Check for updates

ORIGINAL ARTICLE

Machine learning for COVID-19 mortality prediction: enhancing cart models with node-specific odds ratios

Aprendizaje automático para la predicción de mortalidad por COVID-19: mejora de los modelos CART con razones de probabilidad específicas por nodo

*Argelia Pérez-Pacheco[1,3], Alejandra E. Herrera-Suárez[1], Jonathan Vélez-Mata[1], Liza M. Magdaleno-Fourlong[1], Amy D. Avendaño-Carrera[1], Catalina Casillas-Suárez[2], and Adolfo Pérez-García[3]**

[1]Research and Technological Development Unit; [2]Pulmonology Service; [3]Department of Research. Hospital General de Mexico "Dr. Eduardo Liceaga", Mexico City, Mexico

Abstract

Objective: *The objective of the study is to evaluate a classification and regression tree (CART) model combined with odds ratio (OR) to identify key predictors of COVID-19 mortality. Data from 1,432 patients hospitalized at Hospital General de México during the pandemic's 1st year were analyzed.* **Methods:** *A CART model was constructed using demographic, clinical, and laboratory data collected at admission. The model's performance was evaluated using multiple criteria, and ORs were calculated for each node to measure mortality.* **Results:** *Mechanical ventilation emerged as the strongest predictor of mortality. Patients intubated at admission with hospital stays under 17.5 days had the highest mortality rate (99%, OR = 80). Conversely, non-intubated patients hospitalized over 5.5 days with glomerular filtration rates above 66.5 had the lowest mortality (5%, OR = 0.26). The model showed excellent performance, with an F1 score of 0.918, accuracy of 0.903, and area under the curve of 0.955.* **Conclusions:** *The CART model, combined with OR calculations, offers a reliable and comprehensible tool for predicting COVID-19 mortality risk. This model provides practical utility in various healthcare settings, including resource-limited contexts, by focusing on readily available clinical parameters. The results emphasize the significant influence of mechanical ventilation on patient outcomes and the need for timely interventions in high-risk patients.*

Keywords: *COVID-19. Mortality prediction. Classification and regression tree model. Mechanical ventilation. Machine learning. Odds ratio.*

Resumen

Objetivo: *Evaluar un modelo CART con razones de probabilidad para identificar predictores de mortalidad por COVID-19 en 1432 pacientes hospitalizados.* **Métodos:** *Se construyó un modelo CART con datos demográficos, clínicos y de laboratorio al ingreso. Se evaluó el rendimiento y se calcularon las razones de probabilidad para medir el riesgo de mortalidad.* **Resultados:** *La ventilación mecánica fue el predictor más fuerte. Los pacientes intubados al ingreso con estancias de 5.5 días con filtración glomerular > 66.5 tuvieron la menor mortalidad (5%; OR: 0.26). El modelo mostró un excelente rendimiento (F1 = 0.918, precisión de 0.903, AUC = 0.955).* **Conclusiones:** *El modelo CART con razones de probabilidad ofrece una herramienta confiable para predecir la mortalidad por COVID-19, útil en diversos entornos, incluso con recursos limitados. Enfatiza la influencia de la ventilación mecánica y la necesidad de intervenciones oportunas en pacientes de alto riesgo.*

Palabras clave: *COVID-19. Predicción de mortalidad. Modelo CART. Ventilación mecánica. Aprendizaje automático. Razones de probabilidad.*

***Correspondence:**
Adolfo Pérez-García
E-mail: adolfo.perez@salud.gob.mx

Date of reception: 20-02-2025
Date of acceptance: 23-04-2025
DOI: 10.24875/CIRUE.M25000948

Cir Cir (Eng). 2026;94(3):295-303
Contents available at PubMed
www.cirugiaycirujanos.com

2444-0507/© 2025 Academia Mexicana de Cirugía. Published by Permanyer. This is an open access article under the terms of the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

In December 2019, Wuhan, China, became the epicenter of a global health crisis because of the emergence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. This novel virus, which causes coronavirus disease 2019 (COVID-19), belongs to the coronavirus family – a group of viruses capable of infecting both humans and animals. Coronaviruses are associated with several illnesses, from common colds to severe respiratory conditions, such as severe acute respiratory syndrome (SARS) and Middle East Respiratory Syndrome.[1] A frequent and severe complication of these infections is acute hypoxemic respiratory failure or acute respiratory distress syndrome (ARDS).[2] ARDS is a diffuse inflammatory lung injury causing low blood oxygen levels; it is defined as a PaO_2 under 60 mmHg or a SaO_2 under 88%.[3] Pulmonary inflammation cytokines and other proinflammatory molecules damage the capillary endothelium and alveolar epithelium, disrupting the barrier between the capillaries and air spaces. As a result, patients with ADRS suffer severe respiratory distress that requires oxygenation therapy and respiratory support.[4,5]

On September 25, 2022, 612 million confirmed cases and 6.5 million deaths were recorded due to COVID-19 worldwide. In Mexico alone, the global report included 7,702,809 confirmed cases and 334,958 deaths.[6] Between 2020 and 2021, mortality in Mexico secondary to COVID-19 accounted for 47% of all deaths registered in the country, according to data from the National Institute of Statistics and Geography.

