BMC Infectious Diseases

RESEARCH

Open Access

Unraveling relevant cross-waves pattern drifts in patient-hospital risk factors among hospitalized COVID-19 patients using explainable machine learning methods

Fernanda Cristina Barbosa Lana[1*], Carolina Coimbra Marinho[2], Bruno Barbosa Miranda de Paiva[3], Lucas Rocha Valle[1], Guilherme Fonseca do Nascimento[4], Leonardo Chaves Dutra da Rocha[4], Marcelo Carneiro[5], Joanna d'Arc Lyra Batista[6], Fernando Anschau[7], Pedro Gibson Paraiso[8], Frederico Bartolazzi[9], Christiane Corrêa Rodrigues Cimini[10], Alexandre Vargas Schwarzbold[11], Danyelle Romana Alves Rios[12], Marcos André Gonçalves[3] and Milena Soriano Marcolino[2,13]

Abstract

Background Several studies explored factors related to adverse clinical outcomes among COVID-19 patients but lacked analysis of the impact of the temporal data shifts on the strength of association between different predictors and adverse outcomes. This study aims to evaluate factors related to patients and hospitals in the prediction of in-hospital mortality, need for invasive mechanical ventilation (IMV), and intensive care unit (ICU) transfer throughout the pandemic waves.

Methods This multicenter retrospective cohort included COVID-19 patients from 39 hospitals, from March/2020 to August/2022. The pandemic was divided into waves: 10/03/2020-14/11/2020 (first), 15/11/2020-25/12/2021 (second), 26/12/2021-03/08/2022 (third). Patient-related factors included clinical, demographic, and laboratory data, while hospital-related factors covered funding sources, accreditation, academic status, and socioeconomic characteristics. Shapley additive explanation (SHAP) values derived from the predictions of a light gradient-boosting machine (LightGBM) model were used to assess potential risk factors for death, IMV and ICU.

Results Overall, 16,958 adult patients were included (median age 59 years, 54.7% men). LightGBM achieved competitive effectiveness metrics across all periods. Temporal drifts were observed due to a decrease in various metrics, such as the recall for the positive class [ICU: 0.4211 (wave 1) to 0.1951 (wave 3); IMV: 0.2089 (wave 1) to 0.0438 (wave 3); death: 0.2711 (wave 1) to 0.1175 (wave 3)]. Peripheral arterial oxygen saturation to the fraction of inspired oxygen ratio ($SatO_2/FiO_2$) at admission had great predictive capacity for all outcomes, with an optimal cut-off value for death prediction of 227.78. Lymphopenia had its association strength increased over time for all outcomes, optimal

*Correspondence:
Fernanda Cristina Barbosa Lana
fernandacblana@gmail.com

Full list of author information is available at the end of the article

threshold for death prediction of 643×10^9/L. Thrombocytopenia was the most important feature in wave 2 (ICU); overall, values below $143{,}000 \times 10^9$/L were more related to death.

Conclusion Data drifts were observed in all scenarios, affecting potential predictive capabilities of explainable machine learning methods. Upon admission, $SatO_2/FiO_2$ values, platelet and lymphocyte count were significant predictors of adverse outcomes in COVID-19 patients. Overall, inflammatory response markers were more important than clinical characteristics. Limitations included sample representativeness and confounding factors. Integrating the drift's knowledge into models to improve effectiveness is a challenge, requiring continuous updates and monitoring of performance in real-world applications.

Clinical trial number Not applicable.

Keywords COVID-19, Risk factors, Socioeconomic factors, Machine learning, Critical care outcomes, Mortality

Background

The COVID-19 pandemic has presented a significant challenge to healthcare systems worldwide [1, 2]. Throughout the pandemic years, different countries have experienced distinct waves of infections and deaths [3]. Since the first case, more than 770 million confirmed cases of the disease and over seven million deaths have been reported, although the actual number is thought to be higher [4]. Various factors, including the emergence of new SARS-CoV-2 variants, fluctuations in public health measures and changes in population behavior, have played a crucial role in the rates of infectivity, mortality and morbidity, thus, contributing to successive waves of deaths [3, 5, 6].

Although previous studies have investigated risk factors that might be associated with a higher risk of adverse events in hospitalized COVID-19 patients, most of them lack analysis of the impact of different categories of features on adverse outcomes using a robust dataset. Also, they lack analysis of the impact of these features on model output and how this impact evolved over time [3, 7–9].

Additionally, most of the studies use logistic regression to analyze and understand potential risk factors in infectious diseases. Although a popular method, widely used due to its simplicity and interpretability, it has limitations that can affect its applicability and effectiveness [10]. In the context of potential non-linear relationships and complex interactions among variables, machine learning methods differentiate themselves since they allow the identification of interactions that a traditional statistical analysis might fail to fully capture [11]. Studies that employed machine learning methods in COVID-19 patients were mostly conducted in developed countries, primarily with cohort sample sizes of less than 5,000 patients [12].

This study aimed to assess the temporal drifts of clinical, laboratory and socioeconomic characteristics at the time of hospital admission, on in-hospital mortality, need for invasive mechanical ventilation (IMV), and intensive care unit (ICU) transfer in a Brazilian cohort of COVID-19 patients, using advanced non-linear machine learning methods.

Methods

Study design and participants

This study is a substudy of the Brazilian COVID-19 Registry, a multicenter retrospective cohort of consecutive adult (≥18 years-old) COVID-19 patients from 39 different Brazilian hospitals, across 5 states, admitted from March 2020 to August 2022 [13]. The hospitals were not randomly chosen, they were invited to join the cohort. Confirmatory COVID-19 diagnosis followed the World Health Organization guidance [14]. The exclusion criteria were patients who were transferred to a non-participant institution, patients in palliative care, pregnant women, and patients who developed COVID-19 symptoms while admitted for other conditions.

The study was approved by the Brazilian National Research Ethics Committee (*Comissão Nacional de Ética em Pesquisa*– CONEP), Certificate of Presentation for Ethical Assessment 30350820.5.1001.0008, as well as internal approval of ethics boards from each hospital. Individual informed consent was waived by *Comissão Nacional de Ética em Pesquisa* - National Research Ethics Committee due to the pandemic situation and the use of deidentified data, based on medical chart review only. The study adhered to the Declaration of Helsinki.

This reporting followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for observational cohort studies [15] (table S1).

Patient-level data

Demographic information, clinical characteristics, laboratory and outcome data were collected from hospital medical records. The data collection instrument was designed with reference to COVID-19 guidelines from the World Health Organization and the Brazilian Ministry of Health, as previously described [13].

Trained hospital staff or interns collected the patients' information using the Research Electronic Data Capture (REDCap®) online tool [16, 17], hosted at the Telehealth

Center, University Hospital, *Universidade Federal de Minas Gera*is. Variables were described in detail previously [18].

The team involved in data collection underwent online training to ensure process quality. They were also provided with a coding manual, developed for this research, guiding data collection for each variable [13]. In case of any doubts about the accuracy and reliability of the data, the investigators contacted the centers' representatives and asked them to check the information.

Socioeconomic characteristics were analyzed using the gross domestic product (GDP) per capita and average human development index (HDI) of each patients' municipality of origin. Values were obtained from the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística-IBGE) [19].

Hospital-level data

Data regarding hospitals' characteristics included source of income (public, mixed - partly public and partly private - or private), hospitals' accreditation, academic status (accredited medical residency program) and socioeconomic characteristics of each municipality where the hospital was located.

Source of income and academic status were obtained directly through each hospital website and through the open national database from National Registry of Health Facilities (*Cadastro Nacional de Estabelecimentos da Saúde*– CNES) [20]. Hospitals with mixed sources of income were allocated in the private group.

Hospital's accreditation was assessed through hospital's website and double checked by consulting the National Accreditation Organization (*Organização Nacional de Acreditação*– ONA) and the National Health Agency (*Agência Nacional de Saúde* -ANS) [21, 22].

In order to assess the socioeconomic characteristics of the population that attended these hospitals, we analyzed the GDP per capita, and HDI of each hospital municipality, also using data obtained from the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística-IBGE) [19].

Outcomes

The primary outcomes were in-hospital mortality, transfer to ICU and IMV support. For the analysis of the outcome transfer to ICU, patients who were admitted directly to the ICU and those who required mechanical ventilation at the time of hospital admission were excluded. Regarding the outcome IMV, patients who required mechanical ventilation at the time of hospital admission were also excluded.

Data analysis

Statistical analyses were conducted using IBM SPSS software (version 21.0) and R software (version 4.0.0). Descriptive analyses were used to summarize the variables. All the continuous variables exhibited a non-normal distribution. Therefore, they were summarized using medians and interquartile ranges (IQR). Categorical variables were represented through counts and percentages.

The study period was divided in three waves based on literature data of the number of deaths during each epidemiological week in Brazil. The first pandemic wave comprised the period from March 10th, 2020, to November 14th, 2020; the second wave was from November 15th, 2020, to December 25th, 2021, and the third wave was from December 26th, 2021, to March 8th, 2022 [5].

Variables potentially associated with the outcomes were selected based on previous studies as described in a supplementary material (Table S2). Those with more than 50% of missing data were excluded from the analyses.

Potential risk factors were evaluated for each outcome during each pandemic wave using Shapley Additive Explanation (SHAP) values obtained from a Light Gradient-Boosting Machine (LightGBM) model. LightGBM is a non-linear method, based on decision trees, that excels in handling structured data, supporting high-speed computation, and effectively modeling complex interactions and non-linear relationships [11]. SHAP was used in order to interpret the output obtained from the LightGBM model. It is a widely used method for model explainability. Since it assigns each feature an importance value for a particular prediction, it is possible to understand the contribution of individual features to the model's non-linear predictions [23].

This machine learning method is not affected by collinearity, and no imputation is required for the missing data. Boosting models, such as LightGBM, inherently handle missing values without explicit imputation. Rather than simply ignoring missing values, these models can learn patterns associated with missingness, leveraging the fact that data may be missing for informative reasons. Consequently, missing cases were retained within the algorithm [24].

However, to further investigate the impact of missingness, a sensitivity analysis was conducted, using two imputation strategies: (1) mean imputation and (2) missForest imputation.

