

Decision tree analysis as predictor tool for in-hospital mortality in critical SARS-CoV-2 infected patients

Adina Hutanu^{1,2#}, Anca A. Molnar^{1,2#}, Krisztina Pal^{1,2*}, Manuela R. Gabor³, Janos Szederjesi^{4,5}, Minodora Dobreanu^{2,6}

1. Department of Laboratory Medicine, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania
2. Department of Laboratory Medicine, Emergency Clinical County Hospital Targu Mures, Romania
3. Department of Economic Science, Faculty of Economics and Law, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania
4. Department of Anesthesiology and Intensive Care, Emergency Clinical County Hospital Targu Mures, Romania
5. Department of Anesthesiology and Intensive Care, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania
6. Department of Immunology, Center for Advanced Medical and Pharmaceutical Research, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania

ABSTRACT

Identification of predictive biomarkers for the evolution of critically ill COVID-19 patients would represent a milestone in the management of patients and in human and financial resources prioritization and allocation. This retrospective analysis performed for 396 critically ill COVID-19 patients admitted to the intensive care unit aims to find the best predictors for fatal outcomes in this category of patients. The inflammatory and metabolic parameters were analyzed and Machine Learning methods were performed with the following results: (1) decision tree with Chi-Square Automatic Interaction Detector (CHAID) algorithm, based on the cut-off values using ROC Curve analysis, indicated NLR, IL-6, comorbidities, and AST as the main in-hospital mortality predictors; (2) decision tree with Classification and Regression Tree (CRT) algorithm confirmed NLR alongside CRP, ferritin, IL-6, and SII (Systemic Inflammatory Index) as mortality predictors; (3) neural networks with Multilayer Perceptron (MLP) found NLR, age, and CRP to be the best mortality predictors. Structural Equation Modeling (SEM) analysis was complementarily applied to statistically validate the resulting predictors and to emphasize the inferred causal relationship among factors. Our findings highlight that for a deeper understanding of the results, the combination of Machine Learning and statistical methods ensures identifying the most accurate predictors of in-hospital mortality to determine classification rules for future events.

Keywords: COVID-19; Interleukin 6; SEM analysis; Decision tree; Machine Learning algorithms; Neural networks, SEM

Received: 22 March 2023; Accepted: 18 April 2023; Published: 20 April 2023

INTRODUCTION

The evolution of the SARS-CoV-2 infection to a pandemic was one of the greatest challenges facing not only the medical staff, but also humanity for almost two years. The disease severity and variability in symptoms were, on the other hand, one of the biggest challenges for the medical staff, since many COVID-19 patients had an unpredictable evolution with rapid worsening. Many patient- and disease-related factors were incriminated in the poor evolution, from the personal medical history of the patients to the coagulopathy and multiple organ dysfunctions. In this regard, researchers and doctors have tried to identify the most important predictors for disease evolution and fatal outcome in COVID-19 patients. Several risk factors for critical evolution and fatal out-

come related to demographics, preexisting comorbidities, and clinical features were identified in a very recent review. The study concluded that age and comorbidities along with inflammation and liver, renal, and coagulation dysfunctions are significantly related to unfavorable outcomes [1]. Besides laboratory parameters that are critically affected in severe disease forms, clinical risk scores enable the stratification of patients and provide decision-making support. Several clinical scores were analyzed in relation to the disease evolution of COVID-19 patients, like National Early Warning Score 2 (NEWS2), which predicts in-hospital mortality on emergency admission, Systemic Inflammatory Response Syndrome (SIRS), and modified quick sequential organ failure assessment (MqSOFA). While NEWS2 underlines the rapid deterioration of acute patients based exclusively

* Correspondence to: Krisztina Pal, Department of Laboratory Medicine, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania.
E-mail: dr.krisztinapal@gmail.com

These authors contributed equally to this work and share first authorship

on the assessment of physiological parameters and oxygen therapy requirements, SIRS is a score that includes, aside from the respiratory and heart rate, temperature and abnormal white blood cells on admission [2]. The Modified quick Sequential Organ Failure Assessment (MqSOFA) is a SOFA score improved with the SpO₂/FiO₂ ratio, with a good predictive value for outcome evolution of severe SARS-CoV-2 infections [3]. All the above-mentioned scores are calculated mainly based on clinical parameters, with a few exceptions. Because in COVID-19 patients the conversion to more severe disease implies important modifications in laboratory parameters and in line with the fact that severe patients are monitored daily for their clinical evolution, we intended to assess the usefulness of these parameters in mortality prediction. Since several laboratory parameters are questioned daily for critical patients, all these results in conjunction with demographic data, comorbidities, and clinical parameters have tremendous potential to estimate the outcome of the patients. However, this aspect is difficult to identify at a glance, thus, a statistical algorithm might be useful for patient stratification.

By using artificial intelligence, all these data can assist doctors in making the right decision in the shortest time possible. There are many tools covered by artificial intelligence that might be useful for the estimation of disease severity, disease outcome, or response to a specific treatment. One of them is the decision tree algorithm that analyzes in a sequential mode several variables in relation to the outcome. After the detection of the most accurate thresholds for the continuous variables and conversion into categorical variables, the root, classification and nodes are built by selecting attributes in a random manner and consequently creating a split function for classification and node definition [4].

The complex study performed by Karthikeyan et al. using different machine learning models for the assessment of early mortality prediction estimated that the combination of age, neutrophils, high-sensitive C-reactive protein (hsCRP), lymphocytes, and lactate dehydrogenase (LDH) are accurate mortality predictors in COVID-19 patients [5]. Wang et al. developed two decision models for mortality prediction, one based exclusively on clinical parameters and the second based on laboratory parameters in conjunction with age. The latter which included hsCRP, oxygen saturation (SpO₂), lymphocytes, neutrophils in absolute number, D-dimers, aspartate aminotransferase (AST), and renal function evaluated by estimated glomerular filtration rate (eGFR) had a more powerful discrimination rate than the clinical model alone [6].

In this study, we aimed to retrospectively analyze laboratory parameters in conjunction with clinical and

demographic characteristics for 396 patients, 183 survivors and 213 with fatal outcomes, admitted to intensive care units (ICU) with critical SARS-CoV-2 infection. Consequently, we intended to construct a decision tree for risk mortality estimation for critically ill COVID-19 patients admitted to the ICU.

The primary outcome was to characterize the changes in laboratory parameters in critically ill COVID-19 patients. The secondary outcome was to determine the cut-off values for the most representative laboratory parameters from each test-group and to generate a decision tree and a path analysis (SEM) to identify the most predictive laboratory tests for in-hospital mortality prediction based on the recorded data.

METHODS

Data collection

In this retrospective study, data were collected for all patients with COVID-19 admitted to the Intensive Care Unit (ICU) of the Emergency Clinical County Hospital of Târgu Mureș, Romania, from September 2020 to October 2021. The study was approved by the Ethics Committee of the Emergency Clinical County Hospital Târgu Mureș, Romania, no 26973/10.11.2021.

Patients' results were obtained from the Laboratory Information System (LIS). Due to the retrospective design of the study, informed consent was not applicable before data collection. Laboratory results, demographics, medical personal history, and clinical parameters were recorded. The inclusion criteria were: age above 18 years, positive RT-PCR SARS-CoV-2 test results on admission and severe/critical COVID-19 patients who were monitored at least once for IL-6 during ICU admission. In order to divide patients according to comorbidities, the whole group was dichotomized into 2 arms, with diabetes and with other comorbidities (hypertension, chronic ischemic heart disease, obesity, chronic kidney disease, atrial fibrillation, asthma).

The results for IL-6 were recorded for a dynamic description of this pro-inflammatory cytokine and in addition, all biochemical and complete blood count (CBC) parameters were included in the database.

The biochemistry panel for hepatic function assessment consisted of aspartate aminotransferase (AST), alanine aminotransferase (ALT), prothrombin time with international normalized ratio (INR), total bilirubin, direct bilirubin, total protein and albumin. For the evaluation of kidney function, the panel included creatinine, estimated glomerular filtration rate (eGFR), urea, sodium, and potassium.

The inflammatory biomarkers considered were IL-6, C-reactive protein (CRP), ferritin, D-dimers, neutrophils to lymphocyte ratio (NLR), Systemic Inflammatory Index (SII), and fibrinogen. Both indices, SII and NLR are derived from the CBC and were calculated using the formulas exemplified with equations (1) and (2):

$$SII = \frac{\#Neu \times Plt}{\#Lymph} \quad (1)$$

where SII=Systemic Inflammatory Index, #Neu=absolute neutrophils count/ μ L, Plt=platelets, #Lymph=absolute lymphocytes count/ μ L.

$$NLR = \frac{\#Neu}{\#Lymph} \quad (2)$$

where NLR=Neutrophils to Lymphocytes ratio, #Neu=absolute neutrophils count/ μ L, #Lymph=absolute lymphocytes count/ μ L.

The patients' management was in accordance with disease severity, lung impairment and requirement for mechanical ventilation; mortality was recorded at the ICU discharge.

Statistical analysis

The normality of the distribution was verified using the One-Sample Kolmogorov-Smirnov test. To compare the differences between survivors and non-survivors, the Mann-Whitney test was used for data with non-parametrical distribution and the Student t-test for data with normal distribution. The continuous variables were expressed as the median and interquartile range (IQR) for data with non-parametrical distribution, and as mean $\pm$ standard deviation (SD) for the parametrical ones. For the categorical data, analysis using Chi-square bivariate test was performed and the results were reported as relative numbers and percentages.