Globally, the most important risk factors influencing the progression and severity of infection include advanced age and comorbidities, such as diabetes, hypertension, cardiovascular disease, chronic lung disease, and chronic kidney disease.[7,8] The global prevalence of overweight and obesity remains alarmingly high, affecting 62.5% of adults, with 64.1% of men and 60.9% of women falling into this category.[9,10]

During the COVID-19 pandemic, a body mass index (BMI) > 35 kg/m^2 was strongly associated with an increased risk of intensive care unit (ICU) admission. Studies have noted that this BMI threshold not only heightened the likelihood of ICU admission but also increased the need for mechanical ventilation and prolonged hospital stays. Furthermore, overweight and obese individuals were found to have a greater likelihood of developing ARDS, accompanied by higher mortality rates.[11-13]

At the peak of the pandemic, most patients hospitalized for COVID-19 and/or ARDS required ventilatory support because of advanced disease progression. Ventilatory support is a critical intervention for managing respiratory failure by ensuring adequate oxygen delivery and gas exchange.[14] Various ventilatory support methods exist, including non-invasive ventilation and invasive mechanical ventilation, each with distinct benefits and risks.[15]

The decision between non-invasive and invasive ventilation strategies is influenced by several factors, such as the severity of respiratory failure, patient response to initial treatment, and the presence of comorbidities. Although intubation and invasive mechanical ventilation remain contentious due to their association with high morbidity and mortality rates (reported between 81% and 97%), they often play a pivotal role in managing severe cases.[5,16-19]

Another critical predictor of COVID-19 mortality was the duration of hospitalization. Prolonged hospital stay was found to triple the risk of mortality compared with shorter stays.[20] However, Sousa Neto et al.[21] observed that longer ICU stays paradoxically served as a protective factor for certain patients. This finding suggests that extended ICU care may benefit those admitted with extreme disease severity, improving outcomes despite the high initial risk.

The rapid spread of the virus, its capacity for mutation, patient-specific factors, and the natural history of the disease have led to a severe shortage of medical resources and trained personnel. This crisis underscores the urgent need for risk stratification models to optimize healthcare delivery.

Predictive models provide a means to identify patients at higher risk of mortality, enabling timely and targeted interventions. Machine learning (ML) classifiers have proven to be valuable tools for supporting clinical decision-making. As a branch of artificial intelligence, ML leverages large raw datasets to develop high-quality predictive models. Among these, Classification and Regression Trees (CART) are ML-based AI tools that utilize statistical, probabilistic, and optimization techniques to integrate risk factors and generate profiles of patients at high risk of mortality.[22,23] CART models have been applied in various medical domains, including survival analysis,[24] dyslipidemia classification,[25] and prognosis of heart failure.[26] CART can handle non-linear relationships through recursive

partitioning, creating splits based on different thresholds of the predictor variables.[27] Importantly, CART can manage both categorical and continuous variables without requiring extensive preprocessing, which enhances its usability in diverse contexts.[28]

This study aimed to identify the key predictors of mortality in patients admitted to the Hospital General de México "Dr. Eduardo Liceaga" (HGMEL) during the 1st year of the COVID-19 pandemic. Using available clinical and laboratory data, the research employs an easy-to-use and interpretable model developed with accessible software capable of presenting results in a user-friendly, visual format. This approach ensures its applicability in real-world healthcare settings. Identifying predictors of mortality in hospitalized patients with COVID-19 is crucial for improving patient survival and optimizing healthcare resources. By focusing on actionable insights, this study seeks to improve healthcare service delivery and patient outcomes.

Methods

This retrospective study was conducted at the HGMEL. A total of 1432 patients of both sexes, older than 18 years, positive for SARS-CoV-2 by polymerase chain reaction, hospitalized from the second quarter of 2020 to the fourth quarter of 2021, were included. Data were collected upon admission. Demographical data and laboratory test results were registered; the latter consisted of complete blood counts, blood gases, and chemical blood analyses.

The subjects were divided into two groups according to the main outcome: survivors and non-survivors. For the comparison of the variables between groups, the appropriate statistical test was used according to the normality of the data, taking a significant difference $p > 0.05$.

Research involving human subjects complied with all relevant national regulations and institutional policies, which are in accordance with the tenets of the Helsinki Declaration (as revised in 2013), and was approved by the Research Ethics Committee of the HGMEL (Reg. No. DMC-3369-20-20-1).

