Evaluation of potential predictors for mechanical ventilation and mortality in covid-19 patients using self-explainable machine learning methods

1.

a. Abstract:

i. Objective

To determine potential predictors for invasive mechanical ventilation (IMV) and mortality in COVID-19-associated acute respiratory distress syndrome (C-ARDS) using machine learning (ML) techniques.

ii. Design

Single center highly detailed longitudinal observational study.

iii. Setting

Tertiary hospital ICU, during the two first COVID19 pandemic waves in Madrid, Spain.

iv. Patients

280 patients with C-ARDS, not requiring IMV on admission.

v. Interventions

Retrospective and observational study of anonymized time series records of patient clusters up to the 5th day of ICU stay. The researchers applied a dual-step machine learning methodology to identify independent variables associated with IMV and mortality.
The study involved a retrospective analysis of anonymized time series records of patient clusters up to the 5th day of ICU stay.

vi. Main variables

Target: endotracheal intubation and IMV, mortality.
Predictors: demographics, hourly evolution of oxygenation, clinical data, and laboratory results.

vii. Results

The time between symptom onset and ICU admission, the APACHE II score, the ROX index, and procalcitonin levels in blood were potential predictors related to both IMV and mortality. The ROX index was the most significant predictor associated with IMV, while APACHE II, LDH, and DaysSympICU were the most with mortality.

viii. Conclusions

R-part classification identify significant predictors linked with IMV and mortality in C-ARDS patients, including the time between symptom onset and

ICU admission, the severity of the COVID-19 waves, and several clinical and laboratory measures. These findings may help clinicians to better identify patients at risk for IMV and mortality and improve their management.

b. Resumen:

i. Objetivo

Determinar potenciales predictores para ventilación mecánica invasiva (VMI) y mortalidad en el síndrome de distrés respiratorio agudo asociado con COVID-19 (C-SDRA) utilizando técnicas de aprendizaje automático (AA).

ii. Diseño

Estudio observacional longitudinal en un solo centro.

iii. Ámbito

UCI de un hospital terciario durante las dos primeras olas de COVID-19 en Madrid, España.

iv. Pacientes

280 pacientes con C-SDRA que no requieren VMI al ingreso en UCI.

v. Intervenciones

Estudio retrospectivo observacional sobre datos temporales en clústers de pacientes con ingreso en UCI hasta el quinto día de estancia, identificando variables independientes y sus umbrales asociados con VMI y mortalidad.

vi. Principales variables

Objetivo: VMI y Mortalidad.
Predictores: demográficos, variables clínicas, resultados de laboratorio y evolución de la oxigenación.

vii. Resultados

El tiempo entre el inicio de los síntomas y el ingreso en la UCI, la puntuación APACHE II, el índice ROX y los niveles de procalcitonina en sangre eran posibles predictores relacionados tanto con la IMV como con la mortalidad. El índice ROX fue el predictor más significativo asociada con la IMV, mientras que APACHE II, LDH y DaysSympICU fueron los más influyentes en la mortalidad.

viii. Conclusiones

El clasificador R-part identifica predictores significativos vinculados con la VMI y mortalidad en pacientes con C-ARDS, incluido el tiempo entre el inicio de los síntomas y el ingreso en la UCI, la gravedad de las olas de COVID-19 y varias medidas clínicas y de laboratorio. Estos hallazgos pueden ayudar a los médicos a identificar mejor a los pacientes en riesgo de IMV y mortalidad y mejorar su manejo.

Keywords/Palabras Clave: Acute Respiratory Distress Syndrome, Invasive Mechanical Ventilation, COVID-19, Machine Learning, Artificial Intelligence, Predictors.

b. Text

a. Introduction

Invasive mechanical ventilation (IMV) is a cornerstone of organ support in severe COVID-19 patients with acute respiratory distress syndrome (ARDS). As widely experienced in ICUs during the SARS-CoV-2 pandemic, IMV frequently causes complications (1,2) . Hospital services were overwhelmed not only by the surge of patients, but also by scarce human resources and equipment, lack of sufficient mechanical ventilators being probably the most relevant. In surge scenarios, appropriate triage strategies are therefore needed to allocate IMV or alternatives such as high flow nasal prongs. These strategies should be based on the knowledge and understanding of specific potential predictors that could help clinicians to personalize decisions regarding IMV.

There is still considerable controversy regarding who and when to intubate. Several recent studies have addressed the subject (3), although bias cannot be excluded in observational non-randomized trials. A retrospective study suggested that early intubation and IMV is associated with favorable outcomes but included only intubated patients instead of the whole population at risk.

Previous studies have identified covid-19 progression predictors including age, comorbidities, renal function, or immunodeficiency (4) using traditional statistical approaches, where collinearity of data cannot be ruled out. Artificial intelligence (AI) is currently being used for COVID-19 risk stratification (5), studying multiple clinical features to increase effectiveness and efficiency in diagnosis, treatment, and prognosis. Self-explainable Machine learning (ML) techniques can help with risk factor selection through ranking methodologies (6). In this paper, we used regularization models (7) to improve feature selection, and then applied generalized linear mixed-effects model (GLMM) (8–10) tree to elaborate a conceptual model of two different outcomes (endotracheal intubation with IMV and mortality) and compare the significance of variables between the two models. This is a novel methodology, leveraging modern machine learning techniques to provide rigorous and applicable insight into relevant clinical questions when randomized clinical trials are not feasible.

b. Patients and methods

i. Selection and Description of Patients

In our retrospective observational study, we have collected and curated data from our electronic medical records (EMR). We selected patients admitted to our ICU at San Carlos Hospital (HCSC) in Madrid (Figure 1), but were initially not mechanically ventilated.

The database comprises hourly data points for each patient during the first five days. We then applied multi-stage ML algorithms to evaluate which variables turned out to be more relevant in predicting IMV and mortality (Figure 2).

All data were registered in our electronic medical record (ICCA Philips). A total of 12,163 longitudinal sets of hourly clinical and lab data were gathered. Longitudinal sets are grouped in clustered events associated with patients. Each entry contains demographics data, first or second wave admission, time elapsed from start of symptoms to O2 therapy and ICU

admission, APACHE II score, monitoring, blood gases and therapy-related data. We discarded variables with more than 33% of missing values for consistency. We used mode imputation or mean imputation to complete missing values of the remaining variables. Table I and Table II show the predictors that were finally used for the purposes of the study.

Data were anonymized, excluding demographic or temporal information. The study protocol was approved by the local ethics committee (approval code 22/007-E), who waived the need for informed consent due to the retrospective non-interventional nature of the study.

Include Figure 2

ii. Methods and Techniques

Data collected as described above were used to fit the model (11) following four steps for the whole process, as shown in Figure 2. Considering that our data involve a concatenation of longitudinal data for each patient in different events, it was necessary to identify correlations within the cluster when trying to build an accurate prediction model (9).

The different regression approaches to select potential predictors for IMV and mortality risk tested were: LASSO (12), Ridge (13), Elastic-net (14), Boruta (15) and R-Part (16). LASSO, Ridge and Elastic-net perform an automatic predictor selection supported by L1 and L2 regularization terms (17) that minimizes the risk of overfitting, reducing variance and reaching an attenuation effect over the correlation between features. Boruta(18) is a feature selection model based on a Random Forest algorithm that selects all the risk predictors that are relevant for classification purposes defined as *all-relevant problems*. R-Part (16) builds a classification model based on binary trees. R-Part *varImp function* (19) identifies the effect of model predictors based on the loss function mean squared error. In any case, potential predictors have been analyzed and confirmed or rejected based on clinical criteria.

After identifying the optimal set of potential predictors (Figure 10-14 in supplementary material), clustering effects by patient and temporal distribution, as well as cutoff points of the significant variables and their interactions were assessed with GLMM Trees (8–10). To build these trees, we took the entire dataset into account, grouping data by patient and data charting time as random variables to fit the model (11). This fitting methodology avoids both over and underfitting effects that could impact the model's performance (20). Models were implemented based on a 10-fold cross validation strategy using a four-depth-of-layers (full, 5, 10 and 20) strategy. This means the fitting procedure was executed ten times per algorithm implementation.

