patients of the University Clinics of Kinshasa, in the Democratic Republic of Congo. We have conducted a retrospective study of 339 COVID-19 patients admitted to the hospitals of Kinshasa between January 2021 and March 2022. Informed consent was obtained from all participants. The consent process involved participants being informed about the study's aims, procedures, potential risks, and their rights, including the right to withdraw from the study at any time without consequence. The study was approved by the ethics committee of the University of Kinshasa.

To ensure a focused and relevant study population, we established inclusion and exclusion criteria. Inclusion criteria included individuals aged 18 years and over. Exclusion criteria involved patients with incomplete medical records, pregnant women, and patients receiving treatment for cancer, as these conditions could complicate the analysis.

We conducted a retrospective analysis of chest CT images of 339 patients hospitalized for SARS COV-2 pneumonitis at the Diamant medical center in Kinshasa, Democratic Republic of Congo, during the second, third and fourth waves of the COVID-19 pandemic in Kinshasa.

SARS-CoV-2 infection was confirmed by RT-PCR from upper respiratory tract specimens. All patients were managed in accordance with the national management protocol issued by the National Technical Secretariat for the Response to COVID-19 in the DRC. We carried out this study over a period of 13 months, from January 2021 to March 2022. All patients were diagnosed and treated in accordance with the national management protocol disseminated by the National Technical Secretariat of the Response against Covid-19 in the DRC.

A 64-slice TOSHIBA Aquilion CT scanner, commissioned in 2011, was used in all our patients. As acquisition parameters, we used: 120 kV, 100 to 150 mA, 0.6 mm collimation and 1: 1 pitch. The thoracic field of exploration covered the pulmonary apex to the diaphragm on the axial plane taken in free breathing with patients in dorsal recumbency. Native CT images were obtained with 2.5 mm slices then reconstructed with 1.25 mm collimation using a standard algorithm, then sent to the Picture Archiving and Communication System (PACS) for analysis. CT images were evaluated using a lung window with a windowing level of -600 HU and a width of 1500 HU. The soft tissue window level was 40 HU and a width of 300 HU. As for the injected examinations, a dose of iodinated contrast medium of 1 to 1.5 ml/kg body weight was used after obtaining a renal function report (urea and creatinine) and hydrating the patient in any case of suspected pulmonary embolism. All images were stored in PACS and reviewed by three experienced thoracic radiologists.

The following thoracic radiological patterns, suggestive of the diagnosis of COVID-19, were reviewed: (a) ground-glass opacities:

area of pulmonary parenchymal overdensity not obliterating the pulmonary vessels, (b) Crazy paving: appearance of intra- and inter-lobular reticulations within the ground-glass opacities, (c) parenchymal condensations: zone of systematized or non-systematized lung parenchymal overdensity, obliterating the pulmonary vessels, (d) atypical covid-19 lesions: lesions other than subpleural ground-glass opacities, Crazy paving and systematized condensations (e) lesion severity was defined by the extent of the lesions; the latter enables grading of the involvement into : Covid-19 minimal: extent of lesion <10%, Covid-19 moderate: extent of lesion 10-25%, Covid-19 extensive: extent of lesion 25-50%, Covid-19 severe: extent of lesion 50-75% and Covid-19 critical: extent of lesion >75%; (f) progression of pulmonary involvement was staged as follows: Early stage : up to the first 4 days, intermediate stage: 5 to 8 days, late stage: 8 to 13 days, very late stage: beyond 14 days; (g) pulmonary complications were defined as the occurrence of pulmonary embolism and/ or acute respiratory distress syndrome; and (h) associated pathologies: the presence of pathologies other than the above-mentioned complications. Data Analysis was performed in Pyhon.

