Submitted: 15 July, 2024 Accepted: 24 September, 2024 Published: 08 December, 2024 DOI:10.22514/sv.2024.152

Open Access
Signa Vitae

ORIGINAL RESEARCH

Predicting COVID-19 mortality using statistical, machine learning and fuzzy classification methods: insights from a Portuguese cohort study

Cecilia Castro[1], Víctor Leiva[2,*], Pedro Cunha[3], Muhammad Azeem Akbar[4]

[1]Centre of Mathematics, Universidade do Minho, 4710-057 Braga, Portugal
[2]School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, 2362807 Valparaíso, Chile
[3]Emergency Department, Hospital Senhora da Hora, 4835-044 Guimarães, Portugal
[4]Department of Information Technology, Lappeenranta University of Technology, 53850 Lappeenranta, Finland

***Correspondence**
victor.leiva@pucv.cl;
victorleivasanchez@gmail.com
(Víctor Leiva)

Abstract

The prediction of mortality in hospitalized COVID-19 patients using non-invasive and easily accessible measurements remains essential for improving patient outcomes, particularly in fast-paced clinical environments. The present study integrates generalized linear models (GLMs), fuzzy rule-based systems, and advanced machine learning algorithms—as support vector machines (SVMs), gradient boosting machines (GBMs), and random forests (RFs)—to predict COVID-19 mortality. The study was conducted on data from a Portuguese hospital, using patient age, length of stay, maximum oxygen administered, and timing of remdesivir (RDV) therapy as key predictors. Logistic regression provided high predictive performance, with an area under the curve (AUC) of 0.908, while the glmnet model achieved AUC = 0.892. Although ensemble methods such as RF (AUC = 0.922) and SVM (AUC = 0.952) demonstrated high accuracy, logistic regression remained competitive and it is superior due to its interpretability. Fuzzy models identified RDV as an important predictor (13.32% contribution), but with ambiguous effects. The logistic regression model found that delayed RDV administration increases mortality risk. These findings underscore the complexity of RDV impact on outcomes and highlight the importance of combining statistical models with machine learning techniques to enhance clinical decision-making for COVID-19 patients.

Keywords

Advanced predictive analytics; Artificial intelligence; Ensemble methods; Fuzzy rule-based classification; Generalized linear models; Non-invasive clinical predictors; Remdesivir treatment; SARS-CoV-2

1. Introduction

The COVID-19 pandemic has posed unparalleled challenges to the global healthcare community, necessitating sophisticated analyses of complex medical data. Recent advancements in statistical modeling and computational techniques have been pivotal in addressing these challenges. For example, the incorporation of frailties into cure rate regression models [1] and the development of robust autoregressive modeling [2] have provided valuable insights into statistical modeling and diagnostic analytics that can be tailored for COVID-19.

Methodological innovations such as the numerical treatment of new fractional-order differential models [3] and the application of artificial neural networks for solving complex mathematical models related to COVID-19 [4, 5] have expanded the analytical tools available for pandemic response. The integration of internet of things (IoT)-fuzzy intelligent systems for patient management [6] and novel approaches to pandemic modeling [7, 8] demonstrate the potential of combining statistical modeling with computational techniques to enhance pandemic management strategies. This integration can be complemented with advanced statistical techniques and the application of fuzzy and crisp solutions to model pandemic dynamics [8, 9], enriching the analytical spectrum and improving our understanding of COVID-19 broad effects.

The critical role of advanced statistical and machine learning techniques in unraveling the complexities of the pandemic has become increasingly evident. These techniques offer crucial insights into effective pandemic management strategies and individual responses [10]. Studies conducted across various countries have underscored the importance of these techniques in understanding COVID-19 impacts and outcomes [11–16].

In particular, the necessity for precise predictive models is paramount, especially for forecasting mortality among critically ill COVID-19 patients. Accurate mortality predictions have implications for healthcare planning and resource allocation [17], especially for socioeconomic factors that impact patient outcomes, as demonstrated by investigations on the effect of socioeconomic deprivation on mortality rates after

intensive care unit admission [18].

Despite these investigations, current research still lacks models that effectively integrate non-invasive predictors with treatment variables to forecast mortality in hospitalized and severely ill COVID-19 patients, producing a gap in the literature. Addressing this gap, the present study focuses on non-invasive predictors that are accessible and easily interpretable by healthcare professionals, emphasizing their ability to accurately forecast mortality. Specifically, the main objective of our study is to investigate the relationship between several predictor variables and the binary outcome of survival or death in hospitalized COVID-19 patients requiring ventilation. We apply advanced modeling techniques to validate the effectiveness of these predictors in clinical settings, thereby providing actionable insights in medical research and clinical practice.