CART model

The variables to introduce in the model were selected following a literature survey and consensus among experts regarding the most pertinent patient characteristics that could predict mortality. The variables introduced were sex, age, days of hospital stay, BMI, type 2 diabetes mellitus, systemic arterial hypertension, glucose (mg/dL), cardiac rate in beats per minute (CR), respiratory rate in breaths per minute (RR), oxygen saturation in percentage (SpO_2), lactate dehydrogenase (LDH) in IU/L, glomerular filtration rate (GFR), hemoglobin (g/dL), leukocytes ($\times 10^3$), neutrophil/leukocyte ratio (NLR), invasive ventilatory support, and rejection of invasive ventilatory support. Subsequently, the variables underwent an analysis for multicollinearity employing the variance inflation factor approach, with all scoring between 1 and 5. Three variables, specifically BMI, HR, and RR rate, exhibited more than 10% of missing values (10.2, 12.9, and 12.8, respectively). These missing values amounts were considered acceptable on the one hand because even with some missing values, the sample size maintains sufficient statistical power for analysis, and on the other hand because CART models can handle missing data through surrogate splits, a mechanism that uses alternative splitting rules when the primary splitting variable is missing, and maintains all observations in the analysis, even those with missing values,[29] therefore identifying important relationships between variables that could be obscured by imputation.[30] The tree was built considering *deceased* as the dependent variable. The growth method selected was CART with a 10-fold cross-validation as the training method. The maximum tree depth was three levels, with 100 minimal subjects in parent nodes and 50 minimal subjects in child nodes. The Gini impurity measure was used as the splitting criterion, with 0.0001 as the minimum improvement required for a split. The prognostic value for mortality was recorded as a new variable to build a receiver operating characteristic curve. Tetrachoric tables with true/false positives and negatives were used to calculate the Odds Ratio (OR) of each node and the overall sensitivity, specificity, F1 score, positive and negative predictive values (PPV and NPV), positive and negative likelihood ratios (LR+ and LR−), area under the curve (AUC), and accuracy of the model.

All statistical data were analyzed using IBM Statistical Package for the Social Sciences (SPSS)® version 27.

Data availability statement

The data that support the findings of this study are openly available in Figshare at DOI: 10.6084/m9.figshare.28256057.[31] The dataset consists of a csv file. The data can be accessed without restrictions

and are provided under the Creative Commons Attribution 4.0 International License. Any additional information required to reanalyze the data reported in this paper is available from the corresponding author upon request.

Results

A total of 1432 data from confirmed patients of COVID-19 were evaluated, among whom 583 (41%) survived and 849 (59%) died. The sex ratio differed significantly between sexes, with 936 males (66%) and 496 females (34%), $p < 0.001$. The mean age ± standard deviation of patients who died was higher than that of survivors. SpO_2 on admission was significantly lower among patients who died compared to survivors ($p < 0.001$). Among the biochemical findings, elevated LDH levels were observed in patients who died compared with those who survived ($p \leq 0.001$). The NLR was significantly higher in patients who died than in survivors ($p \leq 0.001$). Regarding the number of days of hospital stay, there was no significant difference between the groups. Interestingly, BMI showed no difference between survivors and non-survivors ($p = 0.167$). Demographic data are presented in table 1.

Figure 1 shows the obtained classification tree, which identified patients with a high-risk profile for mortality based on the risk factors present upon admission.

Patients intubated on admission who spent < 17.5 days (Node 9) had the highest mortality (466/470; 99%), with an OR of 80 (confidence interval [CI] 95% = 12.84-32.82, $p < 0.001$). Contrastingly, non-intubated patients who stayed more than 5.5 days and had a GFR > 66.5 (Node 14) had the lowest mortality (5/95; 5%). Moreover, this profile served as a protective factor (OR = 0.26, CI 95% = 0.22-0.32, $p < 0.001$).

The overall performance achieved by the tree was F1 score = 0.918, ACC (accuracy) = 0.903, AUC = 0.955, CI 95% = 0.944-0.965, $p < 0.001$ (Fig. 2). The other performance metrics were PPV = 0.917, NPV = 0.883, LR+ = 7.686, and LR− = 0.091.

Discussion

CART models provide a means of identifying high-risk subgroups to whom prevention and intervention efforts can be targeted, making them useful tools for addressing research questions that cannot be answered by traditional regression methods.[32]

Our model effectively captures the complex, non-linear relationship between length of stay and survival, particularly when considering ventilation status and subsequent physiological parameters.

Similar studies

Previous studies have compared ML algorithms to predict mortality in COVID-19 patients. Zakariaee et al.[33] used Weka 3.9.2 to compare various algorithms for mortality prediction in an 815-subject database, finding that random forest performed best with 97.2% accuracy, 100% sensitivity, 94.8% precision, 94.5% specificity, 97.3% F-score, and 99.9% AUC. Other ML algorithms with an AUC from 81.2% to 93.9% also showed good performance.