Once the potential predictors were identified, we used a GLMM Tree to build conceptual models which most likely explain the two outcome variables and their association. This tree-based algorithm considers not only the different clusters of data but also the temporal characteristics of the dataset. Considering this premise, the model assumes a mixed-effect strategy combining the potential predictors that influence a particular outcome and the relevance of the different potential predictors. It also offers a cut-off value of the variables, a very interesting feature allowing for comparison of the model with subjective evaluation of clinical experience.

GLMM Tree performance metrics were Area Under the Curve of Sensibility-Specificity (AUC), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (21), as well as the deviance, the likelihood statistical (22), and the sensitivity and specificity

parameters. All the regression and GLMM Tree models were fitted with the same subset of variables shown in Table I.

The use of both regressions and GLMM family trees was chosen because it offers a wider vision of the behavior of the potential predictors of IMV and mortality. In fact, the combination allows decision-making with a more explainable and intuitive approach than other ML or black-box modeling strategies. This approach permits us to determine which potential predictors are confirmed or rejected based on their effectiveness in predicting the outcome. We used the same set of variables (Table II) to build a model for mortality, also using the whole cohort. The scripts used and the anonymized database can be accessed in the GitHub repository associated with the study (23).

c. Results

i. Patient characteristics

A total of 154 patients (55 %) required IMV after ICU admission (Figure 1), 65 of 80 patients (81.2 %) during the first and 89 of 200 patients (44.5 %) during the second wave. ICU mortality of the whole cohort was 25.7% (72 of 280 patients), 33.7% (27 of 80 patients) during the first and 22.5% (45 of 200 patients) in the second wave. Table II shows IMV and mortality predictors for the whole patient's cohort. Mean registers per patient was 43.4, for a total of 12,163 hourly registers in the whole database (Figure 12 in complementary material).

Include Table I in horizontal Position

ii. Significance of predictors

R-Part classification achieves the best and most clinically plausible results in selecting the twelve most representative predictors for IMV and mortality from the whole group of available potential predictors (Table II). Concerning this subset of predictors, the final selection is based on decreasing order of importance, according to results reached by the loss function (mean squared error), scaled from 0 to 100 points. Taking into account this premise, the predictors are: days from first symptoms to ICU admission (100), the APACHE II score (92.25), the ROX index (72.46), blood procalcitonin (69.59), serum lactic dehydrogenase (54.45), total serum bilirubin (36.54), the COVID-19 wave (31.18), the dose of corticosteroids administered during the first five days of admission (30.96), lymphocyte count (15.57), pH (13.29), BMI (12.76), C-reactive protein (12.74), time to oxygen therapy (12.42) and body temperature (10.82).

Include Table II in horizontal Position

iii. Modeling Performance

In Table III, the performance of the IMV model is presented. The R-part predictors Regression-GLMTREE pair achieved the highest performance with an AUROC of 0.87, as shown in Figure 8 in the supplementary material. Additionally, the mortality model performed well, with an AUROC of 0.88, as demonstrated in Figure 9 in the supplementary material. Figure 3 illustrates the Mortality decision tree, while Figure 7 in the supplementary material presents the IMV

decision tree. The optimal cut-off point for the prediction model was determined based on the IMV and mortality AUC, using Youden's Index (24), which identifies the point of maximum sum of sensitivity and specificity in ROC curve analysis.

Include Figure 3 in horizontal Position

The trees in Figures 6 and 7 of the supplementary material indicate that oxygenation status (ROX index) has the most significant influence on IMV, with a threshold near 5.2. On the other hand, mortality is mainly influenced by comorbidities (APACHE II score) and LDH, as revealed by the same trees.

Include Table III in horizontal Position

d. Discussion

The present study applied a novel methodology (logistic regression with regularization plus GLMM Tree mixed models) to evaluate the relative importance of several variables as predictors of significant clinical events. Using machine learning and a fine-grained longitudinal multifaceted database, we have established relevant variable value thresholds to support clinical decisions.

Although the model would perform quite well as predictor for IMV and mortality, with good positive predictive values, it is important to emphasize that this is not a predictive model in the classical sense, but an attempt to pinpoint the most important clinical events that represent turning points during the studied process (in this case, clinical management of patients not initially under IMV).

The results of the present study include some highly relevant clinical results. First, the variable sets predicting IMV, and mortality are different. Whereas oxygenation variables are independent predictors of IMV, mortality is associated with increased age and LDH and the presence of comorbidities. The latter variables may be considered markers of two processes: COVID-19-associated inflammation and ICU-acquired superinfection (see Figure 4 in the supplementary material). Secondly, the characteristics of pharmacological therapy, including the administration of steroid drugs, y has little influence on both the need for IMV and mortality. This is a remarkable finding, because the effect of steroids on mortality identified in a previous trial(25) have influenced recommendations, as well as clinical practice, since its publication. It may be speculated that the decision to include and randomize or not at the discretion of the attending physicians, and based on undisclosed criteria, rendered different results by selecting a study subset of COVID-19 cases with different characteristics. In comparison, no inclusion-exclusion criteria or selection process were applied in our “pragmatic” type of cohort. Steroids were given to almost every patient unless a severe contraindication existed, after the results of the RECOVERY trial were made available.

This study demonstrates that predictor-ranking methodologies using self-explainable machine learning may support therapeutic decision-making using observational data, when randomized clinical trials are unfeasible or unethical.

One of the strengths of our study is the quantity and quality of the data set. Collected data have a high level of detail, leveraging the power of strategically devised electronic health records (EHR), which include relevant information in a highly structured and recoverable format. Every effort was made to configure our EHR to optimally gather all relevant information about COVID-19 patients. Also, our anonymized database is available in the repository along with the script we used for statistical analysis, is highly detailed and has been extensively curated to reflect temporal evolution and to improve data quality as much as possible.

The limitations of our study results relate mainly to its single-centered nature and require confirmation in a multicenter dataset to gain external validity. Our methodology would be perfectly suited for a multicenter study, including "center" as a random factor in the second (GLMM Tree) part of the process.

We suggest that future research applying this methodology could focus on designing clinical studies using observational data to answer relevant clinical questions without the logistic requirements of a randomized clinical trial or for hypothesis-generating purposes.

In conclusion, different variables predict IMV and mortality in severe COVID-19 patients, suggesting that the therapeutic decision of when to use IMV has little impact on mortality. Our methodology is a valid option to assess therapeutic decisions using observational data when randomized clinical trials are not feasible or ethical.

C. Contribution of the Authors

SM, MA and AN conceived the presented idea. SM and MA developed the theory and performed the computations. AN conducted an independent literature search to identify potentially relevant studies. MS independently reviewed the search results to identify pertinent articles. MS, AN, TF and VY contributed to the interpretation of the results. SM, MA, AN, MS, FL and AC took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

d. Ethics Disclosures

- Authorship of the paper: Authorship has been limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study.
- Originality and plagiarism: The authors ensure that they have written entirely original works. Work and/or words of others, have been appropriately cited or quoted.
- Disclosure and conflicts of interest: Submission includes disclosure of all relationships that could be viewed as presenting a potential risk.
- Use of patient images or case details: All data related to patients have been anonymized. The studies have passe ethics committee approval. Nº. C.I. 22/007-E Ethical Committee Hospital Clínico San Carlos, Madrid, Spain.

e. Conflict of interest

The authors declare that they have no conflict of interest.

Title

Evaluation of potential predictors for mechanical ventilation and mortality in covid-19 patients using self-explainable machine learning methods

Título

Evaluación de potenciales predictores para ventilación mecánica y mortalidad en pacientes con COVID19 utilizando métodos de aprendizaje automático auto explicativos.

Nombres y filiaciones de los autores

Sergio Muñoz Lezcano[a],
Miguel Ángel Armengol de la Hoz[b]
Alberto Corbi[c]
Fernando López[d]
Miguel Sánchez García[e]
Antonio Nuñez Reiz[f]
Tomás Fariña González[g]
Viktor Yordanov Zlatkov[h]

[a] PhD student of the Program in Computer Science, Universidad Internacional de La Rioja (UNIR), Avenida de La Paz, 137, 26006, Logroño (La Rioja) Spain. ORCID ID 0000-0002-2561-5226. smunozle@gmail.com.