To implement the present investigation, we employ a diverse set of analytical methods, including generalized linear models (GLMs) [19–23] and fuzzy rule-based classification systems (FRBCS) [24–34] that provide cognitive-like insights and clear decision-making rules, balancing statistical robustness with interpretability [35–39]. GLMs, particularly logistic regression, are particularly well-suited for modeling binary outcomes like mortality, providing clear insights into the relationships between predictors and outcomes. We also utilize glmnet, a regularized extension of GLMs, which enhances model generalization, especially in high-dimensional datasets, by integrating elastic net regularization.

In addition to the statistical models, we consider machine learning methods [40], including support vector machines (SVMs), gradient boosting machines (GBMs), and random forests (RFs). These methods excel at capturing non-linear relationships and interactions within the data, often surpassing traditional models like logistic regression and glmnet in predictive performance. Despite their complexity and potential challenges in interpretability, the inclusion of GBM, RF and SVM is essential for a comprehensive analysis, as they can model intricacies that simpler statistical methods might miss.

A novel aspect of our study is the inclusion of remdesivir (RDV) therapy, an antiviral treatment that has been widely used during the COVID-19 pandemic. RDV has been associated with varying outcomes in COVID-19 patients, depending on the timing of its administration. By incorporating the interval from COVID-19 diagnosis to the commencement of RDV therapy into our models, we aim to explore the implications of treatment timing alongside other non-invasive predictors.

The above-mentioned incorporation, grounded in clinical criteria and guidelines, allows us to provide a more comprehensive understanding of how RDV impacts mortality, particularly when combined with other clinical indicators. Therefore, drawing on recent research [2, 41, 42], our study introduces an innovative methodology to predict mortality in COVID-19 patients by integrating non-invasive clinical predictors with treatment variables, such as RDV timing. The proposed machine learning methodology, integrated with advanced statistical models like GLMs and FRBCS, represents an advancement in medical data analytics.

Our study utilizes a dataset from a Portuguese hospital, consisting of patients diagnosed with COVID-19 via polymerase chain reaction who exhibited severe symptoms requiring supplemental oxygen ($\geq$24%). These patients were categorized into two groups: one receiving RDV, where all patients received RDV and 97.1% were previously treated with dexamethasone, and the other one receiving the standard of care (SOC) without RDV. This dataset allows us to investigate the combined effects of RDV treatment and non-invasive clinical indicators on COVID-19 mortality.

Our focus on the implications of RDV treatment, combined with non-invasive clinical indicators, addresses a critical gap in current research. Most existing studies consider treatment effects or non-invasive predictors separately, but our integrated methodology provides a more comprehensive view of the factors affecting COVID-19 mortality. It is important to note that our research represents an initial exploration in the area under study. Our findings lay the groundwork for future research examining these treatment modalities and their impacts on patient outcomes. Therefore, we validate the effectiveness of combining non-invasive predictors with RDV treatment in predicting mortality among severely ill COVID-19 patients. Our investigation contributes to the existing body of knowledge and creates new opportunities for further research. The present analysis provides meaningful insights into COVID-19 mortality prediction, equipping medical professionals with enhanced tools for therapeutic decision-making, and ultimately optimizing patient care strategies during a global health crisis.

The subsequent sections of this article are organized as follows. Section 2 outlines our study population, the dataset, and analytical methodology used. In Section 3, our findings and their discussion are presented. We conclude in Section 4, summarizing our insights and suggesting future research directions.

2. Data and methodology

In this section, to ensure a rigorous and comprehensive analysis, we adopt a dual-pronged approach. We first focus on the demographic and clinical specifics of our cohort under study, followed by an in-depth discussion of the methodological framework. This approach aims to provide a systematic breakdown of our research process.

2.1 Study population and data description

Of the 1200 patients initially screened, 45 were excluded from the study. Reasons for exclusion included missing critical data, failure to meet the inclusion criteria (such as lack of radiological confirmation of pneumonia or not requiring supplemental oxygen), or withdrawal from the study. Therefore, our study included 1155 adult patients hospitalized with radiologically confirmed pneumonia caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), all of whom required supplemental oxygen ($\geq$24%). Among these patients, 843 were treated with RDV, while 312 received SOC, providing a diverse dataset for analysis. Fig. 1 shows the flow of participants in the study.