Additionally, to demonstrate the effectiveness of LightGBM, key evaluation measures were calculated to assess its classification performance for each outcome: micro F1, macro F1, recall (positive class), and precision (positive class). Micro F1 and macro F1 are two ways to aggregate F1 scores across multiple classes: micro F1 calculates the F1 score globally by considering the total true positives, false negatives, and false positives across all

classes, while macro F1 computes the F1 score for each class individually and then takes their unweighted average, giving equal weight to each class regardless of its size. Recall and precision are metrics also used to evaluate the classifier effectiveness in a per-class basis: recall (also known as sensitivity or true positive rate) is the proportion of actual positive instances correctly identified as positive, calculated as Recall = True Positives/(True Positives + False Negatives). Precision measures the proportion of predicted positive instances that are correct, calculated as Precision = True Positives/(True Positives + False Positives) [25]. In this study, these metrics were specifically analyzed for the positive class.

To assess the risk of overfitting, given the large feature set, the behavior of the predictive models during training and validation was analyzed. During the model training phase, the dataset was divided so that 80% of the data was allocated for training and 20% for validation. Loss interactions for both training and validation sets were monitored, considering the three outcomes for all three waves. In a machine learning context, "loss" is a numerical measure of how far the model's predictions deviate from the true values. Graphs were generated to analyze these interactions.

Leveraging SHAP explainability strategies, we obtained the feature importance and the impact of variables on predictions over the waves. We trained gradient boosting algorithms in a 10-fold cross-validation procedure, saving SHAP-values for each sample in the test set of each fold (so as to have SHAP-values for all individual instances). The 10-fold cross-validation experimental procedure was obtained separately for each wave. A diagram depicting this procedure is detailed in Figure S1.

To compute these values, SHAP considers all possible combinations of features and evaluates the change in the model's prediction when each feature is added. This process is akin to distributing the "payout" (model prediction) among the "players" (features) based on their marginal contributions across all possible coalitions (feature combinations). By averaging these contributions, SHAP ensures that the feature attributions are fairly distributed, capturing both interaction effects and the importance of individual features. We can assess how each value of each variable influenced the outcome achieved by the predictive model [23].

In the SHAP graphic, the Y-axis represents the variables of our model in order of importance. The X-axis represents the SHAP values. Positive values indicate support for the reference category (contributes to the outcome), while negative values indicate support for the opposite category. Each point on the graph represents a sample. Finally, the colors represent the increase/decrease in the value of the variable. Shades of red indicate high values and shades of blue indicate lower values [23].

To enhance the clinical applicability of the findings, an experiment was conducted to establish cutoff points for three variables identified as important to the model in predicting mortality. Using a systematic greedy optimization approach, all values observed for the referred variables were evaluated on a linear scale to identify the threshold that offered the best balance between sensitivity and specificity across mortality classes. Additionally, plots illustrating the macroF1 score results across these variables range were generated, which provided insights into the relationship between the cut-off value and prediction performance.

Results

Overall, 16,957 patients were included (Fig. 1). Of those, 6,246 (36.8%) were from the first wave, 9,385 (55.4%) from the second and 1,326 (7.8%) from the third one. The overall median age was 59 (IQR 47–71) years old and 54.7% of the participants were men. The most common comorbidities were hypertension (53.2%), diabetes mellitus (27.0%) and obesity (18.8%). Table S3 shows the non-missing cases for each variable.

Table 1 and Table S4 summarize clinical and demographic characteristics, as well as patient comorbidities in each wave.

There was a low proportion of patients who received specific treatment against SARS-CoV-2: tocilizumab (163 patients, 1.0%); convalescent plasma (86 patients, 0.5%); immunoglobulin (16 patients, (0.1%); sarilumab (6 patients, 0.0%); remdesivir (5 patients, 0.0%). The proportion of other medications used for treatment is available in Table S5. Anticoagulants and systemic glucocorticoids were used by 15,398 (90.8%) and 14,110 patients (83.2%), respectively.

Table S6 shows hospitals' characteristics. The majority of included hospitals (32–82.1%) housed accredited medical residency programs, and 19 (48.7%) received funding exclusively from public sources. A minority of the hospitals possessed quality accreditation ($n = 12$; 30.2%).

Table 2 summarizes the outcomes throughout the pandemic waves.

LightGBM achieved competitive effectiveness metrics across the three outcomes and time periods (waves 1, 2 and 3). Key results for micro F1, macro F1, recall (positive class), and precision (positive class) are summarized in Table S7. The metrics demonstrate a clear temporal drift in the dataset across the three COVID-19 waves. For instance, the consistently decreasing recall for ICU admissions (from 0.4211 in wave 1 down to 0.1951 in wave 3) and IMV (from 0.2669 in wave 2 to 0.0438 in wave 3) points to changes in how patients were classified

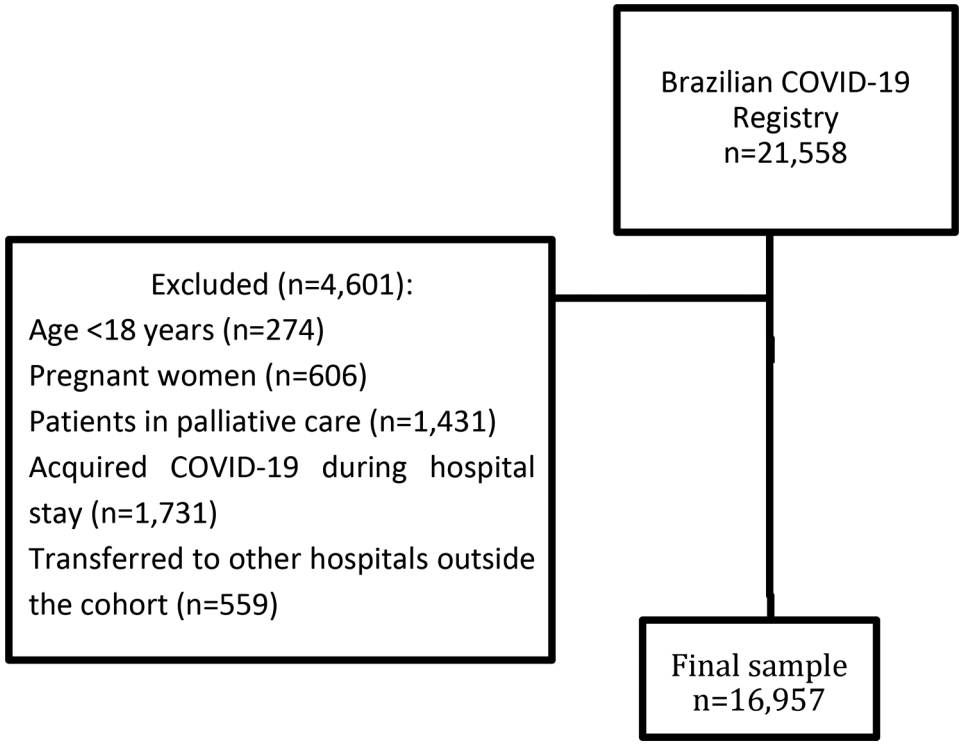

Fig. 1 Flowchart of the patients included in this study

or how the disease manifested over time, which in turn impacted model effectiveness. Additionally, although the micro F1 scores for IMV and death sometimes increased between waves, the substantially lower recalls in Wave 3 (e.g., 0.1175 for death and 0.0438 for IMV) suggest a skew toward correctly identifying negative cases while missing many positives. These shifts in model underscore the presence of temporal drift in this COVID-19 dataset.

ROC curves for each outcome in each wave are available in the supplementary material (Figure S2). In sensitivity analysis, SHAP values obtained from models trained on imputed data were highly similar to those obtained from our original approach without imputation (Figures S3-S8 -).

The risk of overfitting was monitored and represented in Figures S9, S10, S11. The figure demonstrates that loss decreases consistently for both sets at each iteration, with no signs of divergence in any of the waves analyzed or outcomes evaluated. This behavior is in line with what is expected for models with good generalization capacity on unseen data, which indicates that the predictive models used do not show signs of overfitting, even considering a large number of variables.

Impact of variables on the outcome death

Across all waves, $SatO_2/FiO_2$ consistently demonstrated the highest predictive capacity, with higher values being strongly associated with lower mortality rates. Among the most significant predictors, only two represented clinical characteristics: age and heart rate. Age was the second most important variable in waves 1 and 2 and the fourth in wave 3 (Fig. 2). Heart rate emerged as a relevant predictor in wave 2, with higher values correlating with an increased risk of death.

Regarding inflammatory markers, low platelet counts were predictive of mortality in waves 1 and 2. However, a shift in the pattern was observed in wave 3, where extreme platelet values (both high and low) were associated with increased mortality. Lymphopenia consistently ranked among the top predictors of death across all waves. Its importance grew in wave 3, where it became the second most relevant variable.

Arterial pH, while the fifth most important variable in wave 1, showed a decline in predictive capacity over subsequent waves.

Hospital accreditation appeared as a significant predictor in wave 1, suggesting that accredited hospitals experienced lower mortality rates. However, no hospital-related variables were identified among the top predictors in waves 2 and 3.

$SatO_2/FiO_2$ was the first death predictor across all waves, with lower values being predictive of lethality. Higher ages were the second-best predictor for waves 1 and 2 but dropped to fourth place in the ranking during wave 3. Hospital accreditation was a relevant predictor

Table 1 Demographic, clinical and hometown characteristics of a cohort of Brazilian patients admitted to the hospital with COVID-19 ($n = 16,957$)