[b] Big Data Department, PMC-FPS, Consejería de Salud y Consumo, Junta de Andalucía. mangel.armengol@juntadeandalucia.es

[c] Research Institute for Innovation & Technology in Education (iTED), Universidad Internacional de La Rioja (UNIR), Avenida de La Paz, 137, 26006, Logroño (La Rioja) Spain. alberto.corbi@unir.net.

[d] Mathematical Analysis and Applied Mathematics Department. Faculty of Mathematics. Universidad Complutense de Madrid. fclh@ucm.es

[e] Critical Care Department. Hospital Clínico San Carlos. Martín Lagos s/n. 28040 Madrid. Spain. miguel.sanchez@salud.madrid.org

[f] Critical Care Department. Hospital Universitario Clínico San Carlos. Martín Lagos s/n. 28040 Madrid. Spain. antoniodamaso.nunez@salud.madrid.org. ORCID ID 0000-0002-3364-5708

[g] Critical Care Department. Hospital Universitario Infanta Sofía. ORCID ID 0000-0002-4892-8153. tomas_scout@hotmail.com.

[h] Critical Care Department. Hospital Clínico San Carlos. Martín Lagos s/n.

28040 Madrid. Spain. viktoryz1111@hotmail.com.

Autor de correspondencia

Sergio Muñoz Lezcano (smunozle@gmail.com.)

Dirección actual o permanente

C/Oslo, 1, 4B. 28922, Alcorcón, Madrid, España.

Recuento de palabras

- Resumen en español: 234
- Resumen en inglés: 242
- Texto principal (sin incluir la primera página, el resumen/asbtract, la bibliografía, las tablas y las figuras): 2.108

Evaluation of potential predictors for mechanical ventilation and mortality in covid-19 patients using self-explainable machine learning methods

1.

a. Abstract:

i. Objective

To determine potential predictors for invasive mechanical ventilation (IMV) and mortality in COVID-19-associated acute respiratory distress syndrome (C-ARDS) using machine learning (ML) techniques.

ii. Design

Single center highly detailed longitudinal observational study.

iii. Setting

Tertiary hospital ICU, during the two first COVID19 pandemic waves in Madrid, Spain.

iv. Patients

280 patients with C-ARDS, not requiring IMV on admission.

v. Interventions

Retrospective and observational study of anonymized time series records of patient clusters up to the 5th day of ICU stay. The researchers applied a dual-step machine learning methodology to identify independent variables associated with IMV and mortality.
The study involved a retrospective analysis of anonymized time series records of patient clusters up to the 5th day of ICU stay.

vi. Main variables

Target: endotracheal intubation and IMV, mortality.
Predictors: demographics, hourly evolution of oxygenation, clinical data, and laboratory results.

vii. Results

The time between symptom onset and ICU admission, the APACHE II score, the ROX index, and procalcitonin levels in blood were potential predictors related to both IMV and mortality. The ROX index was the most significant predictor associated with IMV, while APACHE II, LDH, and DaysSympICU were the most with mortality.

viii. Conclusions

R-part classification identify significant predictors linked with IMV and mortality in C-ARDS patients, including the time between symptom onset and

ICU admission, the severity of the COVID-19 waves, and several clinical and laboratory measures. These findings may help clinicians to better identify patients at risk for IMV and mortality and improve their management.

b. Resumen:

i. Objetivo

Determinar potenciales predictores para ventilación mecánica invasiva (VMI) y mortalidad en el síndrome de distrés respiratorio agudo asociado con COVID-19 (C-SDRA) utilizando técnicas de aprendizaje automático (AA).

ii. Diseño

Estudio observacional longitudinal en un solo centro.

iii. Ámbito

UCI de un hospital terciario durante las dos primeras olas de COVID-19 en Madrid, España.

iv. Pacientes

280 pacientes con C-SDRA que no requieren VMI al ingreso en UCI.

v. Intervenciones

Estudio retrospectivo observacional sobre datos temporales en clústers de pacientes con ingreso en UCI hasta el quinto día de estancia, identificando variables independientes y sus umbrales asociados con VMI y mortalidad.

vi. Principales variables

Objetivo: VMI y Mortalidad.
Predictores: demográficos, variables clínicas, resultados de laboratorio y evolución de la oxigenación.

vii. Resultados

El tiempo entre el inicio de los síntomas y el ingreso en la UCI, la puntuación APACHE II, el índice ROX y los niveles de procalcitonina en sangre eran posibles predictores relacionados tanto con la IMV como con la mortalidad. El índice ROX fue el predictor más significativo asociada con la IMV, mientras que APACHE II, LDH y DaysSympICU fueron los más influyentes en la mortalidad.

viii. Conclusiones

El clasificador R-part identifica predictores significativos vinculados con la VMI y mortalidad en pacientes con C-ARDS, incluido el tiempo entre el inicio de los síntomas y el ingreso en la UCI, la gravedad de las olas de COVID-19 y varias medidas clínicas y de laboratorio. Estos hallazgos pueden ayudar a los médicos a identificar mejor a los pacientes en riesgo de IMV y mortalidad y mejorar su manejo.

Keywords/Palabras Clave: Acute Respiratory Distress Syndrome, Invasive Mechanical Ventilation, COVID-19, Machine Learning, Artificial Intelligence, Predictors.

b. Text

a. Introduction

Invasive mechanical ventilation (IMV) is a cornerstone of organ support in severe COVID-19 patients with acute respiratory distress syndrome (ARDS). As widely experienced in ICUs during the SARS-CoV-2 pandemic, IMV frequently causes complications (1,2) . Hospital services were overwhelmed not only by the surge of patients, but also by scarce human resources and equipment, lack of sufficient mechanical ventilators being probably the most relevant. In surge scenarios, appropriate triage strategies are therefore needed to allocate IMV or alternatives such as high flow nasal prongs. These strategies should be based on the knowledge and understanding of specific potential predictors that could help clinicians to personalize decisions regarding IMV.

There is still considerable controversy regarding who and when to intubate. Several recent studies have addressed the subject (3), although bias cannot be excluded in observational non-randomized trials. A retrospective study suggested that early intubation and IMV is associated with favorable outcomes but included only intubated patients instead of the whole population at risk.

Previous studies have identified covid-19 progression predictors including age, comorbidities, renal function, or immunodeficiency (4) using traditional statistical approaches, where collinearity of data cannot be ruled out. Artificial intelligence (AI) is currently being used for COVID-19 risk stratification (5), studying multiple clinical features to increase effectiveness and efficiency in diagnosis, treatment, and prognosis. Self-explainable Machine learning (ML) techniques can help with risk factor selection through ranking methodologies (6). In this paper, we used regularization models (7) to improve feature selection, and then applied generalized linear mixed-effects model (GLMM) (8–10) tree to elaborate a conceptual model of two different outcomes (endotracheal intubation with IMV and mortality) and compare the significance of variables between the two models. This is a novel methodology, leveraging modern machine learning techniques to provide rigorous and applicable insight into relevant clinical questions when randomized clinical trials are not feasible.

b. Patients and methods

i. Selection and Description of Patients

In our retrospective observational study, we have collected and curated data from our electronic medical records (EMR). We selected patients admitted to our ICU at San Carlos Hospital (HCSC) in Madrid (Figure 1), but were initially not mechanically ventilated.

The database comprises hourly data points for each patient during the first five days. We then applied multi-stage ML algorithms to evaluate which variables turned out to be more relevant in predicting IMV and mortality (Figure 2).

All data were registered in our electronic medical record (ICCA Philips). A total of 12,163 longitudinal sets of hourly clinical and lab data were gathered. Longitudinal sets are grouped in clustered events associated with patients. Each entry contains demographics data, first or second wave admission, time elapsed from start of symptoms to O2 therapy and ICU

admission, APACHE II score, monitoring, blood gases and therapy-related data (Table IV in supplementary material) . We discarded variables with more than 33% of missing values for consistency. We used mode imputation or mean imputation to complete missing values of the remaining variables. Table I and Table II show the predictors that were finally used for the purposes of the study.

Data were anonymized, excluding demographic or temporal information. The study protocol was approved by the local ethics committee (approval code 22/007-E), who waived the need for informed consent due to the retrospective non-interventional nature of the study.

Include Figure 2

ii. Methods and Techniques

Data collected as described above were used to fit the model (11) following four steps for the whole process, as shown in Figure 2. Considering that our data involve a concatenation of longitudinal data for each patient in different events, it was necessary to identify correlations within the cluster when trying to build an accurate prediction model (9).