Mohammadi-Pirouz et al.[34] developed three classification tree algorithms (Chi-squared Automatic Interaction Detector [CHAID], C5.0, and CART) for 5080 COVID-19-positive subjects using R 4.2.1 and SPSS 26. All models indicated that factors such as ICU hospitalization, intubation, age, kidney disease, BUN, CRP, WBC, NLR, SpO_2, and hemoglobin influenced mortality rates. CART performed best, achieving 0.75 sensitivity, 0.91 specificity, 0.36 precision, 0.90 accuracy, 0.41 F-score, and 0.92 AUC.

A Mexican study[35] evaluated four ML algorithms in 580,570 COVID-19 patients, namely logistic regression, Naive Bayes, decision tree, and random forest, obtaining accuracy values of 0.91, 0.89, 0.91, and 0.91, respectively. An AUC of 0.92 was reported without specifying the algorithm.

These studies undoubtedly provide valuable insights into the application of ML models for COVID-19 prognosis. However, many are limited by small sample sizes and lack comprehensively report of all performance metrics.

Huyut and Üstündağ studied 686 patients in Turkey focused on COVID-19 diagnosis and prognosis using blood-gas data and the CHAID tree algorithm.[36] They achieved a total accuracy of 65.0% in predicting the prognosis of the disease and 68.2% in diagnosing the disease. Other performance metrics were not reported. In this work, the model was developed in SPSS version 25.

ORs

Previous studies using CART trees have used ORs for single features.[37] However, no similar study

Table 1. Demographics all figures represent means ± SD compared by Mann-Whitney's U except where indicated

Variable	Overall (n = 1347)	Survivors (n = 583)	Non-survivors (n = 764)	p
*Sex Female, n (%)	458 (34)	211 (36)	247 (32)	< 0.001
AGE	55.31 ± 14.51	50.72 ± 13.54	58.82 ± 14.25	< 0.001
**Days of hospital stay	11 (6-17)	11 (7-15)	11 (5-18)	0.865
BMI	28.22 ± 5.42	28.26 ± 5.14	28.18 ± 5.62	0.436
*DM, n (%)	505 (40)	176 (39)	329 (61)	0.003
*SAH, n (%)	778 (60)	336 (43)	442 (57)	< 0.001
Glucose	173.41 ± 135.53	146.02 ± 128.99	194.40 ± 136.75	< 0.001
CR	94.01 ± 19.31	89.17 ± 16.3	96.69 ± 20.3	< 0.001
RR	26.07 ± 7.37	23.14 ± 4.87	27.7 ± 8.0	< 0.001
SpO_2	75.52 ± 15.82	81.94 ± 10.30	70.75 ± 17.44	< 0.001
Serum LDH IU/L	506.74 ± 390.26	415.80 ± 170.44	577.74 ± 487.04	< 0.001
GFR	86.63 ± 53.19	101.68 ± 47.41	75.54 ± 54.51	< 0.001
HB	14.03 ± 3.1	14.42 ± 3.04	13.74 ± 3.11	0.021
Leukocytes	11.02 ± 6.67	9.02 ± 5.55	12.51 ± 7.04	< 0.001
NLR	14.36 ± 16.34	8.84 ± 7.84	18.58 ± 19.57	< 0.001
*Invasive ventilatory support, n (%)	708 (52)	59 (8)	649 (92)	< 0.001

*Chi-squared.
**Medians (Interquartile range) compared by the median test.
BMI: body mass index; DM: diabetes mellitus; SAH: systemic arterial hypertension; CR: cardiac rate; RR: respiratory rate; SpO_2: oxygen saturation; LDH: lactate dehydrogenase; GFR: glomerular filtration rate; HB: hemoglobin; NLR: neutrophil-to lymphocyte ratio; SD: standard deviation.

(supervised ML for COVID-19 mortality) combines the calculation of the OR of each profile. This combination significantly enhances the model's utility by:

- Risk quantification, as each patient profile has a precise OR with CIs
- Terminal nodes with specific risk profiles: For non-ventilated patients hospitalized for 5.5 days or less, OR was 0.56, thus acting as a protective factor (Node 7), while for those with 80.5 or less SpO_2 (Node 11), the OR for mortality was 3.36 (77% probability of dying).

While the standard statistical comparison of days of hospital stay found no differences between survivors and non-survivors, the tree and OR calculation showed this variable's importance. For intubated subjects with hospital stays ≤ 25.5 days, the OR was 20.53 (CI 95% = 12.84-32.82, $p < 0.001$), whereas for those with longer stays, the OR reduced significantly (1.36, CI 95% = 0.92-2.00, $p < 0.001$). This represents a 58% difference in the probability of dying.