Variables[§]	All waves Frequency (%) or median (IQR)	Wave 1** Frequency (%) or median (IQR)	Wave 2** Frequency (%) or median (IQR)	Wave 3** Frequency (%) or median (IQR)
Age (years)	59 (47–71)	60 (47–71)	58 (47–69)	69 (57–80)
Men	9,271 (54.7%)	3,471 (55.6%)	5,123 (54.6%)	677 (51.1%)
Comorbidities				
Hypertension	9,026 (53.2%)	3,383 (54.2%)	4,889 (52.1%)	754 (56.9%)
Coronary artery disease	792 (4.7%)	329 (5.3%)	354 (3.8%)	109 (8.2%)
Heart failure	896 (5.3%)	372 (6.0%)	367 (3.9%)	157 (11.8%)
COPD	917 (5.4%)	376 (6.0%)	361 (3.8%)	180 (13.6%)
Diabetes mellitus	4,586 (27%)	1,805 (28.9%)	2,370 (25.3%)	411 (31.0%)
Obesity	3,191 (18.8%)	1,176 (18.8%)	1,872 (19.9%)	143 (10.8%)
Cirrhosis	63 (0.4%)	32 (0.5%)	19 (0.2%)	12 (0.9%)
Chronic kidney disease	748 (4.4%)	311 (5.0%)	319 (3.4%)	118 (8.9%)
Cancer	629 (3.7%)	280 (4.5%)	227 (2.4%)	122 (9.2%)
Clinical characteristics at hospital presentation				
Vaccine	1,539 (16.1%)	0	919 (9.8%)	620 (46.8%)
GCS < 15	1,340 (8.6%)	734 (11.8%)	430 (4.6%)	176 (13.3%)
Heart rate	86 (76–97)	88 (78–100)	85 (76–95)	82 (72–92)
Respiratory rate	20 (18–24)	20 (18–24)	21 (19–25)	20 (18–22)
Temperature (ºC)	36.5 (36.0–37.0)	36.5 (36.0-37.2)	36.4 (36.0-36.8)	36.3 (36.0-36.7)
IMV at hospital presentation	1,056 (6.2%)	413 (6.6%)	587 (6.3%)	56 (4.2%)
$SatO_2/FiO_2$	400.0 (296.9-442.9)	428.6 328.6-452.4)	346.4 (272.2-433.3)	428.6 (339.3-457.1)
Systolic Blood Pressure (mmHg)				
≥90 mmHg without amine	14,317 (96.1%)	5,573 (93.6%)	7,647 (97.9%)	1,097 (97.4%)
<90mmHg without amine	182 (1.2%)	78 (1.3%)	79 (1.0%)	25 (2.2%)
Any value, but with amine	394 (2.6%)	304 (5.1%)	86 (1.1%)	4 (0.3%)
Diastolic blood pressure (mmHg)				
>60mmHg without amine	12,668 (85.5%)	4,869 (82.4%)	6,851 (88.1%)	948 (87.7%)
≤60mmHg without amine	1,748 (11.8%)	736 (12.4%)	842 (10.8%)	170 (15.1%)
Any value, but with amine	394 (2.7%)	304 (5.1%)	86 (1.1%)	4 (0.3%)
Hometown characteristics				
GDP (BRL)	24,993.0 (24,923.0–30,302.7)	28,525.8 (24,922.9–38,670.4)	24,922.9 (24,922.9–24,922.9)	24,922.9 (24,922.9–24,922.9
HDI	0.8 (0.7–0.8)	0.8 (0.7–0.8)	0.8 (0.7–0.8)	0.8 (0.8–0.8)

COPD: chronic obstructive pulmonary disease; GCS: glasgow coma scale; GDP: gross domestic product; (BRL): Brazilian reais; HDI: human development index; IMV: invasive mechanical ventilation

* Final cohort sample

** Wave 1: 10/03/2020-14/11/2020; wave 2: 15/11/2020-25/12/2021; wave 3: 26/12/2021-03/08/2022

[§]Each variable missing cases are available in Table S3

only during the first wave of the pandemic, losing its position in the ranking as the waves progressed.

ALT: alanine aminotransferase; AST: aspartate aminotransferase; CRP: C-reactive protein; FiO2: fraction of inspired oxygen; HCO3: bicarbonate; pCO2: partial pressure of carbon dioxide; pO2: partial pressure of oxygen; pH: potential of hydrogen; SatO2: oxygen saturation; SHAP: Shapley Additive ExPlanation.

Impact of variables on the outcome ICU

$SatO_2/FiO_2$ was consistently inversely associated with ICU transfer and stood out as one of the most influential variables across all three waves. It ranked as the most important predictor in waves 1 and 3, while occupying the second position in wave 2 (Fig. 3).

Age demonstrated a distinct pattern when predicting ICU transfer compared to its role in mortality. Its impact increased progressively over the waves, revealing a complex relationship. In wave 1 and 2, older patients

Table 2 Frequency of the outcomes throughout the COVID-19 pandemic waves in a Brazilian cohort of hospitalized patients ($n = 16,957$)

Outcome	Wave 1* Frequency *n* (%)	Wave 2* Frequency *n* (%)	Wave 3* Frequency *n* (%)
Death	1,034 (16.6%)	1,906 (20.3%)	187 (14.2%)
ICU	2,332 (37.4%)	3,376 (36.0%)	328 (24.7%)
IMV	1,701 (27.2%)	2,576 (27.4%)	185 (14%)

IMV: invasive mechanical ventilation; ICU: intensive care unit

*Wave 1: 10/03/2020-14/11/2020; wave 2: 15/11/2020-25/12/2021; wave 3: 26/12/2021-03/08/2022

exhibited a higher likelihood of ICU transfer, while in wave 3, advanced age was no longer associated with greater risk of ICU admission. Lymphopenia consistently showed an inverse relationship with ICU transfer across all waves, rising to become the second most important variable in wave 3. Conversely, neutrophilia, which was directly associated with ICU transfer in wave 1, lost its predictive relevance in subsequent waves.

AST/ALT ratio consistently ranked among the top 9 predictors throughout the three waves, indicating a direct association with ICU transfer. In contrast, platelet count and respiratory rate were significant predictors only in waves 1 and 2, with their importance diminishing in wave 3.

Interestingly, source of income emerged as the third most impactful variable in wave 2, with private hospitals being a significant positive predictor of ICU transfer. However, no other hospital characteristics ranked among the top predictors of ICU transfer across the waves.

$SatO_2/FiO_2$ was the best predictor for ICU transfer during the first and third waves, though its predictive power was less prominent in the second wave. In this prediction task, inflammatory and respiratory parameters, along with age, remained consistent predictors throughout the pandemic.

ALT: alanine aminotransferase; AST: aspartate aminotransferase; CRP: C-reactive protein; FiO2: fraction of inspired oxygen; HCO3 - bicarbonate; pCO2: partial pressure of carbon dioxide; pO2: partial pressure of oxygen; pH: potential of hydrogen; SatO2: oxygen saturation; SHAP: Shapley Additive ExPlanation.

Impact of variables on the outcome IVM

Similar to findings for other outcomes, higher $SatO_2/FiO_2$ values at admission were consistently associated with a lower likelihood of requiring invasive mechanical ventilation (IMV) across all waves. The temporal trend for respiratory rate revealed an opposite pattern, as its predictive importance decreased consistently from wave 1 to wave 3.

Age at hospital admission was a key predictor of IMV during waves 1 and 2, with older patients showing a higher likelihood of requiring invasive ventilation. However, in wave 3, this trend shifted and advanced age was associated with reduced IMV support.

Lymphopenia remained a significant predictor of IMV support across all waves and emerged as the second most important variable in wave 3. Similarly, neutrophilia was strongly associated with IMV support in wave 1 but lost its predictive relevance in subsequent waves.

Among other laboratory markers, a low platelet count was a notable predictor of IMV support, particularly during wave 2, when it was the most important feature in the model. It also ranked third and fourth in importance in waves 1 and 3, respectively.

Being vaccinated appeared as the ninth most impactful predictor in wave 3. During this period, patients with missing vaccination data (represented in gray in Fig. 4) and those who had been vaccinated exhibited a lower likelihood of IMV support.

Higher $SatO_2/FiO_2$ values were associated with a lower likelihood of requiring IMV support. Lymphopenia was a significant predictor of the outcome across all waves and was the second most important variable in wave 3. Advanced age was a top predictor for IMV support in waves 1 and 2; however, this pattern reversed in the third wave.

ALT: alanine aminotransferase; AST: aspartate aminotransferase; CRP: C-reactive protein; FiO2: fraction of inspired oxygen; GDP: gross domestic product; HCO3 - bicarbonate; pCO2: partial pressure of carbon dioxide; pO2: partial pressure of oxygen; pH: potential of hydrogen; SatO2: oxygen saturation; SHAP: Shapley Additive ExPlanation.

Clinical applicability: features thresholds for death prediction

SHAP graphs revealed $SatO_2/FiO_2$, platelets and lymphocytes count as impactful features for death prediction during the pandemic waves. An additional analysis of the macro-F1 score for mortality prediction across these variables range pointed to cut-off values with best prediction performance. The results revealed that $SatO_2/FiO_2$ below 227.78, thrombocytopenia under $143{,}000 \times 10^9$/L and less than 643×10^9/L lymphocytes are the thresholds with best death prediction capacity for each variable. Plots illustrating the results are available in supplementary material (Figures S12, S13, S14).

Discussion

Main findings

This study leveraged advanced machine learning methods, specifically LightGBM combined with SHAP analysis, to identify key predictors of adverse outcomes in hospitalized COVID-19 patients across distinct waves of the pandemic. By examining mortality, ICU transfer, and