The different regression approaches to select potential predictors for IMV and mortality risk tested were: LASSO (12), Ridge (13), Elastic-net (14), Boruta (15) and R-Part (16). LASSO, Ridge, and Elastic-net perform an automatic predictor selection supported by L1 and L2 regularization terms (17) that minimizes the risk of overfitting, reducing variance and reaching an attenuation effect over the correlation between features. Boruta(18) is a feature selection model based on a Random Forest algorithm that selects all the risk predictors that are relevant for classification purposes defined as *all-relevant problems*. R-Part (16) builds a classification model based on binary trees. R-Part *varImp function* (19) identifies the effect of model predictors based on the loss function mean squared error. In any case, potential predictors have been analyzed and confirmed or rejected based on clinical criteria.

After identifying the optimal set of potential predictors (Figure 10-14 in supplementary material), clustering effects by patient and temporal distribution, as well as cutoff points of the significant variables and their interactions were assessed with GLMM Trees (8–10). To build these trees, we took the entire dataset into account, grouping data by patient and data charting time as random variables to fit the model (11). This fitting methodology avoids both over and underfitting effects that could impact the model's performance (20). Models were implemented based on a 10-fold cross validation strategy using a four-depth-of-layers (full, 5, 10 and 20) strategy. This means the fitting procedure was executed ten times per algorithm implementation.

Once the potential predictors were identified, we used a GLMM Tree to build conceptual models which most likely explain the two outcome variables and their association. This tree-based algorithm considers not only the different clusters of data but also the temporal characteristics of the dataset. Considering this premise, the model assumes a mixed-effect strategy combining the potential predictors that influence a particular outcome and the relevance of the different potential predictors. It also offers a cut-off value of the variables, a very interesting feature allowing for comparison of the model with subjective evaluation of clinical experience.

GLMM Tree performance metrics were Area Under the Curve of Sensibility-Specificity (AUC), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (21), as well as the deviance, the likelihood statistical (22), and the sensitivity and specificity

parameters. All the regression and GLMM Tree models were fitted with the same subset of variables shown in Table I.

The use of both regressions and GLMM family trees was chosen because it offers a wider vision of the behavior of the potential predictors of IMV and mortality. In fact, the combination allows decision-making with a more explainable and intuitive approach than other ML or black-box modeling strategies. This approach permits us to determine which potential predictors are confirmed or rejected based on their effectiveness in predicting the outcome. We used the same set of variables (Table II) to build a model for mortality, also using the whole cohort. The scripts used can be accessed in the GitHub repository associated with the study (23). The database will be published in PhysioNet (24) project in order to disseminate and exchange the anonymized clinical records looking for cooperative project replication.

c. Results

i. Patient characteristics

A total of 154 patients (55 %) required IMV after ICU admission (Figure 1), 65 of 80 patients (81.2 %) during the first and 89 of 200 patients (44.5 %) during the second wave. ICU mortality of the whole cohort was 25.7% (72 of 280 patients), 33.7% (27 of 80 patients) during the first and 22.5% (45 of 200 patients) in the second wave. Table II shows IMV and mortality predictors for the whole patient's cohort. Mean registers per patient was 43.4, for a total of 12,163 hourly registers in the whole database (Figure 12 in complementary material).

Include Table I in horizontal Position

ii. Significance of predictors

R-Part classification achieves the best and clinically most plausible results in selecting the twelve most representative predictors for IMV and mortality from the whole group of available potential predictors (Table II). Concerning this subset of predictors, the final selection is based on decreasing order of importance, according to results reached by the loss function (mean squared error), scaled from 0 to 100 points. Taking into account this premise, the predictors are: days from first symptoms to ICU admission (100), the APACHE II score (92.25), the ROX index (72.46), blood procalcitonin (69.59), serum lactic dehydrogenase (54.45), total serum bilirubin (36.54), the COVID-19 wave (31.18), the dose of corticosteroids administered during the first five days of admission (30.96), lymphocyte count (15.57), pH (13.29), BMI (12.76), C-reactive protein (12.74), time to oxygen therapy (12.42) and body temperature (10.82).

Include Table II in horizontal Position

iii. Modeling Performance

In Table III, the performance of the IMV model is presented. The R-part predictors Regression-GLMTREE pair achieved the highest performance with an AUROC of 0.87, as shown in Figure 8 in the supplementary material. Additionally, the mortality model performed well, with an AUROC of 0.88, as demonstrated in Figure 9 in the supplementary material. Figure 3 illustrates

the Mortality decision tree, while Figure 7 in the supplementary material presents the IMV decision tree. The optimal cut-off point for the prediction model was determined based on the IMV and mortality AUC, using Youden's Index (25), which identifies the point of maximum sum of sensitivity and specificity in ROC curve analysis.

Include Figure 3 in horizontal Position

The trees in Figures 6 and 7 of the supplementary material indicate that oxygenation status (ROX index) has the most significant influence on IMV, with a threshold near 5.2. On the other hand, mortality is mainly influenced by comorbidities (APACHE II score) and LDH, as revealed by the same trees.

Include Table III in horizontal Position

d. Discussion

The present study applied a novel methodology (logistic regression with regularization plus GLMM Tree mixed models) to evaluate the relative importance of several variables as predictors of significant clinical events. Using machine learning and a fine-grained longitudinal multifaceted database, we have established relevant variable value thresholds to support clinical decisions.

Although the model would perform quite well as predictor for IMV and mortality, with good positive predictive values, it is important to emphasize that this is not a predictive model in the classical sense, but an attempt to pinpoint the most important clinical events that represent turning points during the studied process (in this case, clinical management of patients not initially under IMV).

The results of the present study include some highly relevant clinical results. First, the variable sets predicting IMV, and mortality are different. Whereas oxygenation variables are independent predictors of IMV, mortality is associated with increased age and LDH and the presence of comorbidities. The latter variables may be considered markers of two processes: COVID-19-associated inflammation and ICU-acquired superinfection (see Figure 4 in the supplementary material). Secondly, the characteristics of pharmacological therapy, including the administration of steroid drugs, y has little influence on both the need for IMV and mortality. This is a remarkable finding, because the effect of steroids on mortality identified in a previous trial (26) have influenced recommendations, as well as clinical practice, since its publication. It may be speculated that the decision to include and randomize or not at the discretion of the attending physicians, and based on undisclosed criteria, rendered different results by selecting a study subset of COVID-19 cases with different characteristics. In comparison, no inclusion-exclusion criteria or selection process were applied in our “pragmatic” type of cohort. Steroids were given to almost every patient unless a severe contraindication existed, after the results of the RECOVERY trial were made available.

This study demonstrates that predictor-ranking methodologies using self-explainable machine learning may support therapeutic decision-making using observational data, when randomized clinical trials are unfeasible or unethical.

One of the strengths of our study is the quantity and quality of the data set. Collected data have a high level of detail, leveraging the power of strategically devised electronic health records (EHR), which include relevant information in a highly structured and recoverable format. Every effort was made to configure our EHR to optimally gather all relevant information about COVID-19 patients. Also, our anonymized database is available in the repository along with the script we used for statistical analysis, is highly detailed and has been extensively curated to reflect temporal evolution and to improve data quality as much as possible.

The limitations of our study results relate mainly to its single-centered nature and require confirmation in a multicenter dataset to gain external validity. Our methodology would be perfectly suited for a multicenter study, including "center" as a random factor in the second (GLMM Tree) part of the process.

We suggest that future research applying this methodology could focus on designing clinical studies using observational data to answer relevant clinical questions without the logistic requirements of a randomized clinical trial or for hypothesis-generating purposes.

In conclusion, different variables predict IMV and mortality in severe COVID-19 patients, suggesting that the therapeutic decision of when to use IMV has little impact on mortality. Our methodology is a valid option to assess therapeutic decisions using observational data when randomized clinical trials are not feasible or ethical.