In this study, diabetes is defined as a chronic medical condition in which the body either cannot produce enough insulin or cannot effectively use the insulin it produces, leading to elevated levels of glucose in the blood. This definition aligns with the criteria set forth by the World Health Organization.[18]

High blood pressure (hypertension) is identified as a condition where the blood pressure in the arteries is persistently elevated. Hypertension is typically diagnosed when blood pressure readings consistently exceed 140 millimeters of mercury (mmHg) systolic or 90 mmHg diastolic, as per the guidelines of the American Heart Association.[19]

In this study, we primarily focused on Hypertension (HBP) and diabetes due to their prevalent impact in our context. However, recognizing the significance of comprehensive patient histories, we included a category titled ‘other diseases’ to account of various medical antecedents such as obesity, malaria, asthma, gastritis, dyslipidemia. We believe that this approach ensures a holistic view of each patient’s health status, acknowledging the complexity of multiple coexisting conditions.

Statistical Analysis

This section focuses on the analytical methods, model configurations, and statistical evaluations performed on the collected data.

After completing the data cleaning process, we move on to partitioning the features into two distinct variables: a predictive variable, which includes biomarker or comorbidity data along with age, and a target variable, which contains the stage level data. We split then the variables into two distinct phases: a training phase and a testing phase. In the training phase, we utilize a portion of the data, while in the testing phase, we evaluate the model's performance on a separate subset. To ensure a reliable evaluation, it is recommended to choose a test phase size that is less than or equal to 50% of the overall data size.[20] For instance, in our case, we opted for a test phase size of 30%.

The remaining portion of the data is reserved for the training phase, enabling the model to learn patterns and relationships from a substantial dataset. This division into training

and testing phases helps us assess the model's generalization capabilities and provides valuable insights into its performance.

After standardizing the data, we proceed to train the desired machine learning model. In this paper, we test eight different machine learning models configured for optimal performance.

Support Vector Machine (SVM): The configuration used for tuning the SVM model is as follows: Regularization parameter is set for 0.01, the type of kernel used in the algorithm is Linear, and the tolerance parameter for the stopping criterion is set for 0.01.

K-Nearest Neighbor (KNN) with the following settings: the parameter to use for distance calculation is the Minkowski Metric with 7 neighbors, and Uniform Weights.

The decision tree (DT) configured as follows: Gini impurity is used as a criterion to evaluate the quality of a split in a decision tree algorithm, the maximum depth of a tree is set to 1, the maximum number of features is set to the square root of the total number of features in the dataset, which allows the decision tree to explore a diverse range of feature subsets while still maintaining a reasonable level of randomness and avoiding overfitting, and no restriction on the maximum number of leaf nodes.

The Random Forest (RF) tuned with the following parameter settings: a maximum depth of 15 for each tree, a minimum of 3 samples required to split an internal node, a minimum of 3 samples required to be at a leaf node, and a total of 5 trees in the forest.

The Bagging model where the parameter settings are as follows: each decision tree in the Bagging ensemble will be built using a random selection of 3 features from the dataset, each decision tree will be constructed using a random sample of 50 instances from the dataset and the Bagging ensemble will consist of 100 decision trees.

AdaBoost model, which is another ensemble learning method. Here's a summary of the parameter settings: the learning rate is set to 0.01, and the maximum depth allowed for each individual decision tree within the AdaBoost ensemble is set to 3 which should limit the complexity of the weak learners, prevent overfitting and promote generalization, five weak learners are considered.

Gradient Boosting (GBM) model[21] is fine-tuned as follows: the learning rate is set to 0.01, Stagewise Additive Modeling using a Multiclass Exponential loss function (SAMME) is used; an algorithm that updates the weights of the weak learners based on the exponential loss function, 5 weak learners are trained sequentially to minimize the loss function and improve the overall performance of the model.

XGBoost model, which is an optimized implementation of GBM: the learning rate is now set to 10-5. A smaller learning rate typically leads to slower convergence but can improve the model's generalization. The maximum depth is set to 4. The number of boosting rounds is set to 10. Each round focuses on correcting the mistakes made by the previous trees and improves the model's predictive performance. The minimum sum of instance weights needed in a child node to further partition the tree during the tree-building process is configured as 2.

By adjusting these parameters, we have controlled the trade-off between bias and variance, the learning process and the complexity ultimately influencing its performance and generalization ability. The

results and the performance of each model are detailed in Results section 4.