This OR-enhanced analysis strengthens the model's value as a clinical decision support tool, providing clinicians with precise risk estimates for different patient subgroups.

Variables

While other studies found laboratory values, comorbidities, or demographic factors as primary predictors, this model identified mechanical ventilation as the strongest discriminator, even when glucose, hemoglobin, LDH, neutrophils, and lymphocytes were introduced. Prior research has also recognized mechanical ventilation as a critical factor in predicting mortality among patients with COVID-19.[21,38,39] Our study quantifies the dramatic impact of mechanical ventilation, with an OR of 7.55 for ventilated patients compared to 0.26 for non-ventilated patients.

The model's reliance on easily available parameters (mechanical ventilation, days of stay, GFR, SpO_2)

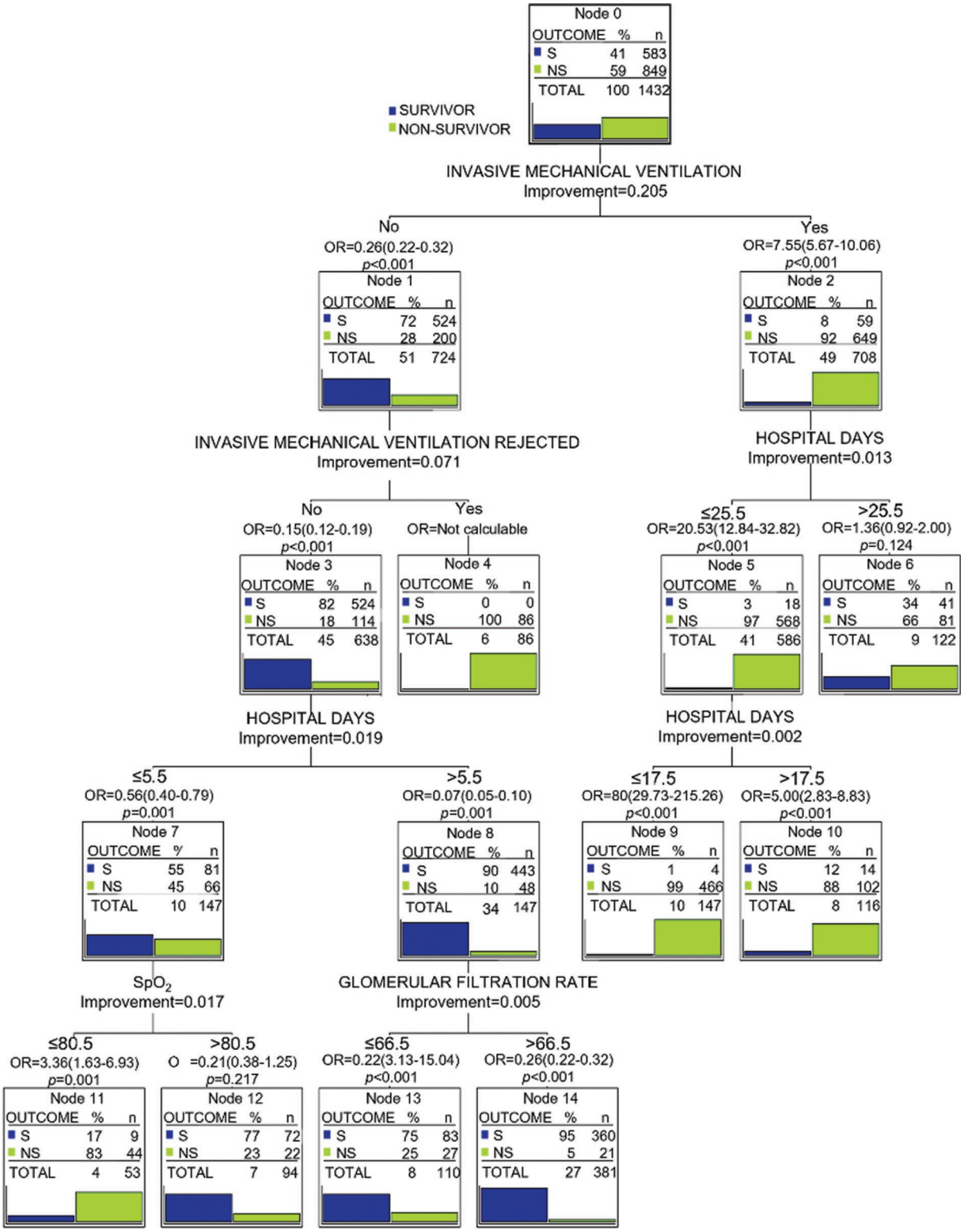

Figure 1. *Classification and regression tree obtained. Improvement is represented as the Gini index. OR is shown for each profile.*

represents a significant practical advantage, as all variables are routinely monitored during standard care without requiring highly specialized tests or extensive patient histories. This emphasis on easily obtainable clinical data, combined with the model's strong performance metrics, makes it valuable for real-world