C. Contribution of the Authors

SM, MA and AN conceived the presented idea. SM and MA developed the theory and performed the computations. AN conducted an independent literature search to identify potentially relevant studies. MS independently reviewed the search results to identify pertinent articles. MS, AN, TF and VY contributed to the interpretation of the results. SM, MA, AN, MS, FL and AC took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

d. Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

e. Conflict of interest

The authors declare that they have no conflict of interest.

f. Acknowledgements

g. References

1. Rajdev K, Spanel A, McMillan S, Lahan S, Boer B, Birge J, et al. Pulmonary barotrauma in COVID-19 patients with ARDS on invasive and non-invasive positive pressure ventilation. Journal of Intensive Care Medicine.2021,36(9), 1013-1017. https://doi.org/10.1177/08850666211019719
2. Estella Á, Vidal-Cortés P, Rodríguez A, Andaluz Ojeda D, Martín-Loeches I, Díaz E, et al. Management of infectious complications associated with coronavirus infection in severe patients admitted to ICU. Med Intensiva Engl Ed. 2021 Nov 1;45(8):485–500. https://.doi.org/10.1016/j.medine.2021.08.013
3. Papoutsi E, Giannakoulis V, Xourgia E, Routsi C, Kotanidou A, Siempos I. Effect of timing of intubation on clinical outcomes of critically ill patients with COVID-19: a systematic review and meta-analysis of non-randomized cohort studies. Crit Care. 2021 Dec;25(1):121. https://doi.org/10.1186/s13054-021-03540-6
4. Chomistek A, Liang C, Doherty M, Clifford C, Ogilvie R, Gately R, et al. Predictors of critical care, mechanical ventilation, and mortality among hospitalized patients with COVID-19 in an electronic health record database. BMC Infect Dis. 2022 Dec;22(1):413. https://doi.org/10.1186/s12879-022-07383-6
5. Aldhoayan M. The Role of Artificial Intelligence and Machine Learning During the Covid-19 Pandemic: A Review. Stud Health Technol Inform. 2022 Jun 29;295:28–32. https://doi.org/10.3233/SHTI220651
6. Rajbahadur G, Wang S, Oliva G, Kamei Y, Hassan A. The Impact of Feature Importance Methods on the Interpretation of Defect Classifiers. IEEE Trans Softw Eng. 2022 Jul;48(7):2245–61. https://doi.org/10.48550/arXiv.2202.02389
7. Ogutu J, Schulz-Streeck T, Piepho HP. Genomic selection using regularized linear regression models: ridge regression, LASSO, elastic net and their extensions. BMC Proc. 2012 May 21;6(2): S10. https://doi.org/10.1186/1753-6561-6-S2-S10
8. Poddar M, Harigovind G. Mixed-effects Model for Classification and Prediction in Longitudinal Data Analysis. In: 2018 International Conference on Bioinformatics and Systems Biology (BSB). 2018. p. 36–9. http://doi.org/10.1109/BSB.2018.8770598
9. Fokkema M, Smits N, Zeileis A, Hothorn T, Kelderman H. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. Behav Res Methods. 2018 Oct;50(5):2016–34. https://doi.org/10.3758/s13428-017-0971-x
10. Seibold H, Hothorn T, Zeileis A. Generalised linear model trees with global additive effects. Adv Data Anal Classif. 2019 Sep 1;13(3):703–25. https://doi.org/10.1007/s11634-018-0342-1
11. Lever J, Krzywinski M, Altman N. Points of Significance: Model selection and overfitting. Nat Methods. 2016 Sep 1;13(9):703–5. https://doi.org/10.1038/nmeth.3968
12. Tibshirani R. Regression Shrinkage and Selection Via the LASSO Journal of the Royal Statistical Society. Series B (Methodological). 1996, 58(1), 267–288. http://www.jstor.org/stable/2346178. Accessed 19 Feb. 2023.
13. Hoerl A, Kennard R. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 1970 Feb 1;12(1):55–67. https://doi.org/10.1080/00401706.1970.10488634
14. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B (Statistical Methodology). 2005;67(2):301–20. JSTOR, http://www.jstor.org/stable/3647580 . Accessed 19 Feb. 2023.
15. Kursa M, Jankowski A, Rudnicki W. Boruta – A System for Feature Selection.

Fundamenta Informaticae. 2010 Jan 1;101(4):271–85. https://doi.org/10.3233/FI-2010-288
16. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods. 2009;14(4):323–48. https://doi.org/10.1037/a0016973
17. Ng A. Feature selection, L1 vs. L2 regularization, and rotational invariance. In: Proceedings of the twenty-first international conference on Machine learning [Internet]. New York, NY, USA: Association for Computing Machinery; 2004 [cited 2022 Aug 19]. p. 78. (ICML '04). https://doi.org/10.1145/1015330.1015435
18. Kursa M, Rudnicki W. Feature Selection with the Boruta Package. Journal of Statistical Software. 2010 Sep 16; 36(11),1–13. https://doi.org/10.18637/jss.v036.i11
19. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008 Nov 10; 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05
20. Deliberato R, Ko S, Komorowski M, Armengol de La Hoz M, Frushicheva M, Raffa J, et al. Severity of Illness Scores May Misclassify Critically Ill Obese Patients. Crit Care Med. 2018 Mar;46(3):394–400. https://doi.org/10.1097/CCM.0000000000002868
21. Vrieze S. Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol Methods. 2012;17(2):228–43. https://doi.org/10.1037/a0027127
22. Chatterjee S, Frohner N, Lechner L, Schöfbeck R, Schwarz D. Tree boosting for learning EFT parameters. Comput Phys Commun. 2022 Aug 1; 277:108385. https://doi.org/10.48550/arXiv.2107.10859
23. Muñoz Lezcano S. Git Hub Code Repository for covid-19 project [Internet]. GitHub. [cited 2023 Jan 4]. Available from: https://github.com/SergioML73/covid-19SC
24. Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J et all. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. 2000; Circulation 101(23):e215-e220. https://doi.org/10.1161/01.CIR.101.23.e215
25. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its Associated Cutoff Point. Biometrical Journal. 2005;47(4):458–72. https://doi.org/10.1002/bimj.200410135
26. RECOVERY Collaborative Group. Dexamethasone in Hospitalized Patients with Covid-19. New England Journal of Medicine.2021. 384(8), 693-704. https://doi.org/10.1056/NEJMoa2021436

h. Tables

Table I. Patient characteristics

Mechanical Ventilation

Variable	N	Overall, N = 12,163[1]	Invasive Mechanical Ventilation		p-value[2]
			No, N = 9,093[1]	Yes, N = 3,070[1]	
Age, years, median (IQR)	12,163	59 (51 – 68)	58 (50 – 67)	63 (54 – 70)	<0.001
Gender, n (%)	12,163				<0.001
Male		8,032 (66)	5,649 (62)	2,383 (78)	
Female		4,131 (34)	3,444 (38)	687 (22)	
Ethnicity, n (%)	12,163				<0.001
Amerindian		4,671 (38)	3,625 (40)	1,046 (34)	
Arab		545 (4.5)	468 (5.1)	77 (2.5)	
Spanish		6,427 (53)	4,591 (50)	1,836 (60)	
Others		520 (4.3)	409 (4.5)	111 (3.6)	
Wave, n (%)	12,163				<0.001
First		1,490 (12)	766 (8.4)	724 (24)	
Second		10,673 (88)	8,327 (92)	2,346 (76)	
Body mass index, Median (IQR)	12,163	27.8 (26.0 – 31.1)	27.8 (26.0 – 31.2)	27.7 (26.0 – 29.4)	0.70
Heart rate, median bpm (IQR)	12,163	73 (65 – 84)	73 (64 – 83)	76 (67 – 87)	<0.001
Temperature in ºC, Median (IQR)	12,163	36.80 (36.50 – 37.10)	36.73 (36.44 – 37.02)	36.97 (36.64 – 37.37)	<0.001
Arterial pressure in mmHg, Median (IQR)	12,163	87 (79 – 95)	87 (80 – 95)	87 (78 – 95)	<0.001
Lactate in mEq/l, Median (IQR)	12,163	1.42 (1.20 – 1.70)	1.42 (1.14 – 1.65)	1.50 (1.33 – 1.80)	<0.001
Procalcitonin, ng/mL Median (IQR)	12,163	0.13 (0.08 – 0.23)	0.13 (0.07 – 0.20)	0.14 (0.13 – 0.35)	<0.001
Eosinophile count per cubic mm, Median (IQR)	12,163	4 (0 – 20)	4 (0 – 22)	4 (0 – 13)	0.011
C reactive protein, mg/LMedian (IQR)	12,163	8 (6 – 11)	8 (4 – 10)	8 (8 – 15)	<0.001
Alkaline phosphatase U/L, Median (IQR)	12,163	82 (68 – 101)	82 (65 – 99)	82 (76 – 104)	0.006
Total bilirubin mg/dL, Median (IQR)	12,163	0.53 (0.44 – 0.62)	0.53 (0.42 – 0.59)	0.53 (0.51 – 0.71)	<0.001
Oxygenation index, Median (IQR)	12,163	5.93 (4.52 – 7.92)	6.18 (5.22 – 8.67)	4.46 (3.62 – 5.93)	<0.001
Creatinine ,mg/dL Median (IQR)	12,163	0.67 (0.59 – 0.78)	0.67 (0.58 – 0.77)	0.67 (0.65 – 0.82)	<0.001