The allocation of RDV within our cohort was based on clinical criteria and the patient condition, following evolving

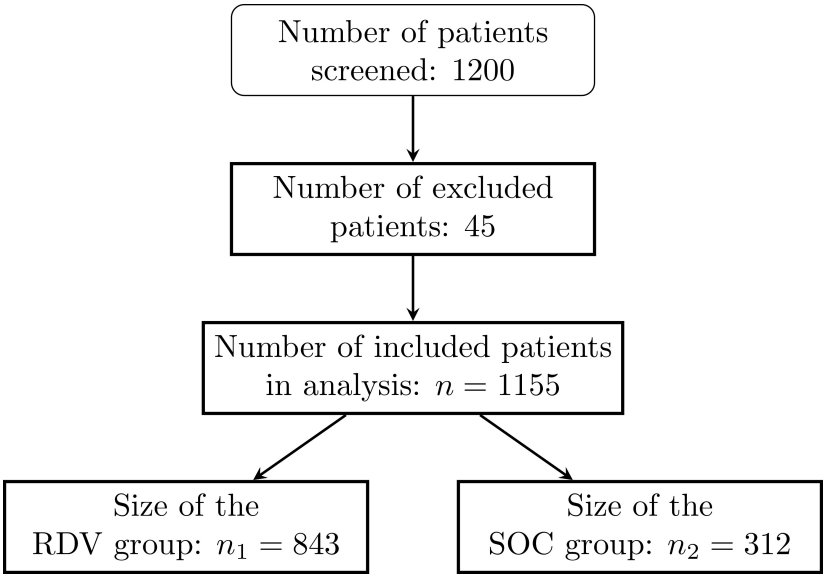

FIGURE 1. Flowchart of participant selection and allocation. RDV: remdesivir; SOC: standard of care.

medical guidelines. Criteria for RDV administration centered around the severity of the patient condition, indicated by factors such as the extent of hypoxemia and the immediate need for supplemental oxygen. Other considerations included the overall health status of the patient, using underlying conditions that could impact the effectiveness of the treatment or introduce potential risks. The timing of RDV therapy initiation was carefully aligned with national health authority recommendations and the latest clinical research findings. This strategic decision-making process reflects the agility of the healthcare system in integrating new therapeutic insights into patient care protocols, an important aspect of our analysis. Stratifying patients into RDV and SOC groups is integral to assess the impact of RDV alongside other non-invasive predictors on patient mortality. Specifically, our models concentrate on patients who received RDV treatment. The timing of RDV initiation—a key variable in our analysis—is relevant only within this patient group. Such timing allows us to isolate and evaluate the impact of RDV treatment on patient outcomes, especially when analyzed alongside key variables that are readily accessible to healthcare professionals and vital for clinical decision-making. Table 1 provides a comprehensive summary of the clinical characteristics of the patients included in the study, stratified by treatment group (RDV and SOC).

Our preliminary data analysis provides a transparent and detailed view of the patient cohort under study. In the RDV group, 97.1% of patients also received dexamethasone, with a mean age of 69.7 years and a standard deviation of 14.4 years, with 61.7% of them being male. In contrast, the SOC group had a mean age of 73.9 years and a standard deviation of 14.5 years, with 50.0% of them being male. Despite similarities in the prevalence of diabetes, hypertension, and chronic lung diseases between the groups, the RDV group had a higher proportion of overweight individuals, whereas the SOC group included more individuals with immunosuppressive conditions and smokers. Notably, patients in the RDV group had an average hospital stay of 4.25 days, shorter than the 5-day average of the SOC group, suggesting a potential benefit of RDV in reducing hospitalization duration. Furthermore, the relative risk of death during the hospital stay for the RDV group was 0.47 compared to the SOC group, indicating that patients treated with RDV had a 53% lower risk of mortality during hospitalization compared to those receiving SOC. This reduction in relative risk suggests a potential survival benefit associated with RDV treatment as part of a multifaceted approach to treating severely ill COVID-19 patients. To build on these findings, we are actively working on a more detailed analysis, which will be presented in an upcoming publication. This future publication will delve into the clinical implications of RDV treatment, specifically addressing how relative risk and other critical metrics impact patient outcomes. We aim to provide a comprehensive perspective that will further enhance understanding in the area under study.

2.2 Machine learning methods

We recall that our primary objective is to investigate the relationship between the predictor variables and the binary outcome of survival or death in hospitalized COVID-19 patients requiring ventilation, presented in Table 1. As previously mentioned in Subsection 2.1, our analysis is focused exclusively on the group treated with RDV, particularly assessing the impact of RDV treatment timing on patient outcomes. To effectively analyze the mentioned variables, we utilize GLMs due to their aptitude for describing binary outcomes like patient survival or death.