To use results from all eligible patients retrieved from the LIS, the missing values for laboratory parameters were replaced using the imputation method by the mean or median of the series for both the decision tree and SEM/path analysis. The optimal cut-off (CO) values were established using ROC Curve analysis (based on Youden's index formula), and further continuous variables were dichotomized into categorical ones and used for the decision tree analysis.

Statistically significant demographic and laboratory parameters, with AUC (Area Under the Curve) > 0.5 and $p \leq 0.05$ were used as independent variables for the assessment of in-hospital mortality.

According to the aim of the research, complementary machine learning methods were applied as predictive modeling techniques, as follows:

1. the decision tree with CHAID algorithm in both cases: (a) no validation and (b) split-sample validation (training and test) by use of random assignments of 50% for the training sample and 50% for the test sample having the advantage that they are structurally the same, respectively the same variables and the same possible categories [7];
2. the decision tree with CRT algorithm due to the advantages of finding boundaries with the more granular level of precision using a very different approach and for Gini coefficient to establish the purity/impurity of the leaf node. For both CHAID and CRT algorithms the minimum number of cases for growth limits were considered as 100 for the parent node and 50 for the child node, while the literature recommended 30 cases for parent and 15 cases for child node [7].
3. neural networks with Multilayer Perceptron (MLP).
4. Structural Equation Modeling (SEM) to confirm correlations and inferred causal relationships among factors[8], for the advantage of the methods to take potential measurement errors into account [8] compared to other multivariate analyses. The minimum conditions of the sample size were respected, the 10 time rule being applied, respectively the sample size to be equal to 10 times the number of independent variables [9]. In our research, 12 independent variables were used for SEM, and the sample size is equal to 396 cases. Furthermore, we have taken into consideration a minimum value of path coefficients between 0.11 and 0.2 with a significance level of 5% and therefore according to Hair Jr. et al we respected the minimum sample size of 155 cases [10]. Referring to the treatment of missing values, we opted not to systematically delete them due to the decreasing of variation in the data and the potential biases introduced as a result [10]. In the case of using decision trees as predictive modeling techniques, not the accuracy of the model is important, but rather the cost of findings, according to the 9 laws of data mining of Khabaza [11].

All variables were transformed into dichotomous ones using the cut-off values from ROC Curve based on Youden's index formula. The dependent variable was mortality (yes/no) and the independent variables introduced in the analysis were: age, gender, comorbidities, medication, co-infections, NLR, IL-6, SII, ferritin, CRP, D-Dimers, blood glucose, fibrinogen, creatinine, AST with the mentioned CO values in Figure 3 for statistically significant variables.

Structural Equation Modeling (SEM) was used to validate the results of the machine learning methods to establish the inferred causal relationship among factors [12] and to find the best predictor for in-hospital mortality among severe COVID-19 patients. For standardized SEM the goodness-of-fit statistics of the model (statistical power of the model) for the estimated model were considered as follows: CMIN/DF (the relative chi-squared) < 15%, Comparative fit index (CFI) $\approx$ 0.9 and RMSEA (Root Mean Square Error of Approximation) $\approx$ 0.1.

Data were analyzed with the licensed software SPSS (Statistical Package for the Social Sciences) 23.0 and SPSS-AMOS 22.0.

RESULTS

Patients' data characterization

This was a retrospective study intending to analyze parameters associated with inflammation, liver and renal function, as well as hematological parameters derived from the complete blood count, for a group of 396 patients with severe forms of COVID-19 infection, hemodynamically unstable and who required ventilatory support, admitted into the ICU of the Emergency Clinical County Hospital Târgu Mureș between September 2020 and October 2021.

Out of the 396 included patients only 183 survived, while for 213 in-hospital mortality was recorded. De-

ceased patients were older ($p < 0.0001$) compared to survivors, while male patients were prevalent in the fatal outcome group (60.6%), with results close to the limit of statistical significance ($p = 0.073$). There were no differences in the duration of hospitalization between survivors and non-survivors. Patients with diabetes mellitus or other co-morbidities, as well as those who developed in-hospital infections, were more likely to be among the patients who died.

The demographics and clinical aspects of the patients included in the analysis are detailed in table 1.

Comparative bivariate analysis of laboratory parameters

Assessment of the biomarkers of inflammatory response

Following bivariate analysis of inflammatory biomarkers between the two groups of patients (survivors and non-survivors), all the studied parameters appear to be significantly higher in the deceased group compared to survivors, $p < 0.001$ for all. While interleukin 6 (IL-6), ferritin, and C-reactive protein (CRP) have double concentration values in the non-survivor group compared to patients with favorable outcomes, for neutrophils to lymphocytes ratio (NLR) and Systemic Inflammatory Index (SII) the values are almost three times higher in the fatal outcome group compared to the survivor group (table 2).

Table 1. Demographic and clinical data of critically ill COVID-19 patients included in the retrospective analysis.

Parameter	Non-survivors (n=213)	Survivors (n=183)	p
Age, years	67.9 $\pm$ 12.8	59.8 $\pm$ 19.1	0.000*
Gender (male) n (%)	129 (60.6%)	95 (51.9%)	0.073**
Diabetes Mellitus n (%)	38 (17.8%)	9 (4.9%)	0.000**
Other Comorbidities n (%) (at least 1)	20 (9.4%)	6 (3.3%)	0.000**
Tocilizumab n (%)	7 (3.3%)	1 (0.5%)	0.052**
Remdesivir n (%)	42 (19.7%)	16 (8.7%)	0.002**
Co-Infections (if present) n (%)	23 (10.8%)	4 (2.2%)	0.001**
Length of hospitalization (Days) median (interquartile range)	10 (7-14)	10 (7-17)	0.497***

Values are expressed as relative numbers and percentages after the application of Chi-square bivariate test, as mean ($\pm$ SD), or as median and interquartile range (IQR);

* student t-test, **Chi-square test, ***- Independent samples Mann-Whitney test.

Table 2. Bivariate analysis of inflammatory parameters for the patients dichotomized in survivors and non-survivors.

Parameter	Non-survivors (n=213)	Survivors (n=183)	p
IL-6, pg/mL	53.25 (22.99-129.95)	20.34 (8.23-44.93)	0.000
Ferritin, ng/mL	1155.0 (666.0-2281.9)	581.1 (214.2-1234.5)	0.000
CRP, mg/L	102.1 (49.7-176.5)	54.0 (12.4-62.6)	0.000
D Dimers, ng/mL	1259.0 (535.0-3891.0)	1042.7 (249.5-1195.7)	0.000
Fibrinogen, mg/dL	498 (391-695.20)	410 (306- 539)	0.001
NLR	14.4 (8.5-23.3)	5.33 (2.42-12.11)	0.000
SII	3303.50 (1749.86-5919.83)	1373.37 (513.84-3359.05)	0.000

CRP= C-reactive protein; NLR= neutrophil to lymphocyte ratio; SII= systemic inflammatory index. Results are expressed as median (IQR) after the application of the Mann-Whitney test; the significance level was set at 0.05.

Assessment of hepatic and renal functions

Comparative analysis of liver function parameters of the two groups of patients, shown in table 3, reveals statistically significant differences regarding AST, total proteins, and albumin serum levels. The group of non-surviving patients had significantly lower total proteins and albumin values ($p=0.01$ for total proteins, and $p<0.0001$ for albumin) and higher activity of AST ($p<0.0001$). In the same comparison analysis, non-surviving patients had higher values of prothrombin time estimated by International Normalized Ratio (PT/INR), close to the limit of statistical significance $p=0.07$ compared to the group of surviving patients, denoting the affected protein synthesis.

In addition to creatinine, eGFR, and urea, electrolyte status, specifically sodium and potassium were considered for renal function evaluation. As can be observed in table 3, there are statistically significant differences for all parameters related to renal and hepatic functions. The non-surviving patients had increased levels of creatinine, urea, sodium, and potassium, compared to the group of surviving patients.

Complete blood count

From the comparative analysis of the CBC parameters of the two groups of patients, it can be discerned that the absolute number of white blood cells (WBC) and neutrophils in the non-surviving group of patients was significantly higher, while the absolute number of lymphocytes was significantly lower in the same group of patients compared to the surviving patients. No statistically significant differences were observed for other hematological parameters, more specifically for the platelets, hemoglobin, hematocrit, or erythrocytes, as can be observed in table 4.

Cut-off thresholds for the parameters included in the study

In addition to transforming continuous variables into categorical variables to dichotomize patients, receiver operating characteristic (ROC) curves were used to establish cut-off thresholds to identify significant predictors of progression to death.

The curves for the most significant parameters ($AUC>0.7$; $p\leq 0.05$) are illustrated in Figures 1 and 2.

Table 3. Bivariate analysis of the parameters related to liver and renal functions for the two groups of critically ill patients included in the study.