Table I. Patient characteristics

Mechanical Ventilation

Variable	N	Overall, N = 12,163[1]	Invasive Mechanical Ventilation		p-value[2]
			No, N = 9,093[1]	Yes, N = 3,070[1]	
Leukocyte count per mm3 , Median (IQR)	12,163	8,925 (7,160 – 10,804)	8,925 (6,857 – 10,548)	8,925 (8,400 – 11,478)	<0.001
Hemoglobin g/l, Median (IQR)	12,163	13.16 (12.28 – 13.96)	13.16 (12.20 – 13.93)	13.16 (12.63 – 14.03)	<0.001
Amylase U/L, Median (IQR)	12,163	63 (50 – 79)	63 (51 – 84)	63 (48 – 64)	<0.001
Lactate dehydrogenase, Median (IQR)	12,163	882 (749 – 1,038)	882 (682 – 964)	939 (882 – 1,172)	<0.001
Lymphocyte count per mm3 , Median (IQR)	12,163	829 (638 – 1,049)	829 (657 – 1,148)	829 (570 – 871)	<0.001
AST U/L, Median (IQR)	12,163	45 (34 – 60)	45 (34 – 63)	45 (33 – 54)	<0.001
Hours from ICU admission to this register, Median (IQR)	12,163	31 (14 – 50)	33 (16 – 51)	23 (9 – 45)	<0.001
APACHE II, Median (IQR)	12,163	13.0 (10.0 – 17.0)	12.0 (10.0 – 16.0)	15.0 (13.0 – 17.0)	<0.001
Days from first symptoms to O2 therapy, Median (IQR)	12,163	7.00 (6.00 – 8.00)	7.00 (6.00 – 8.00)	7.00 (6.00 – 8.00)	0.008
Days from first symptoms to ICU admission, Median (IQR)	12,163	9.0 (8.0 – 11.0)	9.0 (8.0 – 11.0)	9.0 (7.0 – 13.0)	<0.001
Arterial pH, Median (IQR)	12,163	7.43 (7.41 – 7.45)	7.43 (7.41 – 7.46)	7.43 (7.39 – 7.44)	<0.001
Arterial pCO2, Median (IQR)	12,163	38.1 (35.7 – 41.0)	38.1 (35.6 – 40.7)	38.4 (36.0 – 42.4)	<0.001
Type of blood sample, n (%)	12,163				<0.001
Arterial		696 (5.7)	496 (5.5)	200 (6.5)	
BLDO		5 (<0.1)	5 (<0.1)	0 (0)	
Arterial		91 (0.7)	27 (0.3)	64 (2.1)	
Mixed		28 (0.2)	28 (0.3)	0 (0)	
Venous		1,405 (12)	931 (10)	474 (15)	
Venous		9,845 (81)	7,529 (83)	2,316 (75)	
Mixed venous		93 (0.8)	77 (0.8)	16 (0.5)	
Blood gas sat. O2, Median (IQR)	12,163	85 (75 – 91)	85 (77 – 91)	84 (72 – 89)	<0.001
Corticosteroid dose, first 5 days of admission (mg of equivalent methylprednisolone dose), Median (IQR)	12,163	36 (30 – 60)	36 (30 – 60)	36 (30 – 78)	0.32
Melatonin dose in mg/day, n (%)	12,163				0.001
0		4,545 (37)	3,445 (38)	1,100 (36)	
50		3,886 (32)	2,922 (32)	964 (31)	
100		2,167 (18)	1,617 (18)	550 (18)	
200		1,565 (13)	1,109 (12)	456 (15)	
D dimer, ng/ml Median (IQR)	12,163	1,031 (862 – 1,263)	1,031 (804 – 1,232)	1,031 (1,031 – 1,416)	<0.001

Table I. Patient characteristics

Mechanical Ventilation

Variable	N	Overall, N = 12,163[1]	Invasive Mechanical Ventilation		p-value[2]
			No, N = 9,093[1]	Yes, N = 3,070[1]	

Data updated December 7th, 2022

[1] Median (IQR) or Frequency (%)

[2] Welch Two Sample t-test; Pearson's Chi-squared test

Table I. Group of predictors, clinical and biochemical characteristics for regression purposes. This group of predictors will be applied in the selection procedure linked with the five regression algorithms: Ridge, LASSO, Elastic, Boruta and R-part Based on the reached results, the group of predictors are going to be reduced attending to its behavior related to IMV needs. Figures 7, 8, 9, 10 and 11 shows the results from each regression procedure where R-Part was finally selected due to its good balance between model performance and explicability of results.

Table II. Patient characteristics

Mortality

Variable	N	Overall, N = 12,163[1]	Mortality		p-value[2]
			Alive, N = 9,777[1]	Died, N = 2,386[1]	
Days elapsed from first symptoms to ICU admission, Median (IQR)	12,163	9.0 (8.0 – 11.0)	9.0 (8.0 – 11.0)	9.0 (7.0 – 13.0)	<0.001
APACHE II, Median (IQR)	12,163	13.0 (10.0 – 17.0)	12.0 (10.0 – 16.0)	15.0 (13.0 – 17.0)	<0.001
Corticosteroids administered during the first 5d of admission as mg of equivalent methylprednisolone dose, Median (IQR)	12,163	36 (30 – 60)	36 (30 – 60)	36 (30 – 80)	<0.001
Oxygenation index, Median (IQR)	12,163	5.93 (4.52 – 7.92)	5.93 (4.95 – 8.42)	4.58 (3.63 – 5.93)	<0.001
Serum Lactate dehydrogenase, U/L Median (IQR)	12,163	882 (749 – 1,038)	882 (695 – 964)	1,026 (882 – 1,279)	<0.001
Body mass index, Median (IQR)	12,163	27.8 (26.0 – 31.1)	27.8 (26.0 – 31.8)	27.8 (26.0 – 29.4)	<0.001
Temperature in ºC, Median (IQR)	12,163	36.80 (36.50 – 37.10)	36.78 (36.50 – 37.10)	36.86 (36.50 – 37.20)	<0.001
Days elapsed from first symptoms to O2 therapy, Median (IQR)	12,163	7.00 (6.00 – 8.00)	7.00 (6.00 – 8.00)	7.00 (6.00 – 7.00)	0.073
Total bilirubin mg/dL, Median (IQR)	12,163	0.53 (0.44 – 0.62)	0.53 (0.42 – 0.60)	0.53 (0.51 – 0.68)	<0.001
Wave, n (%)	12,163				<0.001
First		1,490 (12)	1,031 (11)	459 (19)	
Second		10,673 (88)	8,746 (89)	1,927 (81)	

Table II. Patient characteristics

Mortality

Variable	N	Overall, N = 12,163[1]	Mortality		p-value[2]
			Alive, N = 9,777[1]	Died, N = 2,386[1]	
Lymphocyte count per mm3, Median (IQR)	12,163	829 (638 – 1,049)	829 (667 – 1,120)	800 (499 – 886)	<0.001
Arterial pH, Median (IQR)	12,163	7.43 (7.41 – 7.45)	7.43 (7.41 – 7.46)	7.43 (7.38 – 7.45)	<0.001
C reactive protein levels mg/L, Median (IQR)	12,163	8 (6 – 11)	8 (5 – 10)	8 (8 – 14)	<0.001
Hours from ICU admission to this register, Median (IQR)	12,163	31 (14 – 50)	31 (14 – 51)	28 (11 – 47)	<0.001

Data updated November 21, 2022

[1] Median (IQR) or Frequency (%)

[2] Welch Two Sample t-test; Pearson's Chi-squared test

Table II. IMV Results. Group of predictors, clinical and biochemical characteristics, used for mortality prediction with GLMM tree algorithm.