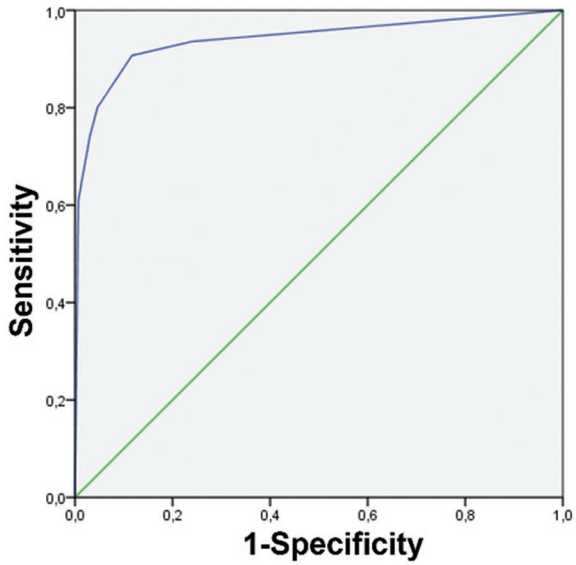

Figure 2. *Receiver operating characteristic curve obtained. Area under curve = 0.955, confidence interval 95% (0.944-0.965), p < 0.001. Sensitivity = 0.91. Specificity = 0.88.*

implementation across various healthcare settings, including resource-limited environments.

Voluntarily rejected mechanical ventilation

We observed that during the period of study, certain patients (or their responsible relatives on their behalf) rejected mechanical ventilation, mainly because of beliefs about COVID-19 and its treatments, shaped by misinformation or distrust in healthcare systems. Research has shown that health misinformation negatively affects individuals' decisions, leading to poor health outcomes and continued viral spread.[40] These beliefs may be influenced by fear and uncertainty in the context of the pandemic and the impact of social media, where even physicians posted misleading information.[41] In our study, a small number of patients rejected mechanical ventilation (86/1432; 6%); however, this decision had a striking impact on their outcome, given that none of them survived (Node 4, 0/86; OR not calculable). Recognizing this finding is crucial for evaluating the impact of misinformation and its outcomes within the context of pandemic outbreaks and could facilitate better-informed discussions on shared decision-making with patients and their families.

Metrics

The model exhibited strong discriminative performance, with balanced accuracy in both positive and negative predictions. It achieved a high sensitivity of 91.97% and specificity of 88.03%, reflecting its excellent capability to accurately identify both mortality and survival cases. Furthermore, the model's PPV of 91.76% and NPV of 88.34% demonstrate its reliability in predicting outcomes. Overall, the model's predictions are accurate approximately 90% of the time, regardless of the outcome.

The LR+ value of 7.686 indicates that a positive prediction is nearly 8 times more likely in true positive cases, whereas the very low LR− value of 0.0912 suggests that false negatives are rare. These values demonstrate strong diagnostic performance.

The F1 score (91.86%) and overall accuracy (90.36%) were fairly high, indicating that the model performed consistently well across all prediction aspects.

Although most published models limit their reporting to basic metrics such as accuracy, sensitivity, and specificity, our model provides a complete performance assessment through nine distinct metrics. Furthermore, the combination of node-specific ORs with these metrics (PPV, NPV, AUC) provides a thorough validation framework that surpasses the partial reporting common in published studies.

Compared with other clinical prediction models, these metrics indicate superior performance, suggesting that the CART model, which also accounts for intubation refusal, could be a highly reliable tool for COVID-19 mortality prediction, risk stratification, and resource allocation planning.

Limitations

This model needs to be tested using data from different centers with lower mortality rates. Potential temporal variations in COVID-19 treatment should be considered. Decisions on ventilation are not exempt from bias. Ultimately, the potential for selection bias must be acknowledged, as the study aimed to identify high-risk profiles during the initial phase of the pandemic; however, the selected participants may not accurately reflect the wider population due to the specific criteria set for inclusion. Finally, there exist confounding variables that may not have been considered in this investigation, including socioeconomic factors or discrepancies in healthcare practices.

Conclusions

The early identification of patients at high risk of mortality poses a significant challenge in medical

practice, especially in the context of SARS-CoV-2 infection, where timely intervention is critical.

By combining the CART model with the OR calculation of the profiles, our approach offers simple performance and interpretation, quickness, accessibility, and reliability, rendering it particularly suitable for implementation in resource-constrained healthcare settings, where the usage of perhaps more precise but indeed more complex tools for AI models (e.g. Python or R) represent a substantial complication, as they require much more specialized knowledge for installation, programming, and interpretation. By promptly identifying high-risk individuals, healthcare professionals can prioritize interventions and allocate resources more effectively, ultimately improving patient outcomes in the face of SARS-CoV-2 infection. Furthermore, insight into patients who voluntarily rejected mechanical ventilation provides real-world evidence of the impact of misinformation and supports clinical teams in counseling patients and families about prognosis.