Fig. 2 Impact of variables on the prediction of death during the waves

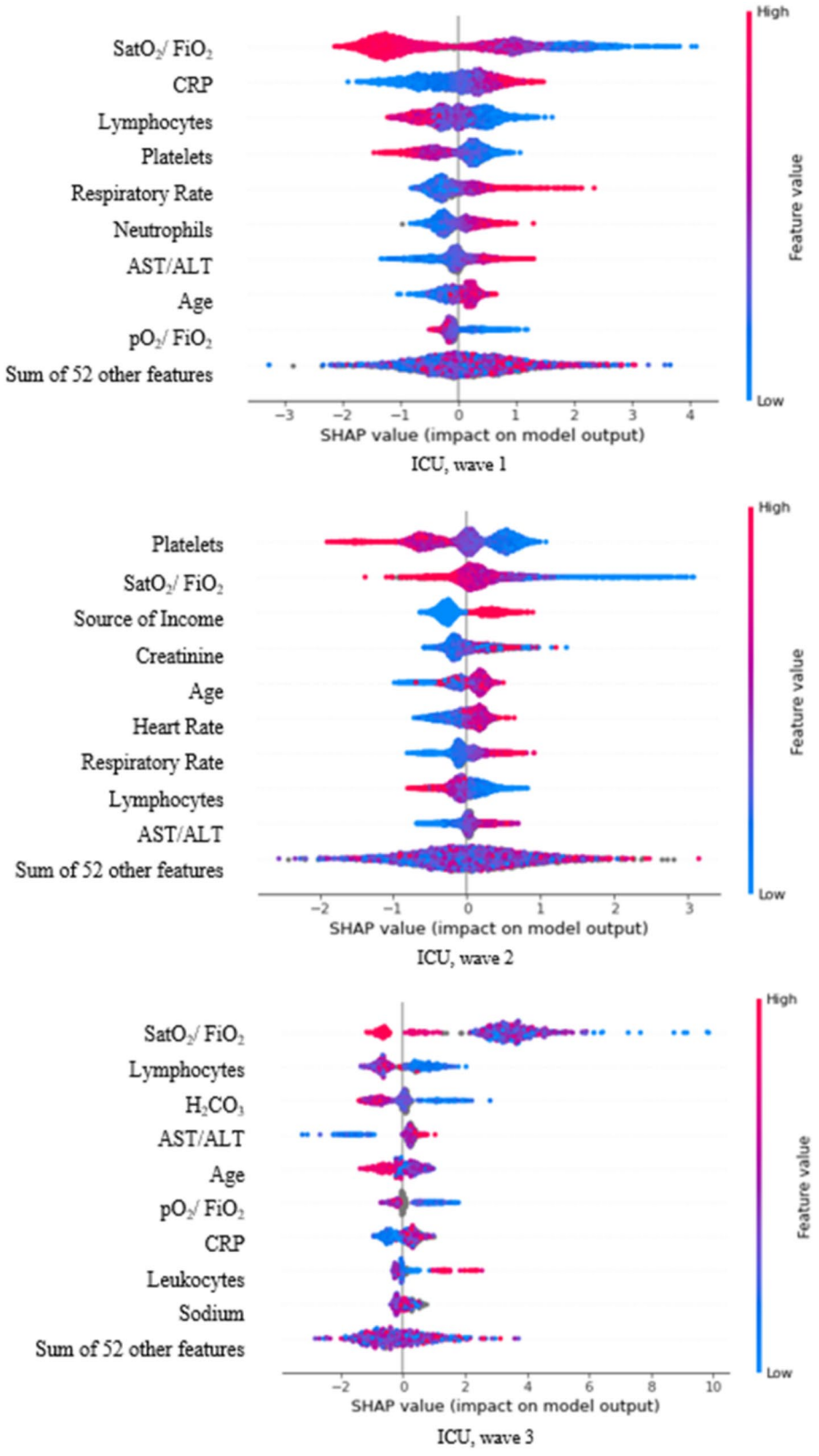

Fig. 3 Impact of variables on the prediction of ICU transfer during the waves

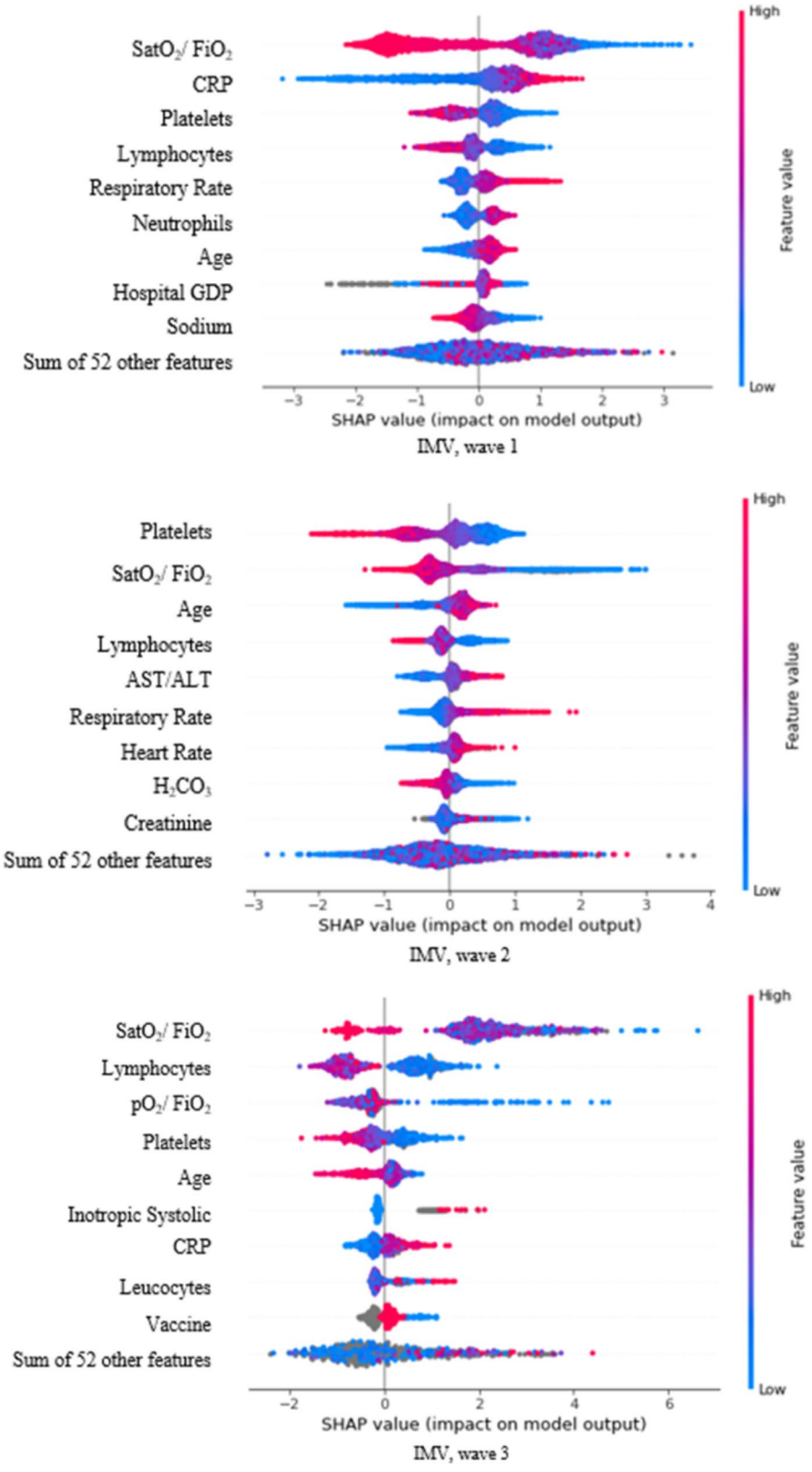

Fig. 4 Impact of variables on the prediction of IMV support during the waves

IMV support as outcomes, this study provided insights into the temporal evolution of key predictors and highlighted the utility of explainable machine learning methods in identifying critical features in patient outcomes. The study also found that $SatO_2/FiO_2$ and age had a great impact on all the outcomes analyzed. Inflammatory response markers, especially platelets and lymphocytes, were more important than most clinical characteristics to predict the outcomes. There were no consistent patterns regarding hospital characteristics, as they only appeared in a few SHAP waves.

Our approach demonstrated the capability of machine learning models to uncover complex and dynamic patterns in healthcare data emphasizing the need to address temporal drift in COVID-19 datasets and highlighting their potential to enhance predictive modeling in infectious diseases. LightGBM combined with SHAP analysis provided critical advantages over traditional statistical approaches, such as logistic regression analysis [11]. SHAP values not only rank variables by importance and indicate the direction of their association with the outcome but also reflect their relative contributions at a given point in time. This approach allowed us to capture temporal shifts in feature relevance, offering valuable insights for real-time clinical decision-making.

Laboratory markers

Our findings suggest that biochemical and physiological assessments play a key role in predicting patient prognosis. Among the 61 variables assessed, $SatO_2/FiO_2$ consistently emerged as the most influential predictor for all outcomes (mortality, ICU transfer, and IMV support) across the different pandemic waves. Higher $SatO_2/FiO_2$ values were strongly associated with reduced risks of adverse outcomes (mortality, ICU transfer, and IMV need), underscoring its utility as a critical marker of disease severity.

Its impact extends beyond its statistical significance, carrying important clinical implications. A decreased $SatO_2/FiO_2$ ratio is a well-established indicator of impaired gas exchange and hypoxemia, often associated with disease severity and the need for advanced respiratory support, including high-flow oxygen therapy or mechanical ventilation [26, 27]. This finding aligns with existing literature, which consistently highlights oxygenation markers, such as pO_2/FiO_2 and $SatO_2/FiO_2$, as critical indicators of disease severity and prognostic tools across pandemic waves [28–32]. Clinically, thresholds such as a $SatO_2/FiO_2$ ratio below 235 has been correlated with pO_2/FiO_2 below 200, and thus have been linked to increased mortality and ICU admission rates [33, 34]. However, our results indicate that the optimal $SatO_2/FiO_2$ for predicting mortality varied slightly across the waves, with the best threshold in all waves being 227.78. These findings underscore the need for continuous reassessment of these thresholds in clinical practice, as the predictive value of physiological markers may shift over time.

The advantage of $SatO_2/FiO_2$ over pO_2/FiO_2 lies in its non-invasive nature, as it relies on pulse oximetry rather than the more invasive arterial blood gas sampling [33, 35]. This makes $SatO_2/FiO_2$ more accessible for continuous monitoring in clinical settings, particularly in resource-limited environments or for patient populations who may not tolerate invasive procedures. However, the $SatO_2/FiO_2$ ratio has a limitation: if oxygen supplementation reaches a saturation of 100% and the provider administers even higher flow than necessary, the $SatO_2/FiO_2$ ratio may overestimate the severity of the condition, as saturation has a ceiling effect at 100%. In contrast, pO_2/FiO_2 remains unaffected by this ceiling, providing a more accurate assessment of oxygenation regardless of supplemental oxygen flow [33].

This limitation could be particularly relevant in patients receiving advanced pre-ICU respiratory support, such as high-flow nasal oxygen (HFNO). HFNO can improve oxygenation and maintain saturations near 100% through precise FiO_2 delivery and a positive end-expiratory pressure (PEEP)-like effect, potentially masking the severity of hypoxemia when using $SatO_2/FiO_2$ alone [36, 37]. Nevertheless, considering the primary objective of risk stratification at the time of hospital admission, the limitation of the $SatO_2/FiO_2$ variable would have minimal impact on the sensitivity of the assessment. Patients using respiratory devices with high delivered FiO_2 values (e.g., high-flow nasal oxygen and face mask) are, in most cases, those with greater severity.

Inflammatory and hematological markers provided significant prognostic value, being more important than most clinical characteristics. These results align with previous studies that highlight the association between immune dysregulation and disease severity [38, 39]. Studies from earlier stages of the pandemic, which also used machine learning models to predict COVID-19 severity based on clinical blood test data, observed the importance of laboratory parameters such as lymphocyte counts and CRP levels [40, 41]. Additionally, a meta-analysis of 38 studies on biomarkers predicting adverse outcomes in COVID-19 hospitalizations confirmed these findings [42]. Our study extends these findings by including data from all three pandemic waves. Although the relative importance of biomarkers varied over time, elevated CRP, low platelet count, and low lymphocyte values were consistently associated with adverse outcomes during the pandemic. These variations of importance may be related to the evolution of clinical management and the characteristics of the predominant circulating variants at each period [42].