Parameter	Non-survivors (n=213)	Survivors (n=183)	p
PT/INR	1.17 (1.07- 1.32)	1.13 (1.03- 1.29)	0.079
AST, U/L	49.10 (34.30-73)	33.75 (21.40- 56.20)	0.000
ALT, U/L	34 (23-53)	29.0 (17.90- 74.55)	0.446
Total Bilirubin, mg/dL	0.46 (0.31- 0.70)	0.45 (0.32- 0.62)	0.766
Direct Bilirubin, mg/dL	0.26 (0.17- 0.47)	0.26 (0.18- 0.37)	0.423
Total Proteins, g/dL	5.66 $\pm$ 0.72	6.01 $\pm$ 0.85	0.010
Albumin, g/dL	3.0 (2.70- 3.30)	3.51 (3.10- 3.77)	0.000
Creatinine, mg/dL	0.99 (0.78-1.89)	0.78 (0.62-1.10)	0.000
eGFR, mL/min/1.73 m ²	69.35 (34.45-97.32)	89.10 (59.61-118.73)	0.000
Urea, mg/dL	77.50 (55.20-131.20)	44.28 (29.45-72.90)	0.000
Sodium, mmol/L	141 (138-144)	139 (136.25-142)	0.000
Potassium, mmol/L	4.33 (3.86-4.85)	4.10 (3.73-4.65)	0.019

PT/INR= prothrombin time/International Normalised Ratio; AST= aspartate aminotransferase; ALT= alanine aminotransferase, eGFR=estimated glomerular filtration rate. Results are expressed as mean $\pm$ SD, or as median and interquartile range (IQR); significance level was set at 0.05.

Table 4. Bivariate analysis of hematological parameters for the two groups of critically ill patients included in the study.

Parameter	Non-survivors (n=213)	Survivors (n=183)	p
White Blood cells, 10 ³ /μL	12.73 (8.60-16.75)	7.59 (5.43-11.63)	0.000
Neutrophils #, 10 ³ /μL	10.73 (6.94-14.67)	5.72 (3.13- 9.76)	0.000
Lymphocytes #, 10 ³ /μL	0.71 (0.47- 1.05)	1.03 (0.62- 1.56)	0.000
Monocytes #, 10 ³ /μL	0.52 (0.32-0.86)	0.55 (0.35-0.82)	0.895
RBC, 10 ⁶ /μL	4.05 (3.49- 4.50)	4.24 (3.63- 4.70)	0.133
Hemoglobin, g/dL	11.76 $\pm$ 2.35	11.91 $\pm$ 2.38	0.422
Hematocrit (%)	36.19 $\pm$ 7.02	36.38 $\pm$ 6.69	0.653
MCV, fL	89.87 $\pm$ 9.99	88.89 $\pm$ 8.07	0.065
MCH, pg	29.41 $\pm$ 2.24	29.06 $\pm$ 3.20	0.480
MCHC, g/dL	32.41 $\pm$ 1.61	32.68 $\pm$ 1.50	0.073
Platelets, 10 ³ /μL	230.50 (159.25-302)	247.50 (184.50-316.75)	0.110

MCV=Mean erythrocyte volume; MCH=Mean erythrocyte hemoglobin; MCHC=Mean erythrocyte hemoglobin concentration, RBC=red blood cells. # Absolute Neutrophils count/μL, Plt=platelets, #absolute Lymphocytes count/μL, #absolute Monocytes count/μL.

Values are expressed as median and interquartile range (IQR) after application of the Mann-Whitney test, or as mean ($\pm$ SD) after application of the student t-test; the significance level was set at 0.05.

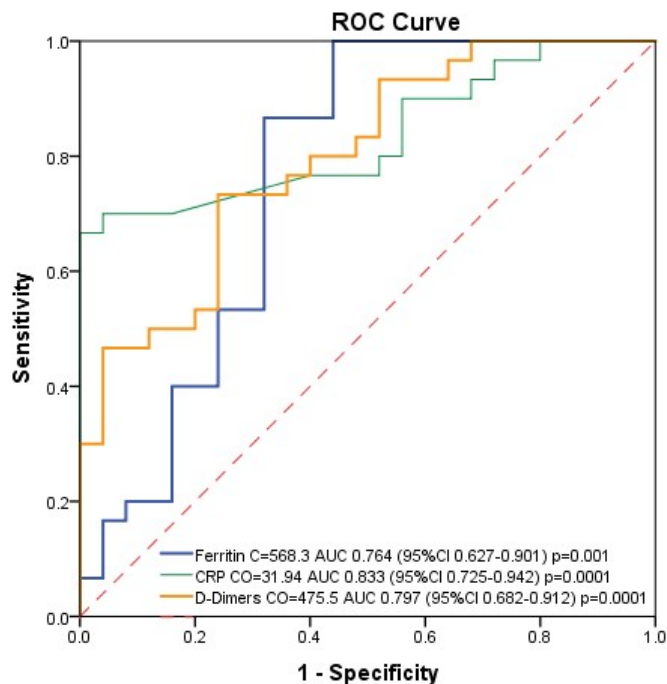


Fig. 1. Receiver Operating Characteristic (ROC) curves and optimal cut-offs for ferritin, C-reactive protein (CRP) and D-dimers for in-hospital mortality prediction in critically ill COVID-19 patients

The risk of death for the patients with severe COVID-19 admitted to the ICU was calculated based on cut-off values using multivariate logistic analysis. Risks were estimated both unadjusted and adjusted for age, sex, diabetes mellitus, and for the presence of hospital-associated infections.

The estimated risk of death expressed as unadjusted and adjusted OR with the corresponding confidence intervals (CI=95%) are presented in table 5.

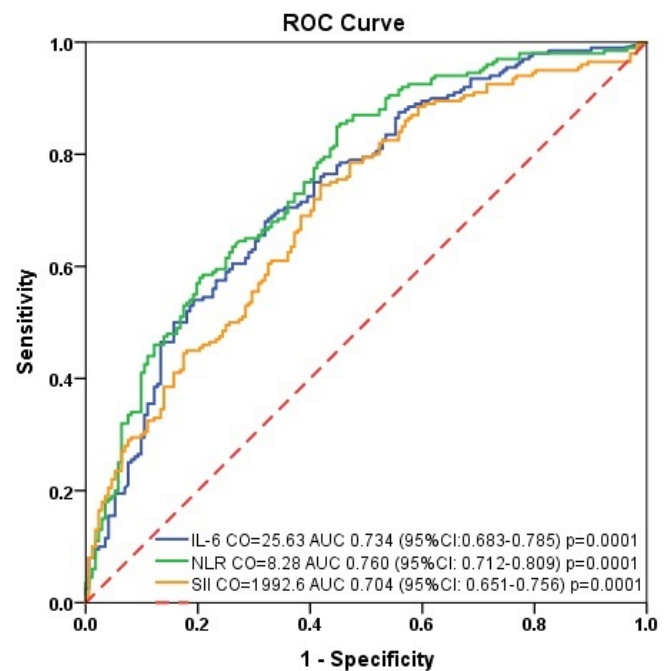


Fig. 2. Receiver Operating Characteristic (ROC) curves for IL-6, Neutrophils/Lymphocytes ratio (NLR) and Systemic Inflammatory Index (SII) for in-hospital mortality prediction in critically ill COVID-19 patients

CHAID (Chi-Square Automatic Interaction Detector), and CRT (Classification and Regression Tree) algorithms for decision tree development for mortality prediction in patients with severe COVID-19

To identify the best mortality predictors for patients with severe forms of COVID-19, the decision tree shown in figure 3 was developed. The Chi-Square Automatic Interaction Detector (CHAID) growing method was

Table 5. The estimated risk of death expressed by unadjusted and adjusted OR for the patients with severe SARS-CoV-2 infection admitted to intensive care wards.

Parameter	OR (CI=95%) unadjusted	p	OR (CI=95%) adjusted	p
Age	1.425 (1.172-1.733)	0.000	2.119 (1.417-3.166)	0.000
<i>Inflammatory parameters</i>				
WBC	1.835 (1.492-2.257)	0.000	3.888 (2.528-5.979)	0.000
Neutrophils	1.940 (1.574-2.393)	0.000	4.471 (2.890-6.917)	0.000
NLR	1.867 (1.543-2.273)	0.000	4.614 (2.958-7.199)	0.000
IL-6	1.762 (1.456-2.132)	0.000	3.798 (2.493-5.785)	0.000
SII	1.732 (1.416-2.119)	0.000	3.482 (2.265-5.353)	0.000
Ferritin	1.600 (1.334-1.920)	0.000	4.096 (2.457-6.828)	0.000
CRP	1.620 (1.308-2.007)	0.000	5.032 (2.564-9.877)	0.000
Fibrinogen	1.408 (1.060-1.871)	0.160	2.114 (1.162-3.845)	0.014
<i>Metabolic parameters</i>				
AST	1.560 (1.258-1.934)	0.000	2.572 (1.663-3.977)	0.000
Glucose	1.167 (1.028-1.326)	0.180	2.240 (1.457-3.442)	0.000
Urea	2.134 (1.698-2.683)	0.000	5.324 (3.354-8.449)	0.000
Creatinine	1.560 (1.254-1.877)	0.000	2.540 (1.668-3.869)	0.000
Sodium	1.298 (1.061-1.589)	0.012	1.769 (1.149-2.725)	0.010
Potassium	1.355 (1.092-1.682)	0.005	1.860 (1.209-2.862)	0.005

WBC= white blood cells; NLR= neutrophils to lymphocytes ratio; SII= inflammatory systemic index; CRP= C-reactive protein; AST= aspartate aminotransferase. Results are predicted for both unadjusted and adjusted OR for age, sex, diabetes mellitus and hospital-associated infections.