GLMs are advantageous as they describe the probability of an outcome (ranging from zero to one) as a function of various predictor variables. This modeling allows us to consider a range of relevant clinical and demographic factors within our patient cohort. In the GLM framework, the binary outcome, denoted as Y, indicates the survival ($Y = 1$) or non-survival ($Y = 0$) of the patients. The predictor variables, collectively referred to as X and presented in Table 1, encompass diverse predictors relevant to patient outcomes.

We use logistic regression, a type of GLM, to analyze the relationship between predictors and patient survival proba-

TABLE 1. Characteristics of the study population stratified by treatment with Portuguese COVID-19 data.

Variable	Class	RDV group ($n_1 = 843$)	SOC group ($n_2 = 312$)
Gender			
	Male	520 (61.7%)	156 (50.0%)
	Female	323 (38.3%)	156 (50.0%)
Smoking status			
	Non-smoker	347 (41.2%)	100 (32.1%)
	Smoker	112 (13.3%)	20 (6.4%)
	Unknown	384 (45.5%)	192 (61.5%)
Immunocompromised			
	No	722 (85.6%)	256 (82.1%)
	Yes	121 (14.4%)	56 (17.9%)
Obesity			
	No	609 (72.2%)	215 (68.9%)
	Yes	234 (27.8%)	97 (31.1%)
Diabetes			
	No	552 (65.5%)	211 (67.6%)
	Yes	291 (34.5%)	101 (32.4%)
Hypertension			
	No	298 (35.3%)	104 (33.3%)
	Yes	545 (64.7%)	208 (66.7%)
Autonomy level			
	Autonomous	598 (70.9%)	191 (61.2%)
	Partially dependent	145 (17.2%)	41 (13.1%)
	Dependent	100 (11.9%)	80 (25.6%)
Pulmonary disease			
	No	690 (81.8%)	281 (90.1%)
	Yes	153 (18.2%)	31 (9.9%)
Mortality			
	Survival	664 (78.8%)	254 (81.4%)
	Non-survival	179 (21.2%)	58 (18.6%)

RDV: remdesivir; SOC: standard of care.

bility $p = P(Y = 1|X = x)$. Logistic regression calculates the odds of survival as an odds ratio (OR), and mathematically is defined as $p/(1 - p)$ providing a direct measure of the probability that a patient survives based on the specific characteristics and treatments. The OR derived from the logistic regression is a crucial metric in our analysis. The OR indicates the strength and direction of the association between each predictor and patient survival. An OR greater than one indicates an increased probability of survival, while an OR less than one suggests a decreased probability. In our logistic regression analysis, we model the logarithm of the OR (log-OR), denoted as Z and its observed value as z, establishing a linear combination of predictors defined by $Z = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, where $\beta_0, \beta_1, \ldots, \beta_k$ are the model parameters, each reflecting the impact of the predictors $X_1, \ldots, X_k$ based on their observed values $x_1, \ldots, x_k$, on patient survival. To convert log-ORs into the probability of survival, p say, we use the sigmoid function stated as:

$$\sigma(z) = p = \frac{1}{1 + \exp(-z)} \quad (1)$$

Which converts a real number into a value at [0, 1]. This conversion translates the linear combination of predictors into a probability value. Such a conversion provides a more intuitive understanding of the probability of a patient survival based on the characteristics and treatments. To train the model, we optimize the parameters β_j, for $j = 0, 1, \ldots, n$, to minimize a cost function formulated as:

$$J(\beta) = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)) \quad (2)$$

Where y_i is the observed class (survival or non-survival) for instance i in the dataset, $\hat{p}_i$ is the corresponding estimated/predicted probability of survival for that instance, and n is the number of instances in the dataset.

The cost function stated in Eqn. 2 is convex, which ensures that an optimization algorithm, such as gradient descent, can find the global minimum, leading to the best set of parameters for the model.

Once trained, the model uses the estimated regression parameters to calculate the survival probability using the function stated in Eqn. 1. Instance $i$belongs to the survival class if $\hat{p}_i > 0.5$, that is, $\hat{y}_i = 1$ when $\hat{p}_i > 0.5$; or $\hat{y}_i = 0$, otherwise, where $\hat{y}_i$ is the estimated value of y_i. Despite being called "regression", logistic regression is mainly used for classification. It is effective for linearly separable classes but its performance decreases with non-linear classes. High correlation among predictors can lead to overfitting, destabilizing the coefficients. Overfitting occurs when a model learns of the training data closely, including its noise, leading to poor performance on unobserved data.