Lymphopenia consistently indicated higher risks for all outcomes, becoming particularly important in wave 3. This finding aligns with previous evidence suggesting that immune suppression plays a significant role in disease progression [43, 44]. A lower lymphocyte count is associated with higher risks of secondary infections, prolonged hospitalization, and increased mortality in other conditions such as sepsis [45, 46], and supports earlier findings that linked lymphopenia with poor COVID-19 outcomes [47]. Interestingly, a higher lymphocyte count was found to be associated with a reduced risk of ICU transfer across all waves, suggesting a potentially protective role of lymphocyte-mediated immune response in mitigating disease severity [47]. Conversely, neutrophil count was directly associated with the need for ICU transfer, particularly during wave 1, which is consistent with previous knowledge that elevated neutrophil levels contribute to the hyperinflammatory response seen in severe COVID-19 cases [48].

Low platelet counts were strongly associated with IMV need and mortality, particularly in wave 2, aligning with evidence of systemic inflammation driving adverse COVID-19 outcomes. Regarding the outcome of death, in wave 3, increased platelet count at admission also emerged as a predictor of mortality. This could be related to chronic inflammatory conditions or iron deficiency, both of which may elevate the risk of poor outcomes [49].

The AST/ALT ratio, a proxy for liver dysfunction, consistently ranked among the top predictors for ICU transfer, emphasizing the multi-organ impact of severe infections. This study used the AST/ALT ratio (De Ritis ratio) rather than each transaminase alone. According to a systematic review of 8 studies (4,606 patients), higher De Ritis ratios were significantly associated with severe disease and mortality in COVID-19 patients, since it reflects the presence of more significant structural and functional liver abnormalities [50]. The study suggests the use of the AST/ALT ratio in association with other prognostic markers to improve its predictive capacity.

Therefore, these biomarkers could aid in early risk stratification and inform clinical decision-making, particularly in identifying patients who may benefit from closer monitoring or targeted interventions. In the context of limited hospital resources and considering the high healthcare costs, early stratification has the potential to assist in the allocation of resources to patients with a higher risk of severe outcomes. This may include measures such as early hospitalization, respiratory monitoring, and timely referral to the ICU. Another way of enhancing their utility could be incorporating these metrics into electronic health records to create trigger alerts for high-risk patients, streamlining triage and ensuring timely interventions. However, clinical decision-making is inherently multifactorial. Rather than relying on rigid thresholds, real-world application requires a comprehensive assessment that takes into account patient history, comorbidities, and other dynamic factors.

Vaccination and pandemic dynamics

Vaccination campaign against COVID-19 in Brazil started on January 17th, 2021, in the beginning of the second wave. Initially, priority was given to some endangered groups, such as healthcare workers, elderly individuals (especially those over 80 years old), people with comorbidities, and indigenous communities. As more vaccines became available, the campaign expanded to include essential workers and the general population, gradually lowering the minimum age for vaccination [51]. By the end of 2021, 90% of the target population (over 177 million people) received at least one dose of the vaccine [52]. Booster doses were introduced in late 2021, mostly addressed to vulnerable groups, and the vaccination effort continued into 2022.

Brazilian vaccination campaign used three types of vaccines: inactivated virus vaccine (CoronaVac®), viral vector vaccine (Covishield; ChAdOx1/AstraZeneca® and Ad26.COV2.S/Jansen®) and messenger RNA vaccine (BNT162b2/Pfizer®) [53]. The impact of the vaccination campaigns could be indirectly observed when analyzing the changes in the profile of hospitalized patients, potentially showing a reduction in the severity of cases, which in turn impacted the outcomes. The mortality peak in wave 2 (20.3%)—the longest and most populous wave—coincided with high ICU transfer (36.0%) and IMV need (27.4%). These rates declined in wave 3, likely reflecting the positive impact of vaccination campaigns, along with improved resource allocation, and evolving treatment strategies [54, 55].

Studies from the same timeframe suggest that emerging variants could have affected clinical outcomes due to differences in transmissibility, immune escape, and disease severity [56, 57]. During the pandemic, genetic sequencing enabled the identification of different SARS-CoV-2 variants, each characterized by distinct virulence and pathogenicity profiles. Initially, strain-specific testing was almost absent. During the second wave, the predominant variant was Delta, whereas the Omicron variant, associated with high pathogenicity but low virulence, drove a surge in COVID-19 cases while causing fewer deaths in wave 3 [5].

Vaccination was an important feature itself in wave 3 regarding the outcome of IMV support. This variable had elevated rates of missing data (available in Table S3) and, as they were incorporated in the analysis, the graph highlighted that the absence of data and presence of vaccination were both factors related to a reduced risk of IMV support. As this data came from medical records, it

is more likely that severe patients had this data checked more frequently than patients with mild symptoms.

Clinical characteristics

In a previous analysis of the first two pandemic waves from our group, age had a great relationship with mortality, but over time this variable became less discriminatory [7]. The present study continued this analysis, including the third wave and revealed that this tendency continued in the third wave regarding the outcome of death. We hypothesize that this may be attributed to the impact of vaccination campaigns or improved management strategies targeting older populations. A recent analysis from our group that compared vaccinated and unvaccinated patients, found out an inverse association between vaccination and in-hospital mortality (adjusted odds ratio [aOR] = 0.42, 95% confidence interval [CI]: 0.31–0.56; $p < 0.001$) and IMV (aOR = 0.40, 95% CI: 0.30–0.53; $p < 0.001$) [58].

In the present analysis, age showed an increasing and complex impact on ICU transfer over the course of the pandemic waves. Aged population sometimes presented a reduced risk of ICU care, especially in wave 3. A turning point was also noticed in the outcome IMV, as, in wave 3, older patients also had a reduced risk of requiring mechanical ventilation.

Advanced age was a major risk factor for unfavorable outcomes. Nevertheless, older adults were the first to be vaccinated [51]. This immunization likely contributed to reducing the impact of age as a risk predictor, although this group of patients continued to have an elevated risk of death. Over time, older patients exhibited a reduced risk of ICU admission and IMV support. Several hypotheses could be proposed to explain these findings: it may be related to ICU bed availability throughout the pandemic, potentially influencing the healthcare team to prioritize younger patients for critical care; alternatively, it may reflect delayed ICU and IMV indications based on patient age.

Respiratory rate's predictive capacity for IMV support decreased across the pandemic waves. We hypothesize that this reduction in the respiratory rate's predictive ability for IMV may be attributed to the phenomenon of 'silent hypoxemia' observed in COVID-19 patients, who often exhibit significant hypoxemia without corresponding respiratory distress [59]. Another possible explanation is that patients were admitted with less severe respiratory conditions, which could have reduced the predictive value of this feature.

Socioeconomic and health system factors

The influence of socioeconomic factors on COVID-19 outcomes has been a subject of growing interest, particularly in understanding health disparities across different regions and income groups [6, 60]. A study that included 119 countries and regions focused on analyzing the impact of income level on mortality during pandemic and observed that mortality rates in developing countries worsened as the pandemic spread, while the death rates declined in developed ones. Although it emphasizes relevant changes in economic influential factors between waves, it does not mention other factors, such as those related to patients or hospitals [6].

While socioeconomic variables, in general, did not show a strong discriminatory capacity for predicting adverse outcomes in the present study, certain trends became evident in specific waves of the pandemic. During the first wave, regional GDP emerged as a risk factor for invasive mechanical ventilation (IMV). Specifically, patients admitted to hospitals located in higher-GDP regions were at a higher risk for orotracheal intubation. This finding could be attributed to the greater availability of ventilatory devices in hospitals in higher-GDP areas [61, 62]. Furthermore, this may reflect the predominance of early intubation protocols recommended at the time, particularly during the initial stages of the pandemic, when management guidelines often advocated for early orotracheal intubation for severe cases. The use of non-invasive ventilation (NIV) and high-flow nasal cannula (HFNC) increased in the later periods of the pandemic. It is probably due to the increased confidence of clinicians in treating COVID-19 patients and the government input acquisition that made this kind of equipment more available [63, 64]. Another relevant change in treatment protocols involved the use of corticosteroids, which increased remarkably and was incorporated as a key part of the treatment, in most protocols during the pandemic [64].

In the second wave, source of income was identified as an important risk factor for ICU transfer. Patients admitted to private hospitals experienced higher rates of ICU transfer compared to those in public institutions. This disparity may be associated with the greater availability of ICU beds in private hospitals, which were better equipped to handle larger numbers of critically ill patients. This suggests that access to healthcare services, particularly intensive care, may have been influenced by socioeconomic factors such as income and the type of hospital (public vs. private). The findings reflect ongoing disparities in healthcare access, with previous studies documenting how differences in socioeconomic status can impact the level of care patients receive [65, 66].

The results of our study align with broader research, including a meta-analysis of 4.3 million patients from 68 studies and two large multicenter cohort studies that also applied machine learning models, one of which included over 4.9 million patients. These studies documented that socioeconomic factors, such as income and

regional GDP, contribute to inequities in healthcare outcomes [67–69]. In regions with higher GDP, the availability of medical resources such as ventilators and ICU beds was more readily accessible, potentially influencing patient management and survival outcomes. Conversely, in lower-GDP regions, limited resources may have led to delayed or less aggressive interventions, which could contribute to worse outcomes [61]. Additionally, comorbidities, such as hypertension or obesity, are more prevalent among lower resource populations [70, 71], thus contributing to worsened disease outcomes [69].

Disparities in public health measures could, also, indirectly impact the results observed in the study. Restrictive measures such as lockdowns and social distancing were widely adopted in the country, although the implementation varied municipally. Over time, with the progressive relaxation of these measures, shifts in hospitalization patterns and disease severity may have occurred [72].

The second wave represented a period of higher mortality, in line with the increased hospital demand for admissions and the collapse of some healthcare systems in the country due to shortages of medical supplies, especially in less developed regions. On the other hand, during the third wave, despite the high numbers of positive cases nationwide, a smaller proportion of these patients required hospitalization. This coincided with the period when most cases were caused by the Omicron variant, which is known to have lower lethality [5].

This analysis highlights the need for policies aimed at addressing healthcare disparities and ensuring equitable access to critical care services, particularly in resource-limited settings. Future studies should further investigate the interplay between socioeconomic factors, healthcare infrastructure, and patient outcomes to better understand and mitigate health inequities.