To detect potential overfitting, we have compared the accuracy of the model in the test phase with its performance in the training phase. By calculating the difference between the two accuracy values, we can determine if there is a significant gap. If we observe a substantial difference in accuracy or notice a significant divergence between the curves, it indicates the presence of overfitting in the model. This means that the model has learned the training data too well, resulting in poor generalization to new, unseen data. Detecting overfitting is crucial to ensure the reliability and effectiveness of the machine learning model. K-fold cross-validation is also used to assess the model's performance on multiple subsets of the data. This has also helped to detect overfitting by evaluating the model's generalization ability across different data partitions.

Results

Table 1 provides a summary of the study population's characteristics, including demographics, comorbidities, biomarkers, and disease severity. It gives an overview of the age distribution, gender representation, prevalence of comorbidities such as high blood pressure (HBP) and diabetes, values of biomarkers such as D-dimer, Polymerase Chain Reaction (PCR), Procalcitonin (PCT), and C-Reactive Protein (CRP), as well as the distribution of disease severity stages. The table provides a snapshot of the study population's profile and helps understand the characteristics of the individuals included in the study.

Table 1. Characteristics of the study population overall

	Median (IQR) or N (%)
Demographics	
Age (y)	52 (41-61)
Male	209 (65 %)
Comorbidities	
HBP	89 (26%)
Diabetes	36 (11 %)
Other diseases	61 (18 %)
NTR	153 (45 %)
Biomarkers	
D-dimer	603.07 (310.98 - 1247.97)
PCR- positive	146 (63 %)
PCT	0.08 (0.05 - 0.23)
CRP	27.5 (7.3 - 83.0)
Disease severity	
Stage 1	179 (53.6%)
Stage 2	124 (37%)
Stage 3	28 (08.4 %)
Stage 4	3 (1%)

IQR, Interquartile range; HBP, high blood pressure; NTR, Nothing to report; PCR, Polymerase chain reaction; PCT, Procalcitonin; CRP,C-reactive proteine;

Table 2 presents the characteristics of the study population categorized by disease severity. The study population's age distribution varies across different disease severity stages. The median age increases as the disease severity progresses, with higher medians observed in Stage 2 and Stages 3 & 4 compared to Stage 1. The table also provides information on the gender distribution among the study population, indicating the number and percentage of males in each disease severity stage. The proportion of males differs among disease severity stages. Stage 2 has a higher percentage of males compared to Stage 1 and Stages 3 & 4. However, the difference is not statistically significant based on the chi-square test (p-value > 0.05).

Moving on to comorbidities, the table presents the prevalence of specific conditions such as HBP, Diabetes, Other diseases, and NTR

(Nothing to Report) for each disease severity stage. The numbers and percentages are provided, highlighting the prevalence of these comorbidities within the different stages. HBP shows a slight increase as disease severity progresses, while Diabetes and Other diseases have a higher prevalence in Stages 3 & 4 compared to other stages. The NTR category (Nothing to Report) shows a decreasing trend as disease severity increases.

Table 2 provides information on biomarkers and their association with disease severity stages in the study population. For D-dimer, the median levels show an increasing trend as disease severity progresses. The proportion of positive D-dimer cases also varies significantly

Table 2. Characteristics of the study population according to their disease severity