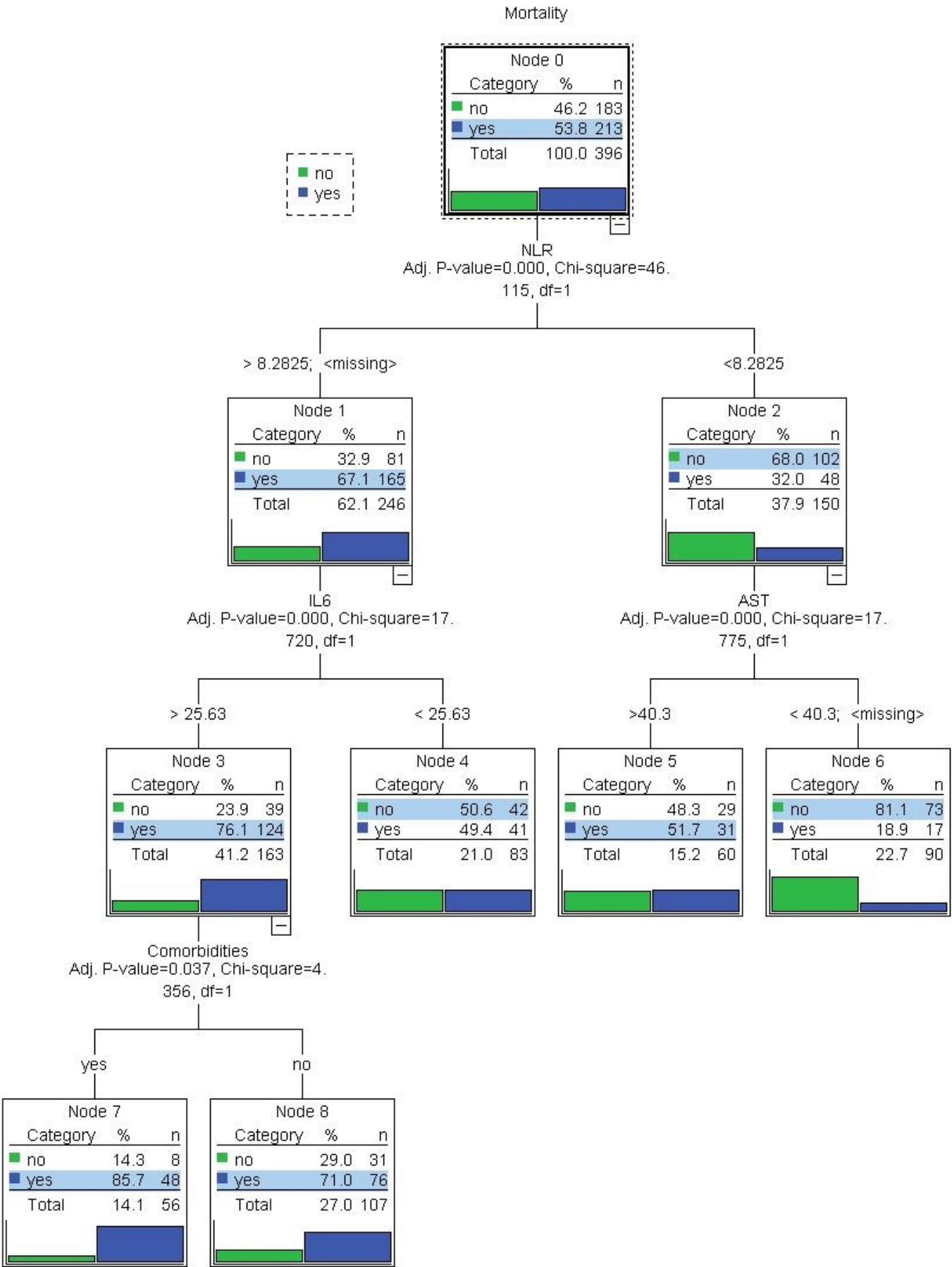


Fig. 3. Chi-Square Automatic Interaction Detector (CHAID) growing method was used for decision tree analysis to assess the mortality prediction in critically ill patients with COVID-19 infection.

adopted using demographic and laboratory parameters as independent variables and mortality as the dependent variable.

The optimal cut-off value was established using ROC Curve analysis ($AUC > 0.5$) and continuous variables were transformed into categorical ones and further used for the decision tree analysis. Furthermore, the Imputation Method was used for missing data, by replacing the missing values with the mean of the series.

The results of the CHAID analysis of the decision tree method with no validation but with adjusted significance values using Bonferroni method are presented in Figure 3.

After applying the sample-split validation method, using random assignment 50% for training samples and 50% for test samples, IL-6 was found as the best predictor for in-hospital mortality (adjusted p-value = 0.000) for both training and test samples.

Of the total of 396 patients included in the predictive analysis, more than half, 53.8% ($n=213$), had fatal outcomes. By applying the decision tree analysis, we were able to estimate the influence of the main laboratory variables on the death rate.

Thus, the first parameter with relevance for fatal outcome prediction was the NLR ratio; among the 246 patients with NLR values above the cut-off of 8.28 (Node 1), a percentage of 67.1% ($n=165$) died, while only 32.9% ($n=81$) survived severe forms of COVID-19 with high significance $p < 0.0001$. Values below the cutoff values for NLR (Node 2) were identified in 37.9% ($n=150$) of patients, of which 68.0% ($n=102$) survived.

The next decisional parameter for Node 1 ($NLR > 8.28$) with a highly significant statistical value (adjusted $p < 0.0001$) was IL-6, with a cutoff-determined value of 25.63 pg/ml. A number of 163 patients were identified above the established cutoff value for IL-6, of which a remarkably high percentage 76.1% ($n=124$) died, and only 23.9% ($n=39$) survived. On the other hand, below the cut-off value, we found a number of 83 patients, of which the deceased and surviving patients were represented in relatively equal percentages 49.4% ($n=41$) and 50.6% ($n=42$), respectively.

Next, for surviving patients with NLR and IL-6 values above the cut-off values of 8.28 and 25.63 pg/ml respectively (Node 3), the next statistically significant decision parameter was the presence or absence of comorbidities, with adjusted p value of 0.030. In this case, the proportion of patients without comorbidities was 27.0% ($n=107$) and that of those with associated conditions was 14.1% ($n=56$). The proportion of surviving patients was higher among those without comorbidities (29.0%), compared to the patients group with comorbidities (14.3%).

On the right hand of the decision tree, for patients in whom the NLR value was below the cut-off of 8.28 (Node 2), the parameter identified as having the highest decision-making power on mortality was the AST value, adjusted p-value < 0.0001 for the cut-off value of 40.3 U/L. Below this threshold of 40.3 U/L we found 22.7% ($n=90$) of the patients, most of whom were survivors 81.1% ($n=73$) vs deceased 18.9% ($n=17$). Above the threshold value of 40.3 U/L for AST, more than half of the patients 51.7% ($n=31$) died, and 48.3% ($n=29$) survived, out of a number of 60 patients ($p < 0.0001$).

According to the decision tree analysis developed with the CHAID method (with adjusted significance values using the Bonferroni method, with no validation), the main predictors of mortality in the case of patients with severe forms of COVID-19 admitted to intensive care units were NLR above 8.28, the concentration of IL-6 greater than 25.63 pg/ml and the presence of at least one comorbidity.

In the present study, by applying complex statistical methods to identify the best predictors for in-hospital mortality in critically ill SARS-CoV-2 infected patients, we opted for the decision tree with adjusted significance values using the Bonferroni method. At the same time, for patients who had NLR values below the cut-off threshold, the main predictor of mortality was AST, greater than 40.3 U/L.

The decision tree with CRT (Classification and Regression Tree) was applied with the same dependent and independent variables from the CHAID algorithm to validate or invalidate the previous results. The advantage of this method compared with CHAID is the normalized importance of the independent variables. The results confirm NLR as the best predictor with a number of 238 subjects included in Node 1 ($NLR > 8.28$) with 162 non-survivors (68.1%) and 76 survivors (31.9%). For Node 2 ($NLR < 8.28$) the group includes a number of 153 subjects (67.7% non-survivors and 32.3% survivors). For the group with $NLR < 8.28$, the next predictor is CRP ($CO = 31.94$ mg/L). Node 3 ($CRP > 31.94$ mg/L) includes 82 subjects, 58.5 % survivors and 41.5% non-survivors and Node 4 ($CRP < 31.94$ mg/L) has a number of 76 subjects (77.6% survivors and 22.4% non-survivors). Figure 4 shows the normalized importance of the independent variables as follows ($>50\%$): CRP, ferritin, NLR and IL-6.