This study was conducted in an upper-middle income country, where half of the participating hospitals exclusively treated patients within the public healthcare system. Another way to assess the indirect impact of socioeconomic characteristics relates to the use of high-cost specific therapies, such as remdesivir. This medication was approved by the Brazilian regulatory agency for hospital use in April 2021. It inhibits coronavirus replication in the body, reducing the infection process. However, data on its impact on mortality was conflicting [73, 74]. These uncertainties, combined with the high cost of the medication, led the Brazilian National Commission for the Incorporation of Technologies in the Unified Health System (CONITEC) to issue a conditional recommendation against the use of the medication in hospitalized COVID-19 patients [75]. Consequently, remdesivir was excluded from the public healthcare system. Furthermore, it was not established as standard treatment in the private healthcare sector. Health insurance providers typically rely on cost-effectiveness analyses and clinical guidelines, which may have contributed to its omission as a routine option for hospitalized patients. This dual limitation of access—both in public and private healthcare—significantly restricted its use in Brazil, despite its availability in other countries. Consequently, the medication was not incorporated into most institutional protocols, resulting in a very limited utilization rate: only 5 out of the 16,957 patients received remdesivir.

Strengths and limitations

This study presents several strengths. First, the dataset utilized is robust and comprehensive, encompassing data from three distinct COVID-19 waves in Brazil, one of the countries most severely impacted by the pandemic. This temporal coverage allows for a deeper understanding of the evolving nature of the disease and the associated clinical outcomes over time. Second, the use of advanced machine learning methods, specifically LightGBM and SHAP, represents a novel approach in this field. Most previous studies relied on logistic regression to analyze and understand potential risk factors in infectious diseases [76, 77], despite its notable limitations. Logistic regression and LightGBM are both widely used statistical and machine learning algorithms, but they differ significantly in their approach and application. Logistic regression is a linear model primarily used for binary and multinomial classification tasks. It predicts the probability of a class using a linear combination of input features and applies a sigmoid function to produce outputs between 0 and 1. Logistic regression is simple, interpretable, and works well when features have a linear relationship with the target variable. However, it struggles with complex, non-linear patterns and interactions between variables, making it less effective for high-dimensional, which is the case of the data we used in our paper. Also, it requires complete datasets or imputation methods for missing data, and may produce biased results in the presence of correlated data [10, 11].

In contrast, LightGBM is an advanced tree-based algorithm designed for speed and efficiency. Unlike logistic regression, LightGBM can capture intricate patterns by constructing decision trees, being well-suited for handling categorical features, missing values, and high-dimensional data. Due to these properties, LightGBM usually produces higher accuracy (better predictive capability) than logistic regression in data with the above characteristics, which is exactly the case for the datasets used in this article. Studies in other areas of medicine have also reinforced the higher accuracy of machine learning predictive models compared to logistic regression [78–80].

LightGBM is, however, more complex and thus harder to interpret than logistic regression. To deal with the

interpretability issues, we couple our LightGBM results with the Explainable Machine Learning SHAP technique to perform our interpretability analyses of the temporal drifts across the COVID-19 waves. Additionally, the model's robustness to missing data, as demonstrated by its stability across different imputation strategies, underscores its reliability in real-world settings, where missingness is often informative and non-random.

This study also has limitations. Since it was a retrospective analysis of patient records review, it suffered from recording bias and missing data. The hospitals participating were not randomly chosen and were mostly located in the wealthiest areas of the country, thus, it is not possible to state that it is representative of the Brazilian reality. Virus sequencing was not carried out in our cohort, therefore, it is impossible to define the specific impact of each variant in our findings.

Unmeasurable confounding factors may also have impacted the results. Healthcare team training was not standardized across hospitals, no uniform treatment protocol was implemented, and regional resource limitations may have influenced therapeutic decisions and, consequently, patient outcomes. Additionally, differences in access to care and regional policies could have introduced significant variability in the management of COVID-19 patients, which might not have been fully accounted for in our analysis. For example, patients in regions with higher resource availability may have had access to more timely interventions or advanced care, potentially affecting outcomes. Similarly, regional policies related to vaccination efforts, or healthcare prioritization might have influenced the severity of illness and patient outcomes.

Although explainable AI techniques have demonstrated to be very useful for drift analyses, there are some limitations. There is risk of bias on the way the data used to train the model was collected and obtained. Incomplete or inconsistent data due to the methods used to obtain the original data in the hospitals participating in the cohort may also affect some of the analyses. While explainable AI techniques used in the present analysis are state-of-the-art, most of the inferences are based on correlations, and we cannot make any claims of causal relationships.

Although the models demonstrated good performance through internal validation techniques, further studies are necessary to externally validate them in different populations and settings, particularly outside Brazil. Finally, the use of IA in health applications should always be moderated by humans; the final decision should be made by health specialists to ensure ethical and contextually appropriate care.

Conclusion

This study demonstrated the ability of machine learning models to identify complex and dynamic patterns in healthcare data. Data drifts were observed in all scenarios, affecting potential predictive capabilities of explainable machine learning methods. Effectively incorporating knowledge about these drifts into predictive models remains a significant technical challenge, requiring continuous updates and performance monitoring in real-world applications.

Biochemical and physiological markers, particularly $SatO_2/FiO_2$, played a significant role in predicting adverse outcomes in COVID-19 patients. $SatO_2/FiO_2$ at admission was the sole variable consistently at the top two most important variables for all outcomes across all three waves; higher values were strongly related to a reduced risk of adverse outcomes. Inflammatory response markers at admission were more important than patients' clinical characteristics in predicting adverse outcomes. Among them, platelets and lymphocytes at admission were the inflammatory response markers with better predictive capacity. The temporal shifts in predictors such as age, lymphopenia, and platelet count suggest evolving clinical patterns and underscore the complexity of COVID-19 progression. These findings could help in early identification of borderline patients at risk of unfavorable outcomes.

Abbreviations

°C	Degrees celsius
ALT	Alanine aminotransferase
ANS	National Health Agency (Agência Nacional de Saúde)
aOR	adjusted odds ratio
AST	Aspartate aminotransferase
BRL	Brazilian reais
CONEP	Comissão Nacional de Ética em Pesquisa– Brazilian National Research Ethics Comittee
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 19
CNES	National Registry of Health Facilities (Cadastro Nacional de Estabelecimentos da Saúde)
CRP	C-reactive protein
K-fold CV	K-fold cross-validation
FiO2	Fraction of inspired oxygen
GDP	Gross domestic product
GCS	Glasgow Coma Scale
HCO3	Bicarbonate
HDI	Human development index
HIV	Human immunodeficiency virus
IBGE	Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística)
IBM SPSS	International Business Machines Corporation Statistical Package for the Social Sciences
ICU	Intensive care unit
IMV	Invasive mechanical ventilation
IQR	Interquartile ranges
LightGBM	Light Gradient-Boosting Machine
LR	Logistic regression
MICE	Multiple imputation with chained equations
mmHg	Millimeters of mercury
N/A	Not applicable
ONA	National Accreditation Organization (Organização Nacional de Acreditação)

OR	Odds ratio
pCO2	partial pressure of carbon dioxide
pO2	partial pressure of oxygen
pH	potential of hydrogen
REDCap®	Research Electronic Data Capture
SX	Supplementary X
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SatO2	Oxygen saturation
SHAP	Shapley Additive ExPlanation
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
ULR	Upper level reference
VTE	Venous thromboembolism

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-025-10766-0.

Supplementary Material 1

Acknowledgements
We would like to thank the hospitals, which were part of this collaboration. We also thank all the clinical staff at those hospitals, all undergraduate students and members of both cohorts, for their efforts in collecting data.

Author contributions
Substantial contributions for the conception or design of the manuscript: FCBL, CCM, BBMP, MAG, LCDR, MSM, LRV. Substantial contributions for data acquisition, analysis or interpretation: FCBL, BBMP, GFN, LCDR, MC, JDLB, FA, PGP, FB, CCRC, AVS, DRAR, MAG, MSM. Drafted the work: FCBL, CCM, BBMP, LRV, MAG, and MSM. Revised the manuscript critically for important intellectual content: FCBL, CCM, BBMP, LRV, GFN, LCDR, MC, JDLB, FA, PGP, FB, CCRC, AVS, DRAR, MAG, MSM. Final approval of the version to be published: FCBL, CCM, BBMP, LRV, GFN, LCDR, MC, JDLB, FA, PGP, FB, CCRC, AVS, DRAR, MAG, MSM.

Funding
This study was supported in part by Minas Gerais State Agency for Research and Development (***Fundação de Amparo à Pesquisa do Estado de Minas Gerais***-FAPEMIG) [grant number APQ-01154–21] and National Institute of Science and Technology for Health Technology Assessment (***Instituto de Avaliação de Tecnologias em Saúde***-IATS) [grant number 465518/2014-1]. LRV was supported by CNPq [grant number 134878/2023–0]. MSM was supported in part by CNPq [grant number 310561/2021–3]. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Data availability
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The study was approved by the Brazilian National Research Ethics Committee (*Comissão Nacional de Ética em Pesquisa*– CONEP), Certificate of Presentation for Ethical Assessment 30350820.5.1001.0008, as well as internal approval of ethics boards from each hospital. Individual informed consent was waived due to the pandemic situation and the use of deidentified data, based on medical chart review only. The study adhered to the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Universidade Federal de Minas Gerais, Av. Presidente Antônio Carlos, Belo Horizonte 6627, Brazil
[2]Department of Internal Medicine, Medical School & Telehealth Center, University Hospital, Universidade Federal de Minas Gerais, Av. Professor Alfredo Balena, Belo Horizonte 110, Brazil
[3]Computer Science Department, Universidade Federal de Minas Gerais, Av. Presidente Antônio Carlos, Belo Horizonte 6627, Brazil
[4]Universidade Federal de São João del-Rei. Praça Dom Helvécio, São João del-Rei 74, Brazil
[5]Hospital Santa Cruz. R. Fernando Abott, Santa Cruz do Sul 174, Brazil
[6]Hospital Regional do Oeste. R. Florianópolis, Chapecó 1448 E, Brazil
[7]Hospital Nossa Senhora da Conceição, Av. Francisco Trein, Porto Alegre 326, Brazil
[8]Instituto Orizonti. Praça Engenheiro Flávio Gutierrez, Belo Horizonte, Brazil
[9]Hospital Santo Antônio, Praça Dr. Márcio Carvalho Lopes Filho, Curvelo 501, Brazil
[10]Hospital Santa Rosália, R. Dr. Onofre, 575, Centro, Teófilo Otoni, Brazil
[11]Department of Internal Medicine, Universidade Federal de Santa Maria, Santa Maria, Brazil
[12]Universidade Federal de São, João del-Rei/ Campus Centro Oeste Dona Lindu, Divinópolis, Brazil
[13]Institute for Health and Technology Assessment. R. Ramiro Barcelos, Porto Alegre 2350, Brazil