	Stage 1	Stage 2	Stages 3 & 4	Chi-square(df)	P-value	CvM
Demographics						
Age (y) (continuous)	50 (39-58)	53 (44-63)	60 (48- 68)			
Age (y) (categorical)						
<20, N=5 (1.49%)	4 (1.20%)	1 (0.30%)	0			
20-40, N=64 (19.05%)	41 (12.35%)	21 (6.33%)	1 (0.3%)			
40-60, N=169 (50.30%)	96 (28.92%)	58 (17.47%)	13 (3.91%)	22.9983 (12.0)	0.0277	0.1520
60-80, N=94 (27.98%)	36 (10.84%)	42 (12.65%)	15 (4.22%)			
>80, N=4 (1.19%)	1 (0.30%)	2 (0.60%)	1 (0.30%)			
Sexe (Male)	101 (56.7%)	85 (68.5%)	21 (67.7%)	6.4559 (3.0)	0.0914	0.1392
Comorbidities						
HBP	42 (23.4%)	34 (27.4%)	11 (35.5%)	5.7122 (3.0)	0.1265	0.1308
Diabetes	14 (7%)	15 (12%)	7 (22.6%)	6.7414 (3.0)	0.0806	0.1421
Other diseases	40 (22.5%)	30 (24.2%)	9 (30%)	4.0553 (3.0)	0.2556	0.1102
NTR	84 (55.2%)	58 (38.15%)	10 (6.6%)	2.4218 (3.0)	0.4896	0.0852
Biomarkers						
D-dimer (continuous)	481.1 (256.5-878.4)	597.6 (329.5-1130.8)	1836.9 (684.6-2251.2)			
D-dimer (categorical)						
Positive D-dimer, N=102 (57.95%)	41 (23.3%)	42 (23.89%)	19 (9.09%)			
Negative D-dimer, N=75 (42.05%)	45 (25.57%)	28 (15.91%)	1 (0.57%)	15.1502 (3.0)	0.0017	0.2934
PCR	81 (62.8%)	51 (63.0%)	13 (59.1%)	1.6921 (3.0)	0.6387	0.0854
PCT (continuous)	0.07 (004-0.16)	008 (0.06-0.22)	0.26 (0.0975-0.44)			
PCT (categorical)						
Positive PCT, N=49 (58.33%)	11 (13.10%)	12 (14.29%)	12 (13.10%)	9.3159 (3.0)	0.0254	0.3330
Negative PCT, N=35 (41.67%)	22 (26.19%)	23 (27.38%)	4 (4.76%)			
CRP (continuous)	13.5 (4.95-42.85)	47.3 (17.2-111.1)	120.6 (52.15-194.2)			
CRP (categorical)						
Positive CRP, N=133 (80.61%)	56 (33.94%)	59 (35.76%)	18 (10.91%)	9.6221 (3.0)	0.0221	0.2415
Negative CRP, N= 32 (19.39%)	23 (13.94%)	8 (4.85%)	1 (0.61%)			

df, Degree of freedom; CvM, Cramér-von mises statistical test

Table 3. Predictive performance of COVID-19 severity

Predictive model	Accuracy	Sensitivity
SVM	0.9086	0.8725
KNN	0.8829	0.8720
Decision Tree	0.8553	0.8140
Random Forest	0.9056	0.8674
Bagging	0.8908	0.8725
AdaBoost	0.8938	0.8594
GBM	0.8791	0.8137
XGBoost	0.9085	0.8235

SVM, Support vector machine; KNN, K-Nearest neighbor; GBM, Gradient boosting

among different disease severity stages.

Table 3 presents the predictive performance of various machine learning models used to predict COVID-19 severity based on biomarkers. It includes metrics such as accuracy that measure the overall correctness of the model in predicting disease severity and sensitivity or true positive rate, that measures the model's ability to correctly identify severe COVID-19 cases. The table compares the performance of SVM, KNN, DT, RF, Bagging, AdaBoost, GBM and XGBoost, which have been fine-tuned as described earlier in the previous section.

Discussion

Regarding Polymerase Chain Reaction (PCR), there is no significant variation in the percentage of positive results across disease severity stages.

As for Procalcitonin (PCT), the median levels demonstrate an increasing pattern with advancing disease severity. The proportion of positive PCT cases also differs significantly across disease severity stages.

The obtained results underscore the utility of C-Reactive Protein (CRP) as it shows higher median levels as disease severity advances. The proportion of positive CRP cases varies significantly among different disease severity stages. This observed trend aligns with the study by Tan et al.[13] that focuses on the correlation between CRP levels and the severity of COVID-19, evidenced through computed tomography (CT) findings. They discovered that CRP levels were significantly higher in patients with severe COVID-19 from the initial stages, predating notable changes in CT images. This suggests that CRP can serve as an early biomarker for predicting COVID-19 severity.