For the case of non-linear data or when a more flexible decision boundary is needed, SVM is preferred over logistic regression. SVMs belong to supervised machine learning designed for classification, but also cater to regression tasks [43, 44]. SVM creates a hyperplane that best separates classes in the data. In a two-dimensional space, this hyperplane is a line. For data that are linearly separable, an SVM finds the line that maximizes the margin between the nearest points (support vectors) of each class. This line is termed the decision boundary, and the area delineated by the margin is known as the "street". For non-linearly separable data, SVM uses a kernel trick, which transforms the data into a high-dimensional space to be partitioned by a hyperplane. Once the hyperplane is determined in this space, it may be mapped back to the original space, resulting in a non-linear decision boundary.

Data often contain noise and classes may overlap, which is challenging to identify a hyperplane that separates the classes. To address such a challenge, SVM introduces a soft margin allowing some points to fall on the wrong side of the margin or even the decision boundary. This wrong classification is avoided balancing maximization of the margin and minimizing misclassification. While SVMs provide classification and regression, we must acknowledge its limitations when dealing with large datasets due to computational demands. However, the concept of optimizing the margin of fit, in classification or regression, remains consistent. SVM regression, addressing non-linearities using kernel trick and its margin-based approach, accommodates as instances as possible within a margin, whose width is regulated by a parameter, balancing complexity of the model and its fit to the data.

In scenarios dealing with large datasets, complex interactions, or the necessity for stability against overfitting, ensemble learning methods, like RF, emerge as powerful solutions. RF is particularly beneficial when the limitations of single-model approaches, such as logistic regression, become evident in handling complex data patterns.

RF creates multiple decision trees during training and aggregate their outputs, following a "wisdom of the crowd" approach. This is represented by the mode of the classes from individual trees [45, 46]. The RF algorithm involves the following steps:

• Step 1—Begin by selecting random samples from the dataset with replacement, known bootstrap samples [47], where reach sample typically includes B instances, with B being the same as the total number of instances n in the original dataset or a different specified number.

• Step 2—Create a decision tree for each bootstrap sample and at each node of the tree considering:

- Step 2.1—Select a subset of variables (features), denoted by m, at random from the total number of variables, k say, in the dataset, where typically m is chosen to be the square root of k.

- Step 2.2—Determine the optimal split on these m variables to divide the node, with the split maximizing the reduction in variance (for regression trees) or minimizing the Gini impurity (for classification trees), and then the node is divided into left and right child nodes. The process of finding the optimal split is mathematically represented as $\Delta I(S, A) = I(S) - \sum_{v \in values(A)} (|S_v|/|S|)\, I(S_v)$, where $I(S)$ is the impurity measure of the original set, values(A) represents the set of all possible values for attribute x, S_v is the subset of S for which attribute x has value v, and $|S_v|/|S|$ is the proportion of instances that have value v for attribute x. The attribute that provides the highest information gain is chosen for the split.

• Step 3—Pass a new object through all the trees in the forest, where each tree casts a vote for a class, and the forest selects the class with the majority of votes from all the trees as its prediction.

A crucial parameter of the RF is the number of trees to generate. A larger number of trees reduces the possibility of overfitting but increases the computational complexity. RFs are robust to outliers, scalable, and capable of modeling non-linear decision boundaries due to their hierarchical structure. Moreover, RF automatically selects the variables that are important in the classification.

RFs can model complex non-linear decision boundaries and do not require input features to be scaled, that is, to have the same range, unlike methods such as SVMs or logistic regression. Despite the power and robustness of an RF, its intricate nature may complicate the decision-making processes. While an RF may struggle with interpretability, it shares with some other models the need for computational resources, particularly when considering large ensembles of deep trees.

Comparatively, stochastic gradient boosting, also known as GBM, combines boosting with regularization to enhance model predictions [48]. Unlike RFs, which construct trees in parallel and focus on reducing variance among their models, GBM builds trees sequentially, concentrating on minimizing bias and improving accuracy of the model by correcting errors from previous trees.

The strength of GBM lies in its ability to combine multiple weak decision-tree-based learners, sequentially refining the model to form a more potent learner. This sequential learning allows GBM to reduce both bias and variance, making it particularly helpful in scenarios where predictive accuracy is paramount. Although each decision tree individually might not yield the best predictions in GBM, they provide improved performance collectively. The trees are constructed to mini-

mize a specific loss function, with every new tree focusing on the previous errors of the model. A distinguishing feature of the GBM is its inclusion of a regularization component that is used to circumvent overfitting. By introducing a regularization term, GBM seeks simplicity of the model, often preferring those ones that balance complexity and performance. The GBM framework emphasizes the model by minimizing the chosen loss function, which evaluates the discrepancy between predicted and observed outcomes.