Neural networks analysis for mortality prediction

For a comprehensive evaluation of the main predictors of in-hospital mortality of patients with severe COVID-19 and to accommodate all the ways in which the values of one predictor may affect the impact of other predictors, we applied the artificial neural network analysis (ANN)

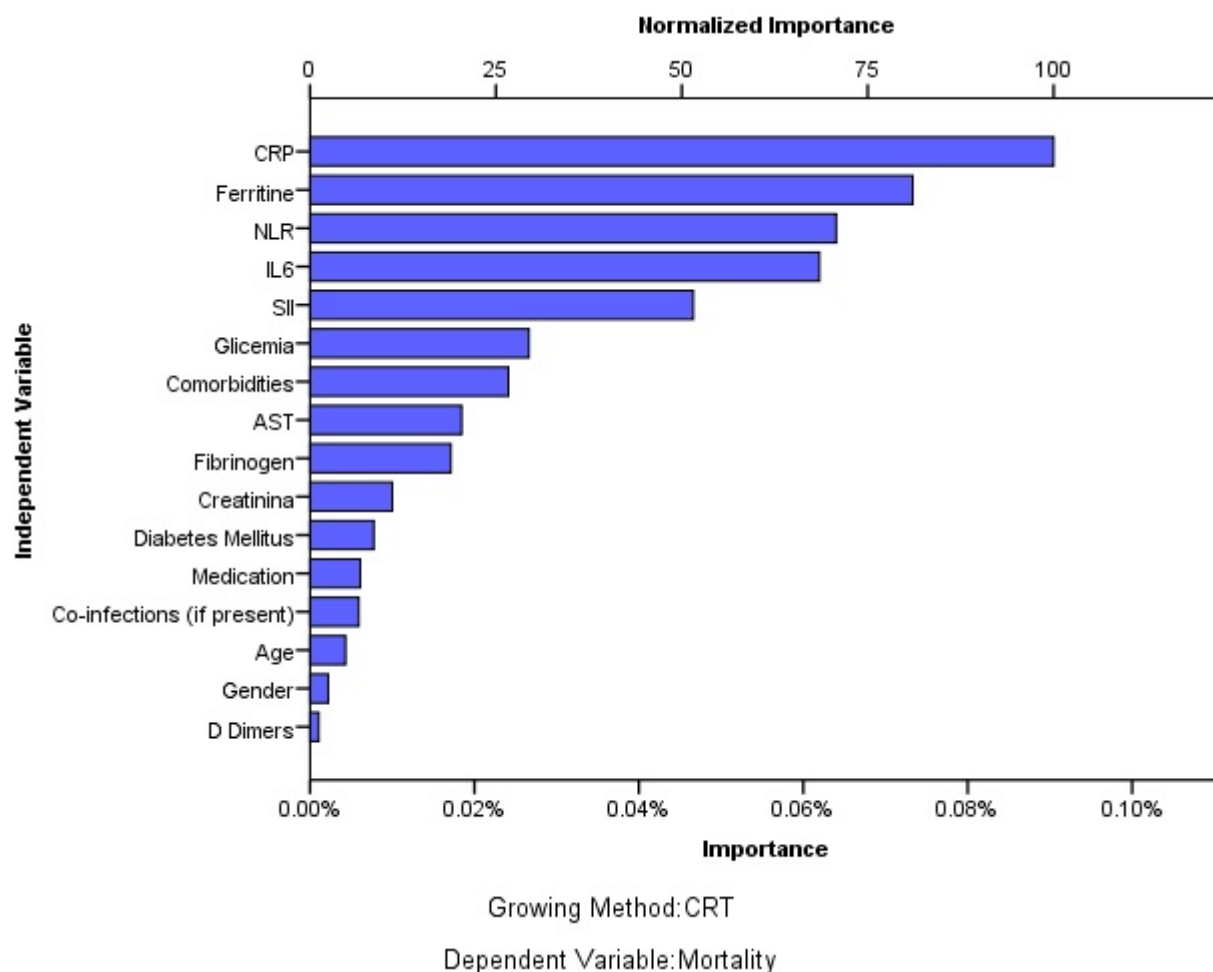


Fig. 4. The normalized importance of the independent variables for decision tree with Classification and regression Tree.

with Multilayer Perceptron (MLP) algorithm, also by using the SPSS software. We opted for the MLP algorithm due to its flexibility and lack of distribution assumption for analyzed data. The neural network analysis indicates the optimal predictors for mortality and the results confirmed NLR as the best predictor (with 100% for normalized importance), followed by age and CRP (Figure 5).

None of the final nodes from the decision tree are “pure” (all the individuals belong to the same group), which means the resulting groups are still heterogeneous and this reveals that there are more causal/association relationships or other influences/good predictors which dichotomize the survivors from non-survivors. Therefore, in the next paragraph, we present the results of multivariate analysis, the SEM – Structural Equation Modeling to complete the results from machine learning methods.

The Structural Equation Modeling (SEM) for mortality prediction

In order to validate The Chi-Square Automatic Interaction Detector (CHAID), CRT and neural networks results, to establish the causal relationship between variables

and to confirm the best predictor for in-hospital mortality among severe COVID-19 patients, a statistical multivariate analysis- the Structural Equation Modeling (SEM) analysis was used. The proposed model contains the endogenous, unobserved variable e_1 (mortality) and 12 exogenous variables (CRP, SII, NLR, IL-6, serum creatinine, AST, INR, age, gender, diabetes mellitus (DM), comorbidities other than DM and hospital-associated infections). The structural SEM model analyzes the causal relationships between variables and identifies among all the parameters introduced in the model those variables with the most direct or indirect influence on the dependent variable, respectively with causal relationship (combination of variables) to death (figure 6).

Regression weights analysis demonstrates the direct relationships of each variable on “death” while the covariances assist to analyze and determine the indirect causes of death due to the existence of statistically significant co-variances/correlations.

As can be seen in figure 6, the variables with a direct effect on “mortality”, statistically significant (bias-corrected) are (in descending order of the values of the “standardized” regression coefficients): comorbidities

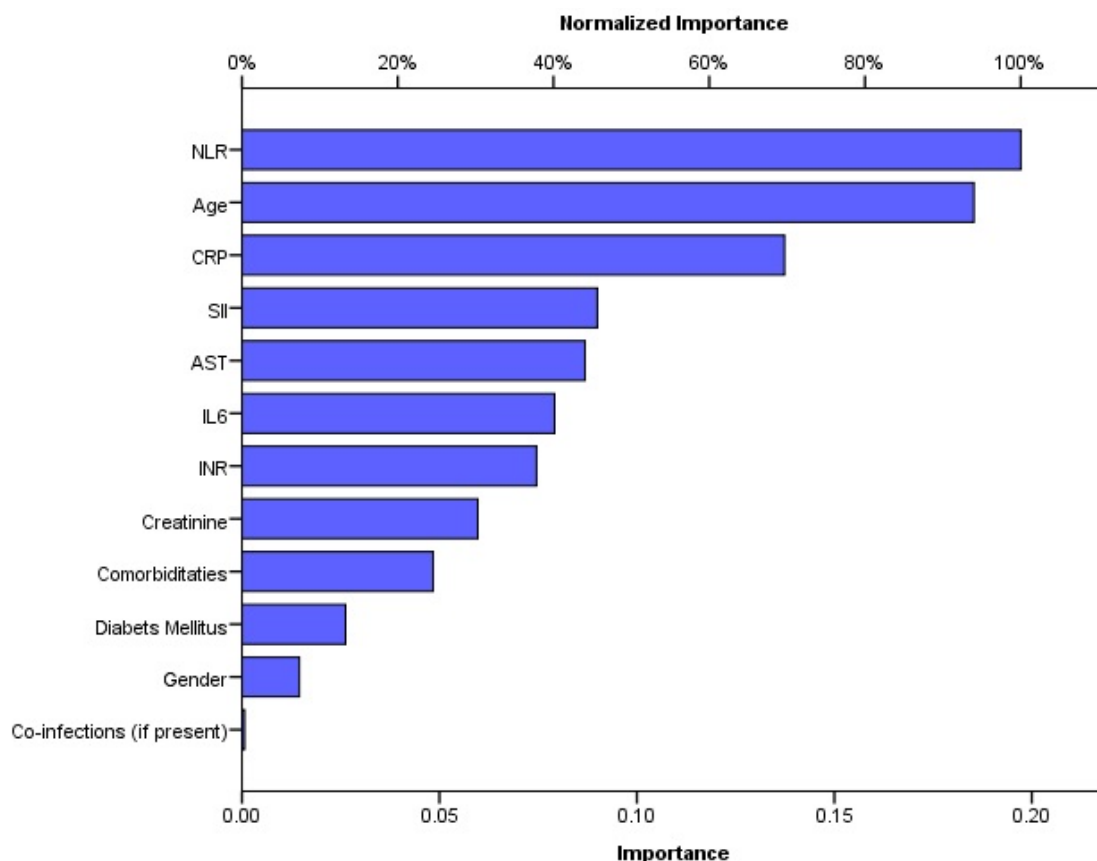


Fig. 5. The normalized importance of the independent variables for Neural Network analysis.