Received: 30 November 2024 / Accepted: 7 March 2025
Published online: 15 April 2025

References

1. World health statistics. 2022: monitoring health for the SDGs, sustainable development goals. World Health Organization; 19 May 2022 [cited 4 May 2024]. Available: https://www.who.int/publications/i/item/9789240051157
2. Cimerman S, Chebabo A, da Cunha CA, Rodríguez-Morales AJ. Deep impact of COVID-19 in the healthcare of Latin America: the case of Brazil. Braz J Infect Dis. 2020;24:93–5. https://doi.org/10.1016/j.bjid.2020.04.005.
3. El-Shabasy RM, Nayel MA, Taher MM, Abdelmonem R, Shoueir KR, Kenawy ER. Three waves changes, new variant strains, and vaccination effect against COVID-19 pandemic. Int J Biol Macromol. 2022;204:161–8. https://doi.org/10.1016/j.ijbiomac.2022.01.118.
4. Coronavirus disease (COVID-19).– World Health Organization. [cited 18 Sep 2024]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019
5. Moura EC, Cortez-Escalante J, Cavalcante FV, Barreto IC, de Sanchez HC, Santos MN. Covid-19: Temporal evolution and immunization in the three epidemiological waves, Brazil, 2020–2022. Rev Saude Publica. 2022;56:105. https://doi.org/10.11606/s1518-8787.2022056004907.
6. Yao L, Graff JC, Aleya L, Ma J, Cao Y, Wei W, et al. Mortality in four waves of COVID-19 is differently associated with healthcare capacities affected by economic disparities. Trop Med Infect Dis. 2022;7. https://doi.org/10.3390/tropicalmed7090241.
7. Paiva BBM, Pereira PD, Gomes VMR, Silva MVR, Valiense C, Marcolino MS et al. Characterizing and Understanding Temporal Effects in COVID-19 Data. Workshop on Healthcare AI and COVID-19. PMLR; 2022. pp. 33–40. Available: https://proceedings.mlr.press/v184/paiva22a.html
8. Souza-Silva MVR, Ziegelmann PK, Nobre V, Gomes VMR, Etges APB da, Schwarzbold S et al. AV,. Hospital characteristics associated with COVID-19 mortality: data from the multicenter cohort Brazilian Registry. Intern Emerg Med. 2022;17: 2299–2313. https://doi.org/10.1007/s11739-022-03092-9
9. Suri A, Singh NK, Perumal V. Association of inflammatory biomarker abnormalities with mortality in COVID-19: a meta-analysis. Bull Natl Salmon Resour Cent. 2022;46:54. https://doi.org/10.1186/s42269-022-00733-z.
10. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: logistic regression. Perspect Clin Res. 2017;8:148–51. https://doi.org/10.4103/picr.PICR_87_17.
11. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems. 2017;30. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
12. Shakibfar S, Nyberg F, Li H, Zhao J, Nordeng HME, Sandve GKF, et al. Artificial intelligence-driven prediction of COVID-19-related hospitalization and death: a systematic review. Front Public Health. 2023;11:1183725. https://doi.org/10.3389/fpubh.2023.1183725.

13. Marcolino MS, Ziegelmann PK, Souza-Silva MVR, Nascimento IJB, Oliveira LM, Monteiro LS, et al. Clinical characteristics and outcomes of patients hospitalized with COVID-19 in Brazil: results from the Brazilian COVID-19 registry. Int J Infect Dis. 2021;107:300–10. https://doi.org/10.1016/j.ijid.2021.01.019.
14. Diagnostic Testing for SARS-CoV-2: Interim Guidance. 2020. Available: https://books.google.com/books/about/Diagnostic_Testing_for_SARS_CoV_2_Interi.html?hl=%26;id=n6Yv0AEACAAJ
15. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335:806–8. https://doi.org/10.1136/bmj.39335.541782.AD.
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inf. 2009;42:377–81. https://doi.org/10.1016/j.jbi.2008.08.010.
17. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inf. 2019;95:103208. https://doi.org/10.1016/j.jbi.2019.103208.
18. Bicalho MAC, Aliberti MJR, Delfino-Pereira P, Chagas VS, Rosa PM da, Pires S et al. MC,. Clinical characteristics and outcomes of COVID-19 patients with pre-existing dementia: a large multicenter propensity-matched Brazilian cohort study. BMC Geriatr. 2024;24: 25. https://doi.org/10.1186/s12877-023-04494-w
19. Instituto Brasileiro de Geografia e Estatística. [cited 18 Sep 2024]. Available: https://cidades.ibge.gov.br/
20. Cadastro Nacional dos Estabelecimentos de Saúde. In: DATASUS [Internet]. [cited 5 Feb 2023]. Available: https://cnes.datasus.gov.br/
21. ONA. In: ONA| Organização Nacional de Acreditação [Internet]. [cited 18 Sep 2024]. Available: https://www.ona.org.br/
22. Informações e Avaliações de Operadoras. In: Agência Nacional de Saúde Suplementar [Internet]. [cited 18 Sep 2024]. Available: https://www.gov.br/ans/pt-br/assuntos/informacoes-e-avaliacoes-de-operadoras
23. Chen H, Janizek J, Lundberg S, I Lee S. True to the model or true to the data? ArXiv. 2020;csLG. https://doi.org/10.48550/ARXIV.2006.16234.
24. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media; 2013. Available: https://play.google.com/store/books/details?id=yPfZBwAAQBAJ
25. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022;12:5979. https://doi.org/10.1038/s41598-022-09954-8.
26. Alberdi-Iglesias A, Martín-Rodríguez F, Ortega Rabbione G, Rubio-Babiano AI, Núñez-Toste MG, Sanz-García A, et al. Role of SpO2/FiO2 ratio and ROX index in predicting early invasive mechanical ventilation in COVID-19. A pragmatic, retrospective, Multi-Center study. Biomedicines. 2021;9. https://doi.org/10.3390/biomedicines9081036.
27. Kim JH, Baek A-R, Lee S-I, Kim W-Y, Na YS, Lee BY, et al. ROX index and SpO2/FiO2 ratio for predicting high-flow nasal cannula failure in hypoxemic COVID-19 patients: A multicenter retrospective study. PLoS ONE. 2022;17:e0268431. https://doi.org/10.1371/journal.pone.0268431.
28. Wick KD, Matthay MA, Ware LB. Pulse oximetry for the diagnosis and management of acute respiratory distress syndrome. Lancet Respir Med. 2022;10:1086–98. https://doi.org/10.1016/S2213-2600(22)00058-3.
29. Hu J, Wang Y. The clinical characteristics and risk factors of severe COVID-19. Gerontology. 2021;67:255–66. https://doi.org/10.1159/000513400.
30. San Martín-López JV, Mesa N, Bernal-Bello D, Morales-Ortega A, Rivilla M, Guerrero M, et al. Seven epidemic waves of COVID-19 in a hospital in Madrid: analysis of severity and associated factors. Viruses. 2023;15. https://doi.org/10.3390/v15091839.
31. Mihajlovic S, Nikolic D, Santric-Milicevic M, Milicic B, Rovcanin M, Acimovic A, et al. Four waves of the COVID-19 pandemic: comparison of clinical and pregnancy outcomes. Viruses. 2022;14. https://doi.org/10.3390/v14122648.
32. Blanca D, Nicolosi S, Bandera A, Blasi F, Mantero M, Hu C, et al. Comparison between the first and second COVID-19 waves in internal medicine wards in Milan, Italy: a retrospective observational study. Intern Emerg Med. 2022;17:2219–28. https://doi.org/10.1007/s11739-022-03052-3.
33. Rice TW, Wheeler AP, Bernard GR, Hayden DL, Schoenfeld DA, Ware LB, et al. Comparison of the SpO2/FIO2 ratio and the PaO2/FIO2 ratio in patients with acute lung injury or ARDS. Chest. 2007;132:410–7. https://doi.org/10.1378/chest.07-0617.
34. Patel S, Singh G, Zarbiv S, Ghiassi K, Rachoin J-S. Mortality prediction using SaO/FiO ratio based on eICU database analysis. Crit Care Res Pract. 2021;2021:6672603. https://doi.org/10.1155/2021/6672603.
35. Lam E, Paz SG, Goddard-Harte D, Pak YN, Fogel J, Rubinstein S. Respiratory involvement parameters in hospitalized COVID-19 patients and their association with mortality and length of stay. Can J Respir Ther. 2022;58:1–8. https://doi.org/10.29390/cjrt-2021-057.
36. Park S. High-flow nasal cannula for respiratory failure in adult patients. Acute Crit Care. 2021;36:275–85. https://doi.org/10.4266/acc.2021.01571.
37. Li J, Albuainain FA, Tan W, Scott JB, Roca O, Mauri T. The effects of flow settings during high-flow nasal cannula support for adult subjects: a systematic review. Crit Care. 2023;27:78. https://doi.org/10.1186/s13054-023-04361-5.
38. Gómez-Romero FJ, Muñoz-Rodríguez JR, Serrano-Oviedo L, García-Jabalera I, López-Juárez P, Pérez-Ortiz JM, et al. Clinical features and mortality of COVID-19 patients admitted to ICU according to SOFA score. Med (Baltim). 2022;101:e29206. https://doi.org/10.1097/MD.0000000000029206.
39. Chen Y, Klein SL, Garibaldi BT, Li H, Wu C, Osevala NM, et al. Aging in COVID-19: vulnerability, immunity and intervention. Ageing Res Rev. 2021;65:101205. https://doi.org/10.1016/j.arr.2020.101205.
40. Aktar S, Ahamad MM, Rashed-Al-Mahfuz M, Azad A, Uddin S, Kamal A, et al. Machine learning approach to predicting COVID-19 disease severity based on clinical blood test data: statistical analysis and model development. JMIR Med Inf. 2021;9:e25884. https://doi.org/10.2196/25884.
41. Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. Mod Pathol. 2021;34:522–31. https://doi.org/10.1038/s41379-020-00700-x.
42. 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 Evid Based Med. 2021;26:107–8. https://doi.org/10.1136/bmjebm-2020-111536.
43. Ponsford MJ, Ward TJC, Stoneham SM, Dallimore CM, Sham D, Osman K, et al. A systematic review and Meta-Analysis of inpatient mortality associated with nosocomial and community COVID-19 exposes the vulnerability of immunosuppressed adults. Front Immunol. 2021;12:744696. https://doi.org/10.3389/fimmu.2021.744696.
44. Vaid N, Ardissino M, Reed TAN, Goodall J, Utting P, Miscampbell M, et al. Clinical characteristics and outcomes of immunosuppressed patients hospitalized with COVID-19: experience from London. J Intern Med. 2021;289:385–94. https://doi.org/10.1111/joim.13172.
45. Amin CA, Liana P, Ahmad Z, Hidayat R, Sari NP, Afifah AR, et al. Lymphocyte levels in predicting the outcome of COVID-19 patient: A prognostic study from single center in Indonesia. J Nepal Health Res Counc. 2021;19:536–42. https://doi.org/10.33314/jnhrc.v19i3.3529.
46. Wang S, Sheng Y, Tu J, Zhang L. Association between peripheral lymphocyte count and the mortality risk of COVID-19 inpatients. BMC Pulm Med. 2021;21:55. https://doi.org/10.1186/s12890-021-01422-9.
47. Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang Y-Q, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct Target Ther. 2020;5:33. https://doi.org/10.1038/s41392-020-0148-4.
48. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol. 2020;20:355–62. https://doi.org/10.1038/s41577-020-0331-4.
49. Mirsaeidi M, Peyrani P, Aliberti S, Filardo G, Bordon J, Blasi F, et al. Thrombocytopenia and thrombocytosis at time of hospitalization predict mortality in patients with community-acquired pneumonia. Chest. 2010;137:416–20. https://doi.org/10.1378/chest.09-0998.
50. Pranata R, Huang I, Lim MA, Yonas E, Vania R, Lukito AA, et al. Elevated de Ritis ratio is associated with poor prognosis in COVID-19: A systematic review and Meta-Analysis. Front Med (Lausanne). 2021;8:676581. https://doi.org/10.3389/fmed.2021.676581.
51. Entenda a ordem. de vacinação contra a Covid-19 entre os grupos prioritários. In: Ministério da Saúde [Internet]. 28 Jan 2021 [cited 29 Nov 2024]. Available: https://www.gov.br/saude/pt-br/assuntos/noticias/2021/janeiro/entenda-a-ordem-de-vacinacao-contra-a-covid-19-entre-os-grupos-prioritarios
52. Retrospectiva. 2021: as milhões de vacinas Covid-19 que trouxeram esperança para o Brasil. In: Ministério da Saúde [Internet]. 30 Dec 2021 [cited 29 Nov 2024]. Available: https://www.gov.br/saude/pt-br/assuntos/noticias/2021/dezembro/retrospectiva-2021-as-milhoes-de-vacinas-covid-19-que-trouxeram-esperanca-para-o-brasil
53. Coronavírus Brasil. [cited 13 Jan 2025]. Available: https://covid.saude.gov.br/
54. Harvey J, Chan B, Srivastava T, Zarebski AE, Dłotko P, Błaszczyk P, et al. Epidemiological waves - Types, drivers and modulators in the COVID-19 pandemic. Heliyon. 2023;9:e16015. https://doi.org/10.1016/j.heliyon.2023.e16015.