Given a training set $\{x_i, y_i\}$, for $i \in \{1, \ldots, N\}$, where x_i represents the predictor variable for instance i and y_i the corresponding label, GBM employs an ensemble of decision trees to compute the prediction w_i for each instance i defined as:

$$w_i = \sum_{j=1}^{T} f_j(x_i),\ i \in \{1, \ldots, N\} \tag{3}$$

Where f_j represents tree j in the ensemble, contributing scores to its leaves, and T indicates the total number of trees.

The efficacy of each tree j is evaluated using an aggregated loss function L_j, which integrates both the prediction error and a regularization component, stated as:

$$L_j = \sum_{i=1}^{N} l(y_i, w_i + f_j(x_i)) + \Omega(f_j),\ j \in \{1, \ldots, T\}$$

Where l represents the individual loss function, assessing the prediction error for each instance i, w_i is defined in Eqn. 3 and denotes the aggregate prediction up to tree j for the sample x_i, while Ω introduces a regularization term to penalize the complexity of the model. Common examples of l include the squared error for regression tasks and the logistic loss for classification tasks. They aim to tune the model predictions to closely align with the observed labels. The function $f_j(x_i)$ minimizes L, leading to constructing each tree, iteratively refining the performance of the model.

The choice between GBM and RF hinges on specific requirements. The trade-off involves considerations of interpretability, computational efficiency, and predictive accuracy. RF may be preferred for its simplicity and robustness, especially when interpretability and computational resources are restricted. Conversely, GBM might be favored in applications where achieving the highest predictive accuracy is paramount and the data are complex.

2.3 Fuzzy methods

Within the continuum of modeling techniques that address data complexities, FRBCS emerge as a versatile alternative to ensemble models like GBM and RF, as well as a broad array of traditional and advanced methodologies [49–51]. By leveraging the concept of fuzziness, FRBCS excel in managing uncertainties and ambiguities that traditional algorithms often fail to address, making it particularly valuable in scenarios where conventional (crisp) logic proves inadequate [39]. The widespread application of fuzzy logic across disciplines such as artificial intelligence, computer science, and mathematics has driven advancements, showing its flexibility and broad applicability.

FRBCS provide good solutions to both classification and regression problems by integrating human-like reasoning with computational precision. At the core of FRBCS is the fuzzification interface, which converts precise inputs into degrees of membership using various functions. This conversion accommodates the inherent vagueness present in data, making FRBCS powerful tools for handling complex scenarios where traditional crisp logic fall short.

In the fuzzification process, we begin with a universe of discourse, χ say, representing the complete set of all possible values that a variable can assume. An element v within this universe represents a specific value of the variable. A fuzzy set A is defined as a subset of this universe, where each element v is assigned a membership degree by the function $\mu_A(v)$. This degree, ranging from 0 to 1, indicates the extent to which the element v belongs to the fuzzy set A, thereby transforming precise input values into a range of possibilities that can be processed by the fuzzy logic system.

Various types of membership functions can be used in the context of the present study. The triangular membership function is defined as:

$$\mu_A(v) = \begin{cases} 0, & v \le a \text{ or } v \ge b; \\ \frac{v-a}{c-a}, & a \le v < c; \\ \frac{b-v}{b-c}, & c < v \le b; \end{cases}$$

Where a is the lower limit, b the upper limit, and c the modal value with $a \le c \le b$.

The trapezoidal membership function is established as:

$$\mu_A(v) = \begin{cases} 0, & v \le a \text{ or } v \ge d; \\ \frac{v-a}{c-a}, & a \le v < c; \\ 1, & c \le v \le c'; \\ \frac{d-v}{d-c'}, & c' < v \le d; \end{cases}$$

Where a is the lower limit, d the upper limit, and $[c, c']$ the interval modal value with $a \le c \le c' \le d$.

The Gaussian membership function is expressed as:

$$\mu_A(v) = \exp\left(-\frac{(v-c)^2}{2\sigma^2}\right)$$

Where c is the mean and σ the standard deviation of the Gaussian model.

Studies have shown that different membership functions can yield varying degrees of accuracy depending on the nature of the input data and the problem domain. The fuzzy sets obtained through fuzzification are then integrated within a fuzzy logic system, which typically includes:

• A fuzzy knowledge base consisting of a database (DB) and a rule-base (RB). The DB defines fuzzy sets with memberships expressed between 0 and 1, while the RB stores fuzzy if-then inference rules (IF-THEN rules). The premises rely on the fuzzy sets defined in the DB, while the conclusions specify the

resulting actions or outcomes.