These findings suggest that biomarkers such as D-dimer, PCT, and CRP could potentially serve as indicators of disease severity in the study population. However, further analysis and interpretation of the detailed values and statistical tests are necessary to establish stronger associations between these biomarkers and disease severity.

Therefore, to further investigate the associations of these factors across different stages of disease severity, we suggest focusing on the predictive performance of various machine learning models for COVID-19 severity. Table 3 presents the predictive performance measures, including accuracy and sensitivity, for each model.

Among the models evaluated, SVM and XGBoost demonstrate the highest accuracy rates of 0.9086 and 0.9085, respectively, indicating their effectiveness in predicting COVID-19 severity. KNN, Random Forest, and Bagging also exhibit relatively high accuracy rates, ranging from 0.8829 to 0.9056. These results are in harmony with those of Ikemura et al.[22] that utilized automated machine learning to develop models predicting COVID-19 patient mortality, demonstrating their efficiency in generating high-performing, interpretable models for clinical decision support. The best models identified were based on Gradient Boosting Machine (GBM) and Extreme Gradient Boosting (XGBoost). We highlight then the potential of tree-based algorithms over deep learning for analyzing clinical data.

We explain the outperformance of tree-based algorithms by the fact that can naturally handle heterogeneous data (which is the case of our data) without the need for extensive preprocessing. This can also be explained by the non-linear relationships between features and outcomes that tree-based models excel at capturing,

This is also confirmed by Xiong et al.[23] where also a tree-based algorithm showed great performance as Random Forest outperformed with high AUC value.

In terms of sensitivity, SVM and Bagging achieve the highest values of 0.8725, closely followed by KNN with a sensitivity of 0.8720. These models demonstrate the ability to correctly identify a significant proportion of severe COVID-19 cases. On the other hand, Decision Tree, AdaBoost, and Gradient Boosting show relatively lower accuracy rates, suggesting potential limitations in their predictive performance for COVID-19 severity.

Based on these findings, SVM, KNN, Random Forest, and XGBoost appear to be the most promising models for predicting COVID-19 severity in the study population.

Conclusion

In this study, we investigated the potential of C-reactive protein (CRP), procalcitonin (PCT), and D-dimer as biomarkers for disease severity in COVID-19 patients in Kinshasa, Democratic Republic of Congo. Our findings demonstrated that CRP, PCT, and D-dimer levels were significantly higher in severe cases compared to non-severe cases. The median levels of these biomarkers varied across different disease severity stages, indicating their potential as indicators of disease progression. Additionally, machine learning methods were employed to predict disease severity based on these biomarkers, achieving high accuracy.

These results suggest that CRP, PCT, and D-dimer can serve as valuable biomarkers for identifying severe COVID-19 cases in Kinshasa. The use of these biomarkers in clinical practice can aid healthcare professionals in making informed decisions about patient management and improve clinical outcomes. However, further research is needed to validate these findings in larger and more diverse populations. The integration of biomarker-based assessments into clinical protocols has the potential to enhance risk stratification and guide personalized treatment strategies for COVID-19 patients.

References

1. Henry B, Oliveira MD, Benoit S, Plebani M, Lippi G. Hemato- logic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clinical Chemistry and Laboratory Medicine (CCLM). 2020; 58(7): p. 1021-1028.

2. Sarah C, Amy C, Rebecca C, Mariah F, Anthea F, Elizabeth G, et al. Making the case for cross-border public health stratagies: a compartivie assessment of Covid-19 epidemiological trends in the Balkan countries across 17 months. Journal of Biostatistics and Epidemiology. 2022; 8(2).

3. Savy N, Moodie EE, Drouet I, Chambaz A, Falissard B, Kosorok MR, et al. Statistics, philosophy, and health: the SMAC 2021 webconference. The International Journal of Biostatistics. 2022.

4. Suo-wen X, Iqra I, Weng , Jian-ping. Endothelial dysfunction in COVID-19: an overview of evidence, biomarkers, mechanisms and potential therapies. Acta Pharmacologica Sinica. 2023; 44(4): p. 695-709.