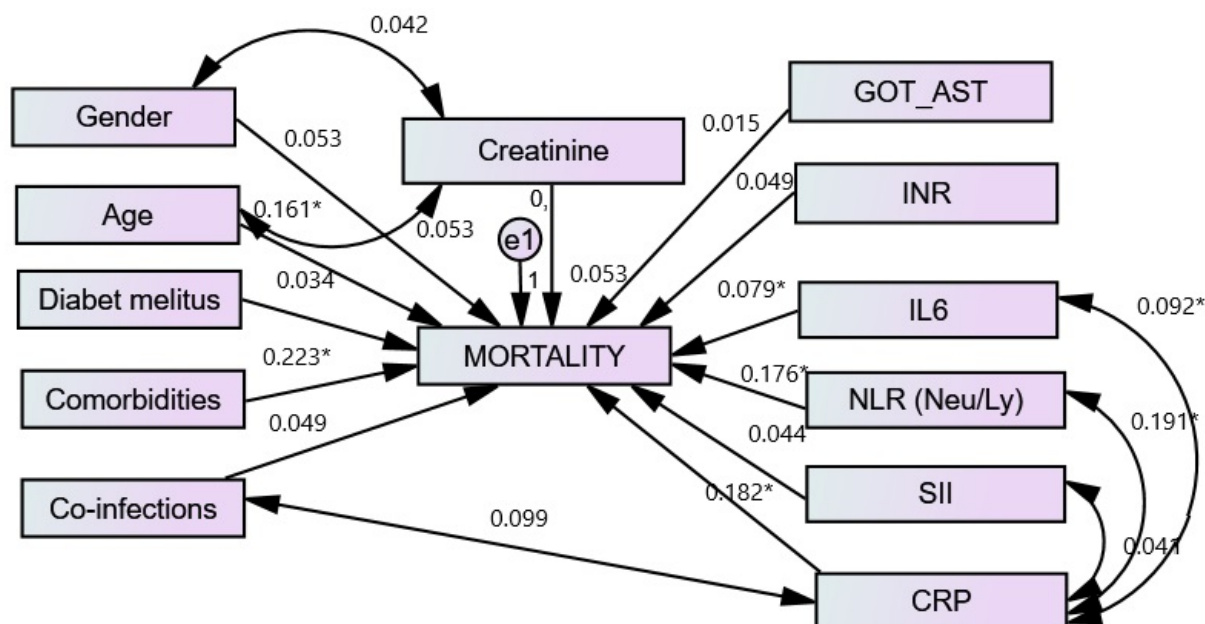


Fig. 6. The structural SEM model analysis showing the causality between variables and the influence of variables with direct or indirect effect on mortality. e1=endogenous variables (mortality), the numbers represent "standardized" regression coefficients with $p \leq 0.05$ for comorbidities, age, CRP, IL-6, and NLR.

(coeff = 0.223, $p = 0.001$), CRP (coeff = 0.182, $p = 0.001$), NLR (coeff = 0.176, $p = 0.029$), age (coeff = 0.161, $p = 0.001$) and IL-6 (coeff = 0.079, $p = 0.001$). Based on the co-variances, the variables with indirect effects on "mortality" can be observed as being creatinine, hospital-associated infections, and gender.

For standardized SEM the goodness-of-fit statistics of the model for the estimated model were considered as follows:

- CMIN/DF (the relative chi-squared) was = 12.612, which means that only 13 % of the fit of data is

reduced by dropping one or more paths.

- Comparative fit index (CFI) = $0.812 \approx 0.9$, the estimated model has no large fit variations from the observed values.
- RMSEA (Root Mean Square Error of Approximation) is $= 0.171 \approx 0.1$ which means that a small percentage was lost between the proposed model and the resulting one. The overall percentage for prediction was 76%.

DISCUSSION

Inflammatory markers and other routine laboratory parameters could represent valuable tools for mortality risk stratification in COVID-19. In many studies, NLR and/or SII are included in the development of a risk score to discriminate severe from mild disease forms or to appreciate the risk of mortality in patients with SARS-CoV-2 infection.

Thus, in our study we retrospectively evaluated the routine biochemical and hematological parameters in critically ill COVID-19 patients admitted to ICU and observed that the NLR value above the threshold of 8.28 determined a 4.6 times greater risk of death compared to patients with NLR values below the established threshold, after adjustment for the considered variables. Comparable results were reported by other studies. A NLR value above 6.82 as reported by Prozan and collab. [13] and a NLR value above 9.1 as reported by Citu et al. were considered unfavorable prognostic factors in COVID-19 patients [14], being the prerogative of an aberrant immune response with increased neutrophils and decreased lymphocytes. NLR was found to be an important risk factor for in-hospital mortality in SARS-CoV-2 infected patients, and an increase of 1 unit for NLR enhances the risk of mortality in COVID-19 hospitalized patients by 8% [15]. Additionally, in geriatric patients with SARS-CoV-2 infection, the 30-day mortality prediction was in relation to NLR, values above 7.8 were found predictive with 83.3% sensitivity and 97.7% specificity [16].

Elevated serum levels of pro-inflammatory cytokines, including IL-6, have been identified in patients with severe COVID-19 compared to individuals with mild disease forms [6], suggesting the key role of a hyperinflammatory response in the pathogenesis and evolution of this viral infection [17]. In a retrospective study performed by Pál et al. on 117 critically ill patients, the Kaplan-Meier survival analysis estimated that patients with IL-6 serum levels above 27.6 pg/ml had a 2.17 greater risk of death compared to patients with lower IL-6 levels [18]. Similarly, in our study, patients with unfavorable outcomes

had twice higher values for IL-6 compared to survivors, and the estimated risk of death expressed by adjusted OR was almost 4-fold higher for patients with high IL-6 serum levels. This finding emphasizes the role of the dysregulated inflammatory response during SARS-CoV-2 infection, mostly with the D variant. Therefore, more than 66% of COVID-19 patients died due to successive alterations of the inflammatory process, while the rest of 33% contracted other in-hospital infections as a result of inappropriate immune response during the evolution of the COVID-19 disease. Similarly, serum levels of CRP, ferritin and the hematological-derived SII were higher in patients who did not survive the infection, similar results being reported by Guan et al. [19].

IL-6 was documented as a strong mortality predictor with even better sensitivity and specificity compared to procalcitonin (PCT), as reported by Andrijevic et al. The authors reported 93.4% risk of mortality for patients with IL-6 above a cut-off of 20.2 pg/ml, close to what we found in our study, compared to only 66.7% risk of mortality for elevated PCT [20].

Although IL-6 is a non-specific proinflammatory marker increasing in many diseases, it is essential that its elevation during severe infections to be highlighted for a prompt antagonist IL-6 intervention to be applied in those patients, by targeting the cytokine *per se*, its cognate membrane receptor or the soluble one [21]. It is also important to mention that IL-6 has a dual role in the inflammatory process, aside from the detrimental role, IL-6 is also an important factor in the immune response integration and in sustaining immunocompetence [22].

Our results revealed IL-6 as a significant in-hospital predictor for ICU patients with SARS-CoV-2 infection in conjunction with other biomarkers, age above 65 years, and associated medical conditions. Elevated IL-6 was found to be a predictor of mortality, in strong relation to patient's age above 65 years and CURB65 (confusion, blood urea nitrogen, raised respiratory rate, low blood pressure, and age above 65 years) score [23].

One of the reasons explaining why increased levels of IL-6 are associated with increased mortality in severe COVID-19 patients is related to immune system hyperactivation and cytokine storm amplified by IL-6 signaling complex that activates STAT3 pathways and consecutive NF- κ B pathway in non-immune cells, as reviewed by Hojyo et al. [24]. After the complex of IL-6 and soluble IL-6 receptor is constituted, a positive IL-6 mediated loop is created for hyperactivation of STAT3 and NF- κ B signaling pathways, known as IL-6 Amplifier [24]. The involvement of the IL-6 Amplifier in the cytokine storm and subsequent fatal outcome was underlined in patients with CAR-T therapy where cytokine release syndrome was countered by blocking the IL-6 receptor [25].

Another reason why IL-6 is associated with fatal outcomes resides in its relationship with the innate immune response. As is already known, a strong innate immune response will favor a rapid viral clearance, while an impaired early immune response will facilitate viral persistence and disease progression. As Rodrigues and her colleagues reviewed, the INF response is abnormal in severe SARS-CoV-2 infection. In addition, excessive cytokine production, including IL-6 is closely related to complement activation, dysfunction of endothelial cells, and thrombus formation, leading to multiple organ damage [26]. Increased complement activity under excessive IL-6 levels results in forming of excessive neutrophil extracellular traps (NETs), consisting of denatured DNA strings or chromatin and toxic granules, and having the role of trapping and inactivating the viruses [26]. Far from being a beneficial factor, NETosis induces thrombotic events which were identified as key factors in lung and other organ failures in SARS-CoV-2 infected patients. Additionally, increased levels of IL-6 could predict the fatal outcome in patients having an 'exhausted' NK phenotype [26].

During SARS-CoV-2 infection, AST and ALT activities are found to be only transiently increased; the mechanism underlying severe liver dysfunction mainly occurs due to secondary inflammation-induced liver damage rather than direct liver damage [27]. Our study revealed that the central tendency of AST in the patients group with unfavorable outcomes was significantly higher compared to surviving patients and that patients with higher AST activity had three times lower survival chance, estimated by Odds Ratio adjusted for age, gender, diabetes mellitus, and hospital-associated infections. Our results are in line with results found in the specialized literature, a recent meta-analysis including 32 initial studies revealed that elevated AST activity is associated with almost three times higher mortality risk in patients with COVID-19, results validated in 18 of the 32 studies, including a total of 6383 patients [27]. A recent article, based on a retrospective study assessing the hepatic involvement in COVID-19 severe evolution, evaluated patients without prior hepatic pathology and revealed that both increased AST and hypoalbuminemia were independent predictors for disease severity and mortality in COVID-19 patients [28]. Moreover, AST value on hospital admission was a good mortality predictor for COVID-19 patients [29]. Aside from the increased AST activity, other biochemical parameters associated with liver damage were altered in severe SARS-CoV-2 infections, according to a recent meta-analysis [30]. In our study, patients with unfavorable outcomes also had significantly lower albumin and total protein serum levels and higher urea levels, while direct and total bilirubin, as well as ALT activity were all without notable differences compared to survivors. A

greater increase in AST versus ALT activity was registered in patients with severe infection, and given the fact that ALT is a more specific marker of hepatocyte injury, while AST is found in many other tissues, it is likely that these changes appear as a result of multiorgan involvement during SARS-CoV-2 infection.