Table III. GLMM Trees Results

Mechanical Ventilation

Regressions	Nº Predictors	AUC	AIC	BIC	Deviance	Log Lik	Sensitivity(recall)	Specificificity
Ridge criteria	25	0.852	9,263.82	9,493.41	9201.82	-4,600.91	0.856	0.689
LASSO criteria	22	0.852	9,263.82	9,4939.41	9,201.82	-4.600.91	0.856	**0.852**
Elastic criteria	20	0.858	9,111.20	9,325.98	9,053.20	-4,526.60	0.750	0.816
Boruta criteria	32	**0.897**	**7,775.23**	**8,004.82**	**7,713.23**	**-3,856.61**	0.858	0.800
R-Part criteria	**13**	0.867	7,830.28	8,059.87	7,758.28	-3,884.14	**0.871**	0.725

Table III. IMV Results. The Akaike Information Criterion (AIC) reports the information score of the whole models: the smaller the ***AIC value***, the better the model fit. AIC is calculated from the number of independent variables to build the model and the maximum likelihood estimate of the model (how well the model reproduces the data). The best-fit model according to AIC is the one that explains the greatest amount of variation using the fewest possible independent variables. ***Bayesian information criterion (BIC)*** (Stone, 1979) is another criteria for model selection that measures the trade-off between model fit and complexity of the model. A lower AIC or BIC value indicates a better fit. The ***log-likelihood (log Lik)*** value of a regression model is a way to measure the goodness of fit for a model. The higher the value of the log-likelihood, the better a model fits a dataset. ***Deviance*** is a goodness-of-fit metric for statistical models, particularly used for GLMs. It is defined as the difference between the Saturated and Proposed Models and can be thought as how much variation in the data does our Proposed Model account for. Therefore, the lower the deviance, the better the model. ***Sensitivity*** is the metric that evaluates a model's ability to predict true positives of each available category. Specificity is the metric that evaluates a model's ability to predict true negatives of each available category. The higher value of sensitivity would mean higher value of true positive and lower value of false negative. For the healthcare domain, models with high sensitivity will be desired. ***Specificity*** is the metric that evaluates a model's ability to predict true negatives of each available category. These metrics apply to any categorical model. Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative). Specificity is a measure of the proportion of people not suffering from the disease who got predicted correctly as the ones who are not suffering from the disease. In other words, the person who is healthy actually got predicted as healthy.

i. Figures

Figure 1. COVID-19 patients admitted during first and second pandemic waves. The cohort comprises 280 severe COVID-19 patients admitted to the ICU Department at HCSC in Madrid, Spain, between March 3, 2020, and February 28, 2021. During this time period, SARS-COV-2 wild-type and subsequently alpha variants were prevalent in Spain. Over the study time period 4,229 covid-19 patients were admitted to HCSC, 405 of whom required ICU admission (first wave: 153, second wave: 252 patients)

Figure 2. Methodology for fitting the machine learning algorithms. In a previous stage, Figure 5 in supplementary material shows the complete workflow, from the cohort selection according to clinical needs to the implementation of the algorithms that have been included in the explanation. The first step involves the cohort selection as well as the initial group of variables considered in this study, The second step consists in the implementation of a statistical study of each variable. This step also involves correlation (Figure 6 in supplementary material) imputation and transformations procedures in order to dispose of the most accurate data in the following steps. The third step analyzed the most significant predictors based on five ML techniques linked with regression analysis. The fourth and last step identifies the behavior of each predictor attending to different proposes. The first one is related to mechanical ventilation needs attending to different settings in the GLMM Tree (depth of layers) looking for the best balance between performance and explainability of the model. The second one is related to the most representative mortality predictors but following the same balance objective.

Figure 3. Mortality Tree Predictors. The predictors appear in different branches attending to their significance in the predictive model. Values in bold letters represent the registries per branch. Values in red bold letters represent the percentage of registries with positive outcome