• An inference engine that processes fuzzy IF-THEN rules using methods tailored to fuzzy rule-based classification systems.

• A defuzzification process that converts fuzzy sets into actionable outputs, potentially integrated within the inference process.

In line with the basis of a fuzzy logic system, our FRBCS implementation employs several specific strategies to effectively address the complexities inherent in medical diagnosis as the following:

• Max-membership principle—This principle directly selects the class with the highest membership degree as the output, avoiding the need to convert fuzzy sets into crisp values. The principle ensures direct and actionable results, particularly in scenarios involving ambiguous or overlapping data.

• Rule weighting (FRBCS.W) strategy—In this strategy, each fuzzy rule is assigned a weight that reflects its relevance in the classification process. By prioritizing the most relevant rules, the model improves prediction accuracy, especially when dealing with complex and diverse data patterns.

• Space partitioning, the chi-square (FRBCS.CHI) method—This method applies the chi-square test to partition the input space, ensuring that the resulting fuzzy partitions are statistically significant. It enhances the alignment between fuzzy rules and the underlying data distribution, leading to more reliable and interpretable classifications.

• Genetic (GFS.GCCL) algorithms—These algorithms optimize the fuzzy rule sets, utilizing cooperative-competitive learning to evolve and refine the rules. Genetic algorithms are particularly effective in exploring large and intricate search spaces, making them ideal for optimizing fuzzy systems in complex applications.

• Pittsburgh (FH.GBML) method—This method focuses on evolving entire populations of fuzzy IF-THEN rule sets simultaneously, optimizing the entire rule set collectively. This holistic method is effective in managing complex classification tasks, ensuring cohesive and accurate outcomes.

In our FRBCS models, defuzzification is seamlessly integrated into the inference process, with the max-membership principle serving as the primary method. This guarantees that outputs are actionable by directly selecting the class with the highest membership degree, without converting fuzzy sets into crisp values. These FRBCS strategies are implemented using a comprehensive software package, as referenced in [52], which supports the deployment of these advanced fuzzy logic methods for complex real-world applications.

The performance of the FRBCS models in medical diagnosis is closely tied to the quality of data and complexity of decision boundaries. The interaction of symptoms demands flexible rule sets that adapt to evolving data patterns. By incorporating machine learning and big data, our methodology aims to enhance the adaptability of fuzzy models, ensuring precision and relevance in dynamic environments – essential aspects for successful real-world applications.

2.4 Performance metrics and validation methods

Evaluating the predictive performance and reliability of a trained model is essential. This evaluation encompasses a variety of metrics and techniques, such as cross-validation or a simple train-test split. We provide an overview of these metrics and techniques, with further details presented in [53].

Before discussing specific metrics, it is important to define the fundamental concepts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These concepts state the outcomes of the predictions of a binary classification model in relation to the observed values and correspond to:

• TP—Correctly predicted positive observations.
• TN—Correctly predicted negative observations.
• FP—Incorrectly predicted positive observations.
• FN—Incorrectly predicted negative observations.

The confusion matrix, a tool to visualize the performance of a binary classification model, summarizes the concepts of TP, TN, FP, and FN, as shown in Table 2.

TABLE 2. Confusion matrix calculation method.

	Predictive positive	Predictive negative
Observed positive	TP	FN
Observed negative	FP	TN

TP: true positive; FN: false negative; FP: false positive; TN: true negative.

Other metrics from Table 2 include precision, sensitivity (recall), accuracy, and the area under the curve (AUC) of the receiver operating characteristic (ROC). Precision and sensitivity focus on positive predictions and are defined as:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{sensitivity (or recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Accuracy provides a broader measure of performance for a model and is established as:

$$\text{accuracy (binary)} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

The AUC is a metric used to evaluate the performance of binary classification models. The ROC curve plots the TP rate (sensitivity) against the FP rate (1 − specificity) at various threshold settings. The AUC is a metric that represents the probability that a randomly chosen positive instance is ranked higher by the model than a randomly chosen negative instance. An AUC of 0.5 indicates no discriminative ability (equivalent to random guessing), while an AUC of 1.0 signifies perfect discrimination. This metric is particularly helpful in medical research as it provides an aggregate measure of performance

across all classification thresholds, assessing a model ability to distinguish between positive and negative cases irrespective of the decision threshold.