Hypoalbuminemia was also evidenced in COVID-19-infected patients; it is well-known that albumin is a protein synthesized in the liver with a serum half-life of approximately 21 days, being the main source of amino acids for tissue reconstructive actions/reactions [31]. It has been reported that hypoalbuminemia was seen predominantly in severe COVID-19 cases compared to mild cases, probably due to the systemic inflammation, and not necessarily due to hepatocellular dysfunction alone, as supported by the fact that AST and ALT were not found to have predictive value for the outcome in the study performed by Huang and collab [32].

A strong relation between mortality and the age of patients was found by Grasselli et al., a 10-year increase in patient age was found to be significantly associated with mortality, and patients over 64 years of age had significantly lower survival probability compared to younger patients [33]. Also, the risk of death among patients older than 50 years was 2.76 times higher compared to younger patients after adjustment [34]. Similarly, in our study according to SEM analysis, age above 66.21 years was associated with an increased mortality rate.

Regarding renal function, a prevalent complication reported in the literature in patients with COVID-19 infection is acute kidney injury (AKI) [35], with an incidence in hospitalized patients of over 20% and over 50% incidence in intensive care units [36]. When AKI occurs, the dialysis rates may increase to 30%, with the survival rate significantly reduced [37]. According to a recent study developed by Yildirim et al., it would be recommended that albuminuria be routinely evaluated using the albumin/creatinine ratio in spot urine on hospital admission for COVID-19-related AKI patients who do not meet AKI criteria on hospital admission [38].

Serum creatinine is a biomarker that reflects renal function, this function being altered through a series of immunological and pathophysiological mechanisms in the pathology of COVID-19. In our study, the chance of death for patients with elevated creatinine levels was almost three times higher, adjusted OR 2.540 (CI=95%: 1.668- 3.869), $p=0.000$. In a prospective cohort study of 701 patients with COVID-19 infection, it was found that during hospitalization, the incidence of acute kidney injury and patient death was significantly higher in patients with elevated baseline serum creatinine levels compared to patients with normal baseline values.

The mechanism by which this occurs is speculated to be through hematogenous spread and accumulation of the virus in the kidney, causing necrosis of the kidney cells [39].

A high amount of data and information are generated for a single patient during hospitalization, especially in severe COVID-19 patients, where accommodation in ICU lasts for at least 10 days in the most favorable scenarios. In this regard, clinicians must deal with a plethora of results and appreciate the importance of modified laboratory parameters in the clinical context of each individual. However, by using Machine Learning methods (such as decision trees or neural networks), a better perspective of how all these results might impact the clinical evolution could be gained, helping doctors in critical clinical settings to decipher the best approach for each patient. It is acknowledged that diagnostic charts or clinical diagnostic pathways are popular among physicians. In this light, the results generated by Machine Learning methods could be particularly useful, since the software will be able to recognize a pattern after running the training set, group data with similar characteristics determining a similar outcome and finally generate an algorithm. In contrast to this method, the conventional statistical tests assume that the data and outcome are to some degree known and the model is created by the user [40].

The advantages of using decision tree compared with classical statistical methods include:

- Allowing the prediction of individuals to distinct categories, based on their measures according to one or more predictor variables [7,41]
- Allowing utilization of both categorical and continuous types of data by using different algorithms (CHAID, CRT – Classification and Regression Tree)
- Classifying the individuals in homogenous groups by independent variables.
- Helping to identify characteristics of group in relation to independent variables regarding the dependent variable.

The advantages of using SEM are:

- Enabling researchers to simultaneously model and estimate complex relations among multiple dependent and independent variables [10],
- Obtaining a more precise measurement of the theoretical concepts of interest [42]
- Working efficiently with a small sample size with no normal distribution assumption needed, [10] SEM being in fact a non-parametrical method.

The limits of the research are linked to: (1) the relatively small number of samples (survivors and non-survivors), the international literature recommends a number of 100 cases for the analyzed sample size, while our research includes 396 cases (183 non-survivors and 213 survivors) and therefore could be considered a pilot research for in-hospital mortality; (2) the validation of machine learning methods, especially for decision tree through the crossvalidation method by using minimum values of 10% for testing and 90% for training, because the risk in the output is the average risk of all the trees.

We consider that it is crucial to find as many useful biomarkers as possible for disease evolution and mortality prediction, and that this aspect is especially important for IL-6, which is the trigger of the cytokine storm with devastating effects on patients with SARS-CoV-2 infection. On the other hand, it is important that its elevation during severe infections be highlighted for a prompt blocking of IL-6 action to be applied. Hence, we consider the investigation of IL-6, along with other inflammatory biomarkers, of great importance in patients with severe forms of COVID-19, since the therapeutic approach could represent the turning point in the management of the ICU patient.

Furthermore, the results we obtained using an advanced statistical approach (including decision trees and neural networks, such as machine learning methods and Structural Equation Modeling) could serve as a model for further research and suggest identifying prognostic biomarkers in other conditions with severe evolution.

CONCLUSIONS

Monitoring of hospitalized COVID-19 patients generates a tremendous number of parameters, some of them of great clinical importance. The decision tree analysis could represent a valuable tool to assist clinicians in patient management and to provide clinical decision support. In our study, we found that inflammatory biomarkers, as well as the renal and hepatic function, were altered in patients with severe COVID-19 and additionally using a statistical algorithm that utilized the CHAID method we found that NLR, IL-6 and AST were the main in-hospital mortality predictors for critically ill COVID-19 patients. Additionally, after the sample-split validation method was applied, IL-6 remained the best predictor for in-hospital mortality, for both training and test samples. Using a combination of Machine Learning methods together with statistical methods ensures identifying the most accurate predictors of in-hospital mortality.

ABBREVIATIONS

AKI – Acute Kidney Injury
 ALT – Alanine Aminotransferase
 AST – Aspartate Aminotransferase
 ANN- Artificial Neural Network Analysis
 AUC – Area Under the Curve
 CAR-T- Chimeric Antigen Receptor T-cell therapy
 CBC – Complete Blood Count
 CFI – Comparative Fit Index
 CHAID – Chi-Square Automatic Interaction Detector
 CO – Cut-off
 CRP – C-reactive Protein
 CRT – Classification and Regression Tree
 eGFR- Estimated Glomerular Filtration Rate
 hsCRP – high-sensitive C-reactive Protein
 ICU – Intensive Care Unit
 IL-6 – Interleukin-6
 INR – International Normalized Ratio
 IQR – Interquartile Range
 LDH – Lactate Dehydrogenase
 LIS – Laboratory Informational System
 MCV – Mean Erythrocyte Volume
 MCH- Mean Erythrocyte Hemoglobin
 MCHC- Mean Erythrocyte Hemoglobin Concentration
 MLP – Multilayer Perceptron
 MqSOFA – Modified quick Sequential Organ Failure Assessment
 NETs – Neutrophil Extracellular Traps
 NEWS2 – National Early Warning Score 2
 NF- κ B – Nuclear Factor Kappa B
 NLR – Neutrophils to Lymphocyte Ratio
 OR – Odds Ratio
 PLT – Platelets
 PT – Prothrombin Time
 RBC – Red Blood Cells
 RMSEA – Root Mean Square Error of Approximation
 ROC – Receiver Operating Characteristic
 SD – Standard Deviation
 SEM – Structural Equation Modeling
 SII – Systemic Inflammatory Index
 SIRS – Systemic Inflammatory Response Syndrome
 SpO₂ – Oxygen Saturation
 SPSS – Statistical Package for the Social Sciences

STAT-3 – Signal Transducer and Activator of Transcription 3

WBC – White Blood Cells

ACKNOWLEDGMENT

This work was supported by George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Târgu Mureș Romania, Research Grant number 10126/1/17.12.2020.

We would like to express our sincere gratitude to Prof. Marius Mărușteri for his invaluable recommendations and experienced insights that have significantly contributed to refining and improving this work.

The authors would also like to thank Dr. Ionut Branea and Dr. Timár Ágota-Evelyn for their important contribution to this study.