Acknowledgments

Raúl Serrano-Loyola for the resources provided and Lucía Monserrat Pérez-Navarro, Diana Sofia Cabrera-Orejuela, Daniel E. Colli-Trejo, Melanie A. Mafara-Castillejos, Karla R. Rodríguez-Francisco, Carlos E. Maldonado-Barrientos, and Abraham Meneses-Juarez for data handling.

Funding

The authors declare that they have not received funding.

Conflicts of interest

The authors declare no conflicts of interest.

Ethical considerations

Protection of humans and animals. The authors declare that no experiments involving humans or animals were conducted for this research.

Confidentiality, informed consent, and ethical approval. The authors have obtained approval from the Ethics Committee for the analysis of routinely obtained and anonymized clinical data, so informed consent was not necessary. Relevant guidelines were followed.

Declaration on the use of artificial intelligence. The authors declare that artificial intelligence was used in the writing of this manuscript. Paperpal was used throughout the manuscript to correct and improve English writing.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382:727-33.
2. Meng L, Qiu H, Wan L, Ai Y, Xue Z, Guo Q, et al. Intubation and ventilation amid the COVID-19 outbreak: wuhan's experience. Anesthesiology. 2020;132:1317-32.
3. Lagina M, Valley TS. Diagnosis and management of acute respiratory failure. Crit Care Clin. 2024;40:235-53.
4. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020;8:475-81.
5. 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;323:1061-9.
6. Zhang J, Dong X, Liu G, Gao Y. Risk and protective factors for COVID-19 morbidity, severity, and mortality. Clin Rev Allergy Immunol. 2022;64:90-107.
7. Najera H, Ortega-Avila AG. Health and institutional risk factors of COVID-19 mortality in Mexico, 2020. Am J Prev Med. 2021;60:471-7.
8. Casillas N, Ramón A, Torres AM, Blasco P, Mateo J. Predictive model for mortality in severe COVID-19 patients across the six pandemic waves. Viruses. 2023;15:2184.
9. Okunogbe A, Nugent R, Spencer G, Powis J, Ralston J, Wilding J. Economic impacts of overweight and obesity: current and future estimates for 161 countries. BMJ Glob Health. 2022;7:e009773.
10. Murray CJ, Aravkin AY, Zheng P, Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet. 2020;396:1223-49.
11. Földi M, Farkas N, Kiss S, Zádori N, Váncsa S, Szakó L, et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: a systematic review and meta-analysis. Obes Rev. 2020;21:e13095.
12. Lighter J, Phillips M, Hochman S, Sterling S, Johnson D, Francois F, et al. Obesity in patients younger than 60 years is a risk factor for COVID-19 hospital admission. Clin Infect Dis. 2020;71:896-7.
13. Gong MN, Bajwa EK, Thompson BT, Christiani DC. Body mass index is associated with the development of acute respiratory distress syndrome. Thorax. 2010;65:44-50.
14. Abd-Elsayed, Alaa, Mena Abdelmalak, Kevin W. Chung, 'Mechanical Ventilation', in Alaa Abd-Elsayed (ed.), Miscellaneous, in Alaa Abd-Elsayed (ed.), Advanced Anesthesia Review (New York, 2023; online edn, Oxford Academic, 1 Mar. 2023), https://doi.org/10.1093/med/9780197584521.003.0359
15. Pierson DJ. Complications associated with mechanical ventilation. Crit Care Clin. 1990;6:711-24.
16. Weiss P, Murdoch DR. Clinical course and mortality risk of severe COVID-19. Lancet. 2020;395:1014-5.
17. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395:1054-62.
18. Du RH, Liu LM, Yin W, Wang W, Guan LL, Yuan ML, et al. Hospitalization and critical care of 109 decedents with COVID-19 pneumonia in Wuhan, China. Ann Am Thorac Soc. 2020;17:839-46.
19. Ñamendys-Silva SA. Respiratory support for patients with COVID-19 infection. Lancet Respir Med. 2020;8:e18.
20. Chalkias A, Huang Y, Ismail A, Pantazopoulos I, Papagiannakis N, Bitterman B, et al. Intubation decision based on illness severity and mortality in COVID-19: an international study. Crit Care Med. 2024;52:930-41.
21. Sousa Neto AL, Mendes-Rodrigues C, Pedroso RD, Röder DV. Revisiting the COVID-19 pandemic: mortality and predictors of death in adult patients in the intensive care unit. Life (Basel). 2024;14:1027.
22. Singhal A, Shukla G. Decision Trees: Classification and Regression Trees (CART). Bhubaneswar: National Institute of Science Education and Research (NISER); 2023. p. 16.
23. Karthikeyan A, Garg A, Vinod PK, Priyakumar UD. Machine learning based clinical decision support system for early COVID-19 mortality prediction. Front Public Health. 2021;9:626697.
24. Dauda K. Development and modeling of decision tree for survival data with multiple events using deviance and cox-snell residuals within node homogeneity technique. Int J Adv Heal Sci Technol. 2022;2:183-90.