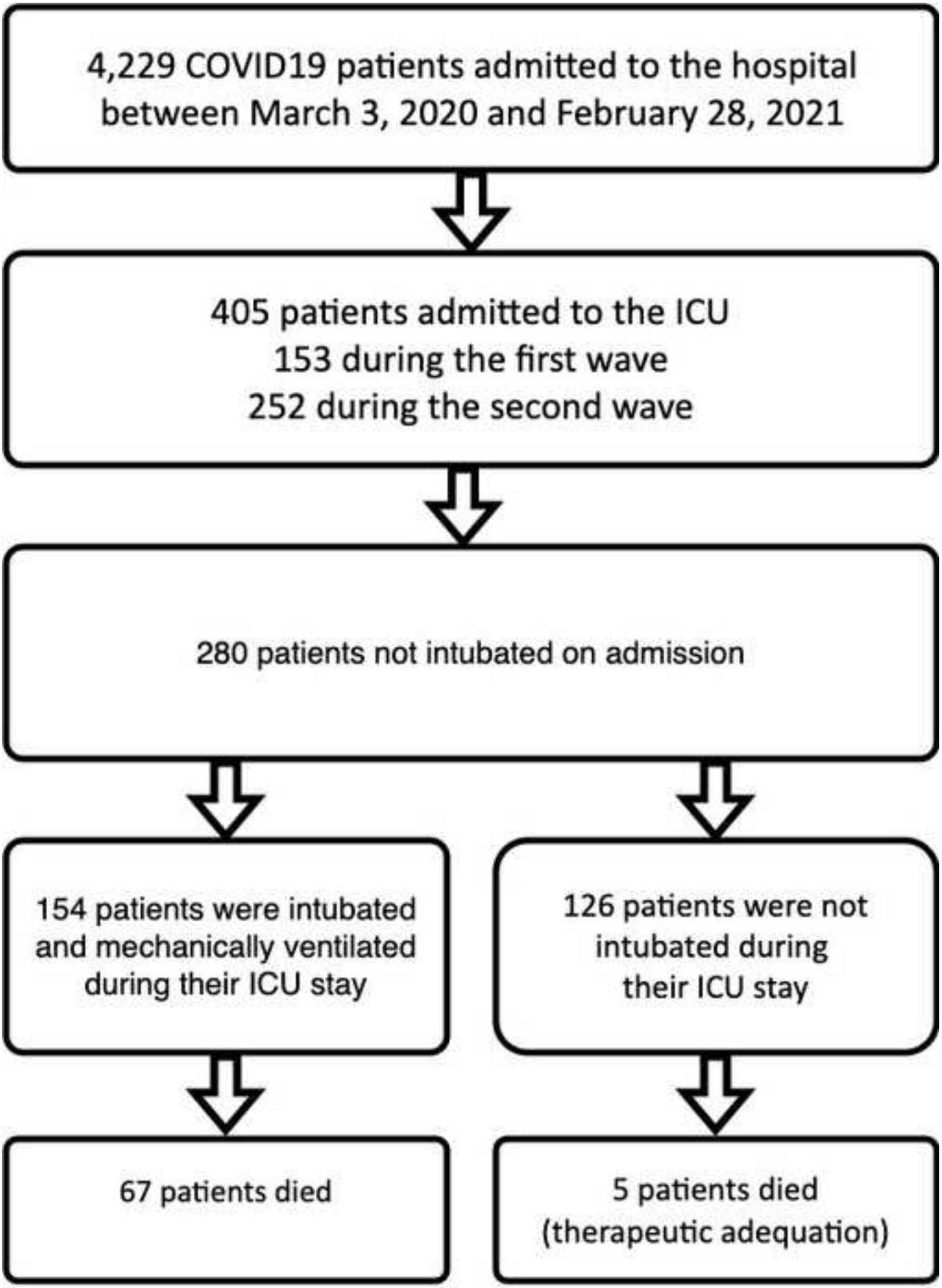


Figura / Figure

Pulse aquí para acceder/descargar;Figura / Figure;Fig2_Methodolgy.jpg

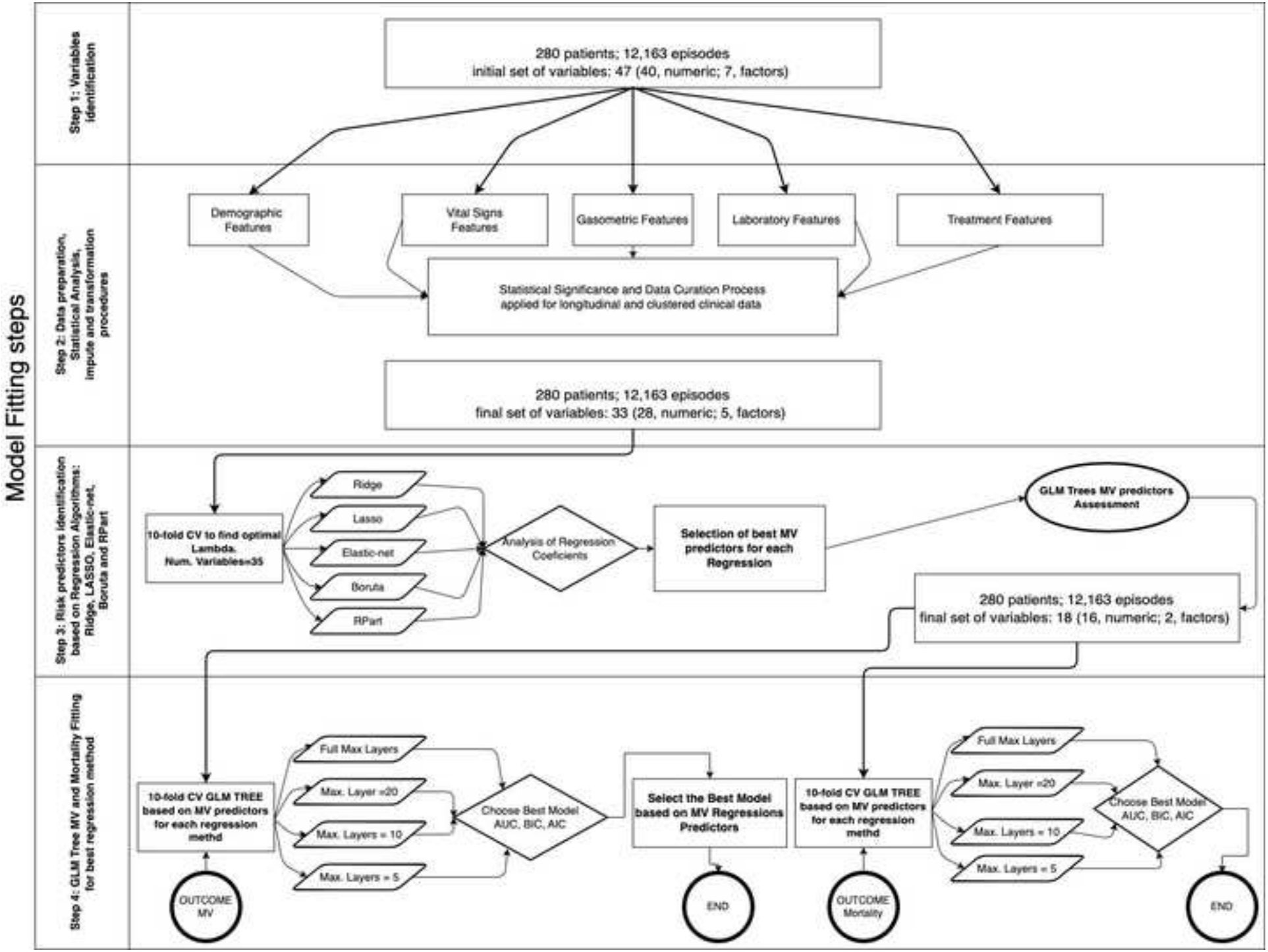
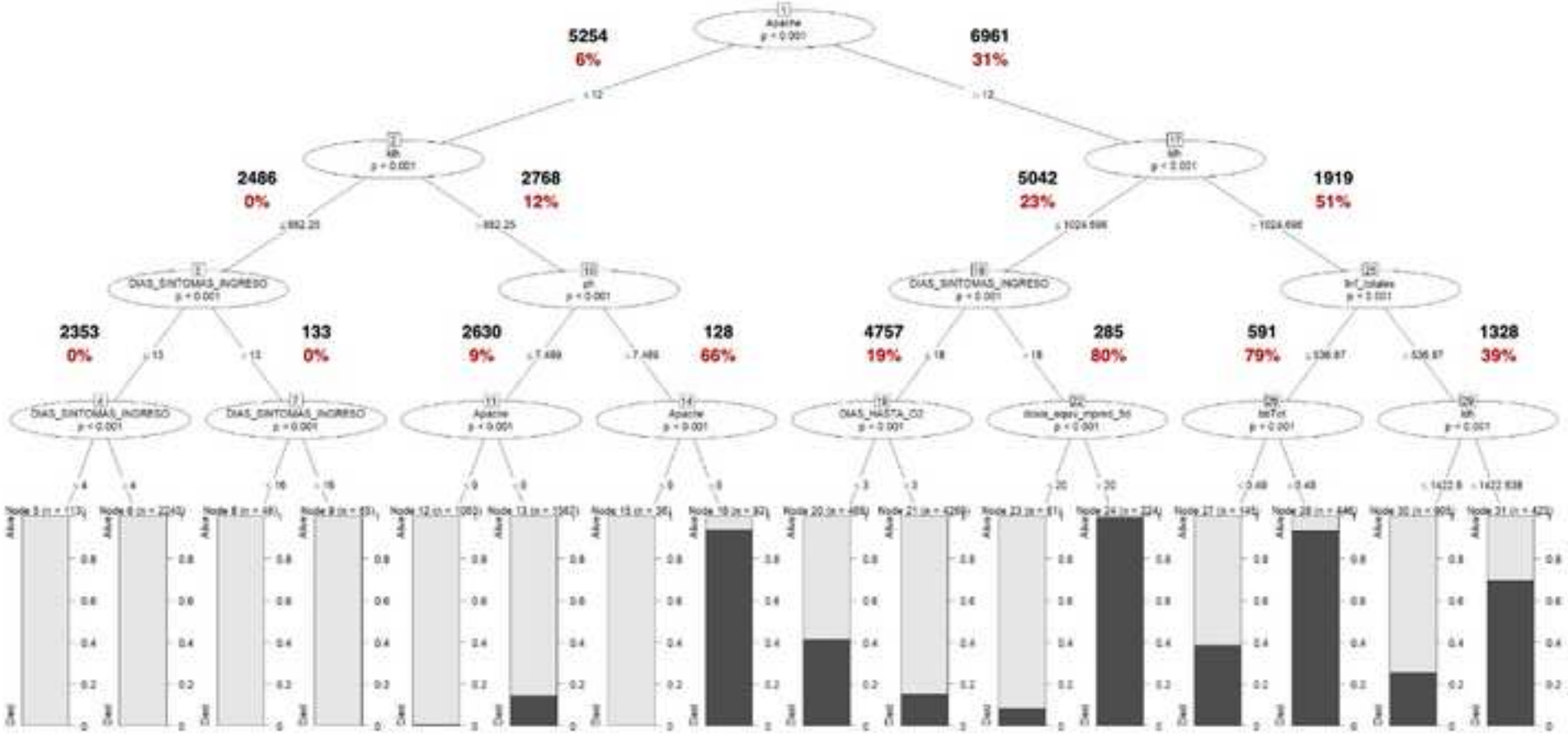


Figura / Figure

Pulse aquí para acceder/descargar;Figura / Figure;Fig3_MortalityTree.jpg



Material adicional (documentos, videos...)

Pulse aquí para acceder/descargar

Material adicional (documentos, videos...)

AdditionalMaterials02192023.docx

Contribution of the Authors

SM, MA and AN conceived the presented idea. SM and MA developed the theory and performed the computations. AN conducted an independent literature search to identify potentially relevant studies. MS independently reviewed the search results to identify pertinent articles. MS, AN, TF and VY contributed to the interpretation of the results. SM, MA, AN, MS, FL and AC took the lead in writing the manuscript. All authors provided critical feedback and