AUTHORS' CONTRIBUTION

AH – Conceptualization and methodology; writing the original draft; writing-reviewing and editing

AAM – Data curation; writing the original draft; writing-reviewing and editing

KP – Data curation; writing-reviewing and editing

MRG – Data curation; software and statistical analysis; writing-reviewing and editing

JS – Conceptualization and methodology; software and statistical analysis; writing-reviewing and editing

MD – Conceptualization and methodology; writing-reviewing and editing

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PUBLISHER'S NOTE

Editorial Office stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

REFERENCES

1. Wolff D, Nee S, Hickey NS, Marschollek M. Risk factors for Covid-19 severity and fatality: a structured literature review. *Infection*. 2021;49(1):15-28. DOI: 10.1007/s15010-020-01509-1
2. Myrstad M, Ihle-Hansen H, Tveita AA, Andersen EL, Nygård S, Tveit A, et al. National Early Warning Score 2 (NEWS2) on admission

- predicts severe disease and in-hospital mortality from Covid-19 - a prospective cohort study. *Scand J Trauma Resusc Emerg Med.* 2020;28(1):66. DOI: 10.1186/s13049-020-00764-3
3. Guarino M, Perna B, Remelli F, Cuoghi F, Cesaro AE, Spampinato MD, et al. A New Early Predictor of Fatal Outcome for COVID-19 in an Italian Emergency Department: The Modified Quick-SOFA. *Microorganisms.* 2022;10(4):806. DOI: 10.3390/microorganisms10040806
 4. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision Trees: An Overview and Their Use in Medicine. *J Med Syst.* 2002;26(5):445-63. DOI: 10.1023/A:1016409317640
 5. Karthikeyan A, Garg A, Vinod PK, Priyakumar UD. Machine Learning Based Clinical Decision Support System for Early COVID-19 Mortality Prediction. *Front Public Heal.* 2021;9:626697. DOI: 10.3389/fpubh.2021.626697
 6. Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and laboratory predictors of in-hospital mortality in patients with COVID-19: a cohort study in Wuhan, China. *Clin Infect Dis.* 2020;71(16):2079-88. DOI: 10.1093/cid/ciaa538
 7. McCormik K, Salcedo J, Peck J. WA. SPSS Statistics for Data Analysis and visualization. Indianapolis, IN: John Wiley & Sons, Inc. 2017:355-92.
 8. Hsu M. Structural Equation Modeling with IBM SPSS Amos. In: IBM Corporation, editor. IBM Software Business Analytics. Somers NY; 2010.
 9. Barclay D., Higgins C, Thompson R. The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration. *Technol Stud.* 1995;2(2):285-309.
 10. Hair, Jr., J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., Ray, S. Partial least squares structural equation modeling (PLS-SEM) using R- a workbook, SPRINGER, 2021. DOI: 10.1007/978-3-030-80519-7
 11. Khabaza T. 9 Laws of Data Mining [Internet]. 2010 [cited 2023 Apr 17]. Available from: http://khabaza.codimension.net/index_files/9laws.htm
 12. Jöreskog KG, Wold H. The ML and PLS technique for modeling with latent variables: historical and comparative aspects. In Wold H, Jöreskog KG (Eds.) *Systems under indirect observations, part I.* Amsterdam, North-Holland. 1982:263-270.
 13. Prozan L, Shusterman E, Ablin J, Mitelpunkt A, Weiss-Meilik A, Adler A, et al. Prognostic value of neutrophil-to-lymphocyte ratio in COVID-19 compared with Influenza and respiratory syncytial virus infection. *Sci Rep.* 2021;11(1):21519. DOI: 10.1038/s41598-021-00927-x
 14. Cîțu C, Gorun F, Motoc A, Sas I, Gorun OM, Burlea B, et al. The Predictive Role of NLR, d-NLR, MLR, and SIRI in COVID-19 Mortality. *Diagnostics.* 2022;12(1):122. DOI: 10.3390/diagnostics12010122
 15. Liu Y, Du X, Chen J, Jin Y, Peng L, Wang HHX, et al. Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19. *J Infect.* 2020;81(1):e6-12. DOI: 10.1016/j.jinf.2020.04.002
 16. Önal U, Gülhan M, Demirci N, Özden A, Erol N, Işık S, et al. Prognostic value of neutrophil-to-lymphocyte ratio (NLR) and lactate dehydrogenase (LDH) levels for geriatric patients with COVID-19. *BMC Geriatr.* 2022;22(1):1-6. DOI: 10.1186/s12877-022-03059-7
 17. Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. *Clin Infect Dis.* 2020;71(15):762-8. DOI: 10.1093/cid/ciaa248
 18. Pál K, Molnar AA, Huțanu A, Szederjesi J, Branea I, Timár Á, et al. Inflammatory Biomarkers Associated with In-Hospital Mortality in Critical COVID-19 Patients. *Int J Mol Sci.* 2022;23(18):10423. DOI: 10.3390/ijms231810423
 19. Guan X, Zhang B, Fu M, Li M, Yuan X, Zhu Y, et al. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a retrospective cohort study. *Ann Med.* 2021;53(1):257-66. DOI: 10.1080/07853890.2020.1868564
 20. Andrijevic I, Matijasevic J, Andrijevic L, Kovacevic T, Zaric B. Interleukin-6 and procalcitonin as biomarkers in mortality prediction of hospitalized patients with community acquired pneumonia. *Ann Thorac Med.* 2014;9(3):162-7. DOI: 10.4103/1817-1737.134072
 21. Jones SA, Hunter CA. Is IL-6 a key cytokine target for therapy in COVID-19? *Nat Rev Immunol.* 2021;21(6):337-9. DOI: 10.1038/s41577-021-00553-8
 22. Rose-John S, Winthrop K, Calabrese L. The role of IL-6 in host defence against infections: immunobiology and clinical implications. *Nat Rev Rheumatol.* 2017;13:399-409. DOI: 10.1038/nrrheum.2017.83
 23. Guirao JJ, Cabrera CM, Jiménez N, Rincón L, Urra JM. High serum IL-6 values increase the risk of mortality and the severity of pneumonia in patients diagnosed with COVID-19. *Mol Immunol.* 2020;128:64-8. DOI: 10.1016/j.molimm.2020.10.006
 24. Hojyo S, Uchida M, Tanaka K, Hasebe R, Tanaka Y, Murakami M, et al. How COVID-19 induces cytokine storm with high mortality. *Inflamm Regen.* 2020;40:37. DOI: 10.1186/s41232-020-00146-3
 25. Hirano T, Murakami M. COVID-19: A New Virus, but a Familiar Receptor and Cytokine Release Syndrome. *Immunity.* 2020;52(5):731-3. DOI: 10.1016/j.immuni.2020.04.003
 26. Rodrigues PRS, Alrubayyi A, Pring E, Bart VMT, Jones R, Coveney C, et al. Innate immunology in COVID-19-a living review. Part II: dysregulated inflammation drives immunopathology. *Oxford Open Immunol.* 2020;1(1):iqaa005. DOI: 10.1093/oxfimm/iqaa005
 27. Malik P, Patel U, Mehta D, Patel N, Kelkar R, Akrmah M, et al. Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis. *BMJ Evidence-Based Med.* 2021;26(3):107-8. DOI: 10.1136/bmjebm-2020-111536
 28. Kodavoor Vadiraj P, Thareja S, Raman N, Karantha SC, Jayaraman M, Vardhan V. Does Raised Transaminases Predict Severity and Mortality in Patients with COVID 19? *J Clin Exp Hepatol.* 2022;12(4):1114-23. DOI: 10.1016/j.jceh.2022.01.004
 29. Pozzobon FM, Perazzo H, Bozza FA, Rodrigues RS, de Mello Perez R, Chindamo MC. Liver injury predicts overall mortality in severe COVID-19: a prospective multicenter study in Brazil. *Hepatol Int.* 2021;15(2):493-501. DOI: 10.1007/s12072-021-10141-6
 30. Parohan M, Yaghoubi S, Seraji A. Liver injury is associated with severe coronavirus disease 2019 (COVID-19) infection: A systematic review and meta-analysis of retrospective studies. *Hepatol Res.* 2020;50(8):924-35. DOI: 10.1111/hepr.13510
 31. Rothschild MA, Oratz M, Schreiber SS. Serum albumin. *Hepatology.* 1988;8(2):385-401. DOI: 10.1002/hep.1840080234
 32. Huang J, Cheng A, Kumar R, Fang Y, Chen G, Zhu Y, et al.

- Hypoalbuminemia predicts the outcome of COVID-19 independent of age and co-morbidity. *J Med Virol.* 2020;92(10):2152. DOI: 10.1002/jmv.26003
33. Grasselli G, Greco M, Zanella A, Albano G, Antonelli M, Bellani G, et al. Risk Factors Associated With Mortality Among Patients With COVID-19 in Intensive Care Units in Lombardy, Italy. *JAMA Intern Med.* 2020;180(10):1345-55. DOI: 10.1001/jamainternmed.2020.3539
 34. Leulseged TW, Hassen IS, Ayele BT, Tsegay YG, Abebe DS, Edo MG, et al. Laboratory biomarkers of COVID-19 disease severity and outcome: Findings from a developing country. *PLoS One.* 2021;16(3): e0246087. DOI: 10.1371/journal.pone.0246087
 35. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA.* 2020;323(11):1061-9. DOI: 10.1001/jama.2020.1585
 36. Nadim MK, Forni LG, Mehta RL, Connor MJ, Liu KD, Ostermann M, et al. COVID-19-associated acute kidney injury: consensus report of the 25th Acute Disease Quality Initiative (ADQI) Workgroup. *Nat Rev Nephrol.* 2020;16(12):747-764. DOI: 10.1038/s41581-020-00356-5
 37. Fabrizi F, Alfieri CM, Cerutti R, Lunghi G, Messa P. COVID-19 and Acute Kidney Injury: A Systematic Review and Meta-Analysis. *Pathogens.* 2020;9(12):1-16. DOI: 10.3390/pathogens9121052
 38. Yildirim C, Ozger H S, Yasar E, Tombul N, Gulbahar O, et al. Early predictors of acute kidney injury in COVID-19 patients. *Nephrology (Carlton).* 2021;26(6), 513-521. DOI: 10.1111/nep.13856
 39. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int.* 2020;97(5):829-38. DOI: 10.1016/j.kint.2020.03.005
 40. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. *Knee Surgery, Sport Traumatol Arthrosc.* 2022;30(3):753-7. DOI: 10.1007/s00167-022-06896-6
 41. Rakotomalala R. Arbres de Décision. *Rev Modul.* 2005;33:163-87.
 42. Cole DA, Preacher KJ. Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychol Methods.* 2014;19(2):300-15. DOI: 10.1037/a0033805