55. Liu Q, Qin C, Liu M, Liu J. Effectiveness and safety of SARS-CoV-2 vaccine in real-world studies: a systematic review and meta-analysis. Infect Dis Poverty. 2021;10:132. https://doi.org/10.1186/s40249-021-00915-3.
56. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of Spike protein. J Med Virol. 2022;94:1641–9. https://doi.org/10.1002/jmv.27526.
57. Chatterjee S, Bhattacharya M, Nag S, Dhama K, Chakraborty C. A detailed overview of SARS-CoV-2 Omicron: its Sub-Variants, mutations and pathophysiology, clinical characteristics, immunological landscape, immune escape, and therapies. Viruses. 2023;15. https://doi.org/10.3390/v15010167.
58. al. de M et. Comprehensive statistical analysis reveals significant benefits of COVID-19 vaccination in hospitalized patients: propensity score, covariate adjustment, and feature importance by permutation. BMC Infect Disease. 2024;24. https://doi.org/10.1186/s12879-024-09865-1
59. Laredo M, Vandiedonck C, Miró Ò, Del González J, Alquézar-Arbé A, Jacob J, et al. Are there differences in the relationship between respiratory rate and oxygen saturation between patients with COVID-19 and those without COVID-19? Insights from a cohort-based correlational study. Emerg Med J. 2023;40:805–9. https://doi.org/10.1136/emermed-2022-212882.
60. Khanijahani A, Iezadi S, Gholipour K, Azami-Aghdash S, Naghibi D. A systematic review of Racial/ethnic and socioeconomic disparities in COVID-19. Int J Equity Health. 2021;20:248. https://doi.org/10.1186/s12939-021-01582-4.
61. Beck da Silva Etges AP, Bertoglio Cardoso R, Marcolino MS, Brasil Ruschel K, Coutinho AP, Pereira EC, et al. The economic impact of COVID-19 treatment at a Hospital-level: investment and financial registers of Brazilian hospitals. J Health Econ Outcomes Res. 2021;8:36–41. https://doi.org/10.36469/jheor.2021.22066.
62. McMahon DE, Peters GA, Ivers LC, Freeman EE. Global resource shortages during COVID-19: bad news for low-income countries. PLoS Negl Trop Dis. 2020;14:e0008412. https://doi.org/10.1371/journal.pntd.0008412.
63. Riera J, Barbeta E, Tormos A, Mellado-Artigas R, Ceccato A, Motos A, et al. Effects of intubation timing in patients with COVID-19 throughout the four waves of the pandemic: a matched analysis. Eur Respir J. 2023;61. https://doi.org/10.1183/13993003.01426-2022.
64. Carbonell R, Urgelés S, Rodríguez A, Bodí M, Martín-Loeches I, Solé-Violán J, et al. Mortality comparison between the first and second/third waves among 3,795 critical COVID-19 patients with pneumonia admitted to the ICU: A multicentre retrospective cohort study. Lancet Reg Health Eur. 2021;11:100243. https://doi.org/10.1016/j.lanepe.2021.100243.
65. McHenry RD, Moultrie CEJ, Quasim T, Mackay DF, Pell JP. Association between socioeconomic status and outcomes in critical care: A systematic review and Meta-Analysis. Crit Care Med. 2023;51:347–56. https://doi.org/10.1097/CCM.0000000000005765.
66. Nordberg P, Jonsson M, Hollenberg J, Ringh M, Kiiski Berggren R, Hofmann R, et al. Immigrant background and socioeconomic status are associated with severe COVID-19 requiring intensive care. Sci Rep. 2022;12:12133. https://doi.org/10.1038/s41598-022-15884-2.
67. Carvantes-Barrera A, Díaz-González L, Rosales-Rivera M, Chávez-Almazán LA. Risk factors associated with COVID-19 lethality: A machine learning approach using Mexico database. J Med Syst. 2023;47:90. https://doi.org/10.1007/s10916-023-01979-4.
68. Rahman R, Kang J, Rousseau JF, Ding Y. Analyzing Impact of Socio-Economic Factors on COVID-19 Mortality Prediction Using SHAP Value. AMIA Annu Symp Proc. 2022;2022: 892–901. Available: https://www.ncbi.nlm.nih.gov/pubmed/37128465
69. Magesh S, John D, Li WT, Li Y, Mattingly-App A, Jain S, et al. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: A Systematic-Review and Meta-analysis. JAMA Netw Open. 2021;4:e2134147. https://doi.org/10.1001/jamanetworkopen.2021.34147.
70. Seifu CN, Fahey PP, Atlantis E. Unhealthy diet pattern mediates the disproportionate prevalence of obesity among adults with Socio-Economic disadvantage: an Australian representative Cross-Sectional study. Nutrients. 2021;13. https://doi.org/10.3390/nu13041363.
71. Sarki AM, Nduka CU, Stranges S, Kandala N-B, Uthman OA. Prevalence of Hypertension in Low- and Middle-Income Countries: A Systematic Review and Meta-Analysis. Medicine (Baltimore). 2015;94: e1959. https://doi.org/10.1097/MD.0000000000001959
72. da Silva LLS, Lima AFR, Polli DA, Razia PFS, Pavão LFA, de Cavalcanti MAF. Medidas de distanciamento social Para O Enfrentamento Da COVID-19 no Brasil: caracterização e análise epidemiológica Por Estado. Cad Saúde Pública. 2020;36:e00185020. https://doi.org/10.1590/0102-311X00185020.
73. Reddy Vegivinti CT, Pederson JM, Saravu K, Gupta N, Barrett A, Davis AR, et al. Remdesivir therapy in patients with COVID-19: A systematic review and meta-analysis of randomized controlled trials. Ann Med Surg (Lond). 2021;62:43–8. https://doi.org/10.1016/j.amsu.2020.12.051.
74. Tanni1 SE. Silvinato2 A, Floriano3 I, Bacha4 HA, Barbosa5 AN, Bernardo6 WM. Use of Remdesivir in patients with COVID-19: a systematic review and meta-analysis. J Bras Pneumol. 2022;48:e20210393–20210393. https://doi.org/10.36416/1806-3756/e20210393.
75. Diretrizes Brasileiras Para Tratamento Hospitalar Do Paciente Com COVID-19-. Capítulo 2: Tratamento Medicamentoso. [cited 13 Jan 2025]. Available: https://www.gov.br/conitec/pt-br/midias/relatorios/diretrizesbrasileiras_tratamentohospitalarpaciente_capii.pdf
76. Dessie ZG, Zewotir T. Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients. BMC Infect Dis. 2021;21:855. https://doi.org/10.1186/s12879-021-06536-3.
77. Li J, Huang DQ, Zou B, Yang H, Hui WZ, Rui F, et al. Epidemiology of COVID-19: A systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. J Med Virol. 2021;93:1449–58. https://doi.org/10.1002/jmv.26424.
78. Singal AG, Mukherjee A, Elmunzer BJ, Higgins PDR, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol. 2013;108:1723–30. https://doi.org/10.1038/ajg.2013.332.
79. Lee H-C, Yoon SB, Yang S-M, Kim WH, Ryu H-G, Jung C-W, et al. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. Logistic regression model. J Clin Med. 2018;7:428. https://doi.org/10.3390/jcm7110428.
80. Xue Y, Chen S, Zhang M, Cai X, Zheng J, Wang S, et al. The prediction models for high-risk population of stroke based on logistic regressive analysis and LightGBM algorithm separately. Iran J Public Health. 2022;51:999–1009. https://doi.org/10.18502/ijph.v51i5.9415.

Publisher's note