25. Costanza MC, Paccaud F. Binary classification of dyslipidemia from the waist-to-hip ratio and body mass index: a comparison of linear, logistic, and CART models. BMC Med Res Methodol. 2004;4:7.
26. Umehara T, Katayama N, Kaneguchi A, Iwamoto Y, Tsunematsu M, Kakehashi M. Models to predict prognosis in older patients with heart failure complicated by pre-frailty and frailty: a pilot prospective cohort study. Hong Kong Med J. 2022;28:356-66.
27. Orfanoudaki A, Chesley E, Cadisch C, Stein B, Nouh A, Alberts MJ, et al. Machine learning provides evidence that stroke risk is not linear: the non-linear Framingham stroke risk score. PLoS One. 2020;15:e0232414.
28. Shim EJ, Yoon MA, Yoo HJ, Chee CG, Lee MH, Lee SH, et al. An MRI-based decision tree to distinguish lipomas and lipoma variants from well-differentiated liposarcoma of the extremity and superficial trunk: classification and regression tree (CART) analysis. Eur J Radiol. 2020;127:109012.
29. Hyunjoong K, Loh W. Classification trees with unbiased multiway splits. J Am Stat Assoc. 2001;96:589-604.
30. Twala BE, Jones MC, Hand DJ. Good methods for coping with missing data in decision trees. Pattern Recognit Lett. 2008;29:950-6.
31. Pérez-Pacheco A, Herrera-Suárez EA, Vélez Mata J, Avendaño-Carrera AD, Casillas-Suárez C, Pérez García A. Dataset of 1432 COVID-19 patients from Hospital General de México "Dr. Eduardo Liceaga" https://figshare.com/articles/dataset/Dataset_of_1432_COVID-19_patients_from_Hospital_General_de_M_xico_Dr_Eduardo_Liceaga_/28256057?-file=51881855
32. Esteban C, Arostegui I, Moraza J, Aburto M, Quintana JM, Pérez-Izquierdo J, et al. Development of a decision tree to assess the severity and prognosis of stable COPD. Eur Respir J. 2011;38:1294-300.
33. Zakariaee SS, Naderi N, Ebrahimi M, Kazemi-Arpanahi H. Comparing machine learning algorithms to predict COVID19 mortality using a dataset including chest computed tomography severity score data. Sci Rep. 2023;13:11343.
34. Mohammadi-Pirouz Z, Hajian-Tilaki K, Sadeghi Haddat-Zavareh M, Amoozadeh A, Bahrami S. Development of decision tree classification algorithms in predicting mortality of COVID-19 patients. Int J Emerg Med. 2024;17:126.
35. Sánchez-Gálvez AM, Sánchez-Gálvez S, Álvarez-González R, Rojas-Alarcon F. Covid-19 mortality risk prediction model using machine learning. Comput Sist. 2023;27:881-8.
36. Huyut MT, Üstünda H. Prediction of diagnosis and prognosis of COVID-19 disease by blood gas parameters using decision trees machine learning model: a retrospective observational study. Med Gas Res. 2022;12:60-6.
37. Zimmerman RK, Balasubramani GK, Nowalk MP, Eng H, Urbanski L, Jackson ML, et al. Classification and regression Tree (CART) analysis to predict influenza in primary care patients. BMC Infect Dis. 2016;16:503.
38. Soares TH, De Moraes NH, Soares KP, Carvalho MM, Holanda AS, Rodrigues LF, et al. Factors associated with mortality of patients with COVID-19 on invasive mechanical ventilation: a retrospective cohort study in a university hospital in Northeastern Brazil. An Acad Bras Cienc. 2024;96:e20231355.
39. Li H, Ashrafi N, Kang C, Zhao G, Chen Y, Pishgar M. A machine learning-based prediction of hospital mortality in mechanically ventilated ICU patients. PLoS One. 2024;19:e0309383.
40. Caceres MM, Sosa JP, Lawrence JA, Sestacovschi C, Tidd-Johnson A, Ul Rasool MH, et al. The impact of misinformation on the COVID-19 pandemic. AIMS Public Health. 2022;9:262-77.
41. Sule S, DaCosta MC, DeCou E, Gilson C, Wallace K, Goff SL. Communication of COVID-19 misinformation on social media by physicians in the US. JAMA Netw Open. 2023;6:e2328928.