5. Giovanni P, Monia M, Cristel R, Aldo T, Tomris O. Biomarkers associated with COVID-19 disease progression. Critical reviews in clinical laboratory sciences. 2020; 57(6): p. 389-399.

6. Karimollah HT, Zahra G, Vahid N. Statistical Considerations in Combining Multiple Biomarkers for Diagnostic Classification: Logistic Regression Risk Score Versus Discriminant Function Score. Journal of Biostatistics and Epidemiology. 2022.

7. Maria CF, Giampiero F, Marco L, Antonio A, Enea B, Carla P, et al. Investigating biomarkers for COVID-19 morbidity and mortality. Current Topics in Medicinal Chemistry. 2023.

8. Marc V, Dmitry S, Marie-Christine B, Frédérique D, François M, Florence H, et al. Prognostic value of cellular population data in patients with COVID-19. Informatics in Medicine Unlocked. 2023; 38: p. 101-207.

9. Marin BG, Aghagoli G, Lavine K, Yang L, Siff EJ, Chiang SS, et al. Predictors of COVID-19 severity: a literature review. Reviews in medical virology. 2021; 31(1).

10. Siddharth S, Kuldeep S, Patel , B S, S PF, Mohammed O, et al. Elevated D-dimer levels are associated with increased risk of mortality in coronavirus disease 2019: a systematic review and meta-analysis. Cardiology in review. 2020; 28(6): p. 295-302.

11. Amit K, Era K, Kiran T, Pramod K, Ganesh C, Aradhana K, et al. Procalcitonin as a predictive marker in COVID-19: A systematic review and meta-analysis. Plos one. 2022; 17(9).

12. Takayuki Y, Mako W, Takahiro Y, Nitin C, Takahisa M, Hirotaka M, et al. Value of leukocytosis and elevated C-reactive protein in predicting severe coronavirus 2019 (COVID-19): A systematic review and meta-analysis. Clinica chimica acta. 2020; 509.

13. Chaochao T, Ying H, Fengxia S, Tan K, Ma Q, Chen Y, et al. C-reactive protein

correlates with computed tomographic findings and predicts severe COVID-19 early. Journal of medical virology. 2020; 92(7).

14. OpenAI. ChatGPT: Language Generation Model. [Online].; 2023 [cited 2023 07 08. Available from: https://chat.openai.com/.

15. Jingyuan L, Yao L, Pan X, Lin P, Haofeng X, Chuansheng L, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. Journal of translational medicine. 2020; 18(1).

16. Fei Z, Ting Y, Ronghui D, Guohui F, Ying L, Zhibo L, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The lancet. 2020; 395(10229).

17. Benkeser D, Mertens A, Colford J, Hubbard A, J M. A machine learning-based approach for estimating and testing associations with multivariate outcomes. The international journal of biostatistics. 2020; 17(1).

18. Pan American Health Organization (2024). Diabetes. [Online]. Available from: https://www.paho.org/fr/sujets/diabete.

19. World Health Organization (2024). Hypertension: Fact sheet. [Online]. Available from: https://www.who.int/fr/news-room/fact-sheets/detail/hypertension#:~:text=On%20consid%C3%A8re%20qu'une%20personne,ou%20%C3%A9gale%20%C3%A0%2090%20mmHg.

20. Guyon I, Makhoul J, Schwartz R, Vapnik V. What size test set gives good error rate estimates?. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1),.; 1998. p. 52-64.

21. Griesbach C, Safken B, Waldmann E. Gradient boosting for linear mixed models. The International Journal of Biostatistics. 2021; 17(2).

22. Ikemura K, Bellin E, Yagi Y, Billett H, Saada M, Simone K, et al. Using automated machine learning to predict the mortality of patients with COVID-19: prediction model development study. Journal of medical Internet researc. 2021.

23. Xiong Y, Ma Y, Ruan L, Li D, Lu C, Huang L. Comparing different machine learning techniques for predicting COVID-19 severity. Infectious diseases of poverty. 2022; 11(1).