In addition to AUC, the area under the precision-recall curve (AUPRC) is especially informative for evaluating the performance of models on imbalanced datasets. Unlike the AUC-ROC, which considers both FPs and FNs equally, the AUPRC focuses on the performance concerning the positive class (the minority class). In imbalanced datasets, where the negative class dominates, the AUPRC offers a better measure of model ability to correctly identify the positive cases by emphasizing precision (the proportion of TPs among all predicted positives) and recall. A high AUPRC indicates that the model performs well in distinguishing the minority class from the majority class, which is crucial in medical research where the consequences of FNs can be particularly severe, such as in disease detection or risk prediction.

We also calculate bootstrapped confidence intervals (CIs) for accuracy, sensitivity, specificity, and AUC. Bootstrapping is a non-parametric statistical method that enables us to estimate the distribution of these metrics by repeatedly sampling, with replacement, from the original data. Bootstrapping allows for the calculation of CIs without relying on parametric assumptions. By generating multiple bootstrap samples and computing the desired metric for each, we obtain an estimate of the metric variability, making bootstrapping especially helpful when the data distribution is unknown or the sample size is small.

Additionally, the choice between cross-validation and a train-test split depends on the dataset and the type of modeling. Cross-validation, particularly *k*-fold, is preferred when seeking reliable performance, especially with small datasets, as it maximizes the use of data for both training and testing, providing a measure of model performance. However, it is computationally more intensive than a train-test split, which might be more suitable for very large datasets or when a quicker model assessment is required due to its computational efficiency. This ensures that the chosen method aligns with the specific requirements and constraints of our study, balancing accuracy, and computational resources.

2.5 Methodology and software

Next, we outline the methodology and software tools employed in our study, emphasizing computational considerations and the rationale behind our methodological choices. We utilized the R software and caret library [54] for machine learning and data modeling, focusing on mortality outcome classification. Although caret provides access to a variety of models, we tailored our methodology to address the unique challenges posed by COVID-19 data, selecting models that best fit our research questions and data characteristics.

For model training and evaluation, we employed the train function within caret, which allows the application of different modeling techniques through its versatile method parameter. Our methodology included models such as GLM with stepwise selection based on the Akaike information criterion (AIC)—glmStepAIC, glmnet, GBM (gbm), RF (rf), and SVM with a radial basis function kernel (svmRadial).

Elastic net regularization, used by the glmnet model, combines the strengths of both least absolute shrinkage and selection operator (LASSO) and ridge regressions. LASSO regression, employing the L1 norm (sum of absolute values of coefficients), is effective in variable selection by shrinking some coefficients to zero, so simplifying the model. Ridge regression, using the L2 norm (sum of squares of coefficients), helps to manage multicollinearity by distributing the coefficients among correlated predictors. The elastic net produces balancing, offering a model that generalizes well, particularly when dealing with high-dimensional datasets.

We optimized and evaluated the models based on the following criteria:

- Stopping criteria—Model training was halted using a predefined threshold for improvement in predictive accuracy on a validation set, aiming to minimize overfitting.
- Computational cost—Assessed by tracking the training time and resource usage for each model, prioritizing efficiency while maintaining performance.
- Solver for optimization—For models requiring optimization, such as glmnet and GBM, we used gradient descent and its variants, selected for their effectiveness in minimizing the cost function of the model architecture.

Given that the fuzzy methods utilized in our study do not directly provide a means to determine the relevance of individual variables due to their fuzzification rather than assigning direct weights, we employed permutation feature importance. This involves modifying the values of each predictor to assess its impact on model accuracy, allowing us to identify the variables most relevant in predicting COVID-19 mortality outcomes.

The pseudo-code for feature importance analysis has the following steps:

- Step 1—For each feature in the dataset:
- Step 1.1—Permute the values of the feature.
- Step 1.2—Evaluate model performance with the permuted data.
- Step 1.3—Calculate the change in performance.
- Step 2—Rank features based on performance change.
- Step 3—Identify key features influencing model accuracy.

These steps help to discern the relative importance of variables, thereby facilitating focused analysis and informed decision-making. For transparency and enabling replication of our findings, the detailed code, including the permutation feature importance analysis, is available upon request. The feature importance analysis is outlined in Algorithm 1 and illustrated in a flowchart in Fig. 2.

Algorithm 1: Permutation feature importance analysis.

Input: Given dataset and pre-trained fuzzy model.

Output: Ranked features based on importance.

1. Start with a pre-trained fuzzy model that has been fitted to the dataset.

2. For each feature in the dataset do:

2.1 Create a copy of the original dataset.

2.2 Permute the values of the current feature in the copied dataset.

2.3 Evaluate the model using the permuted dataset and a specified performance metric.

2.4 Record the change in performance compared to the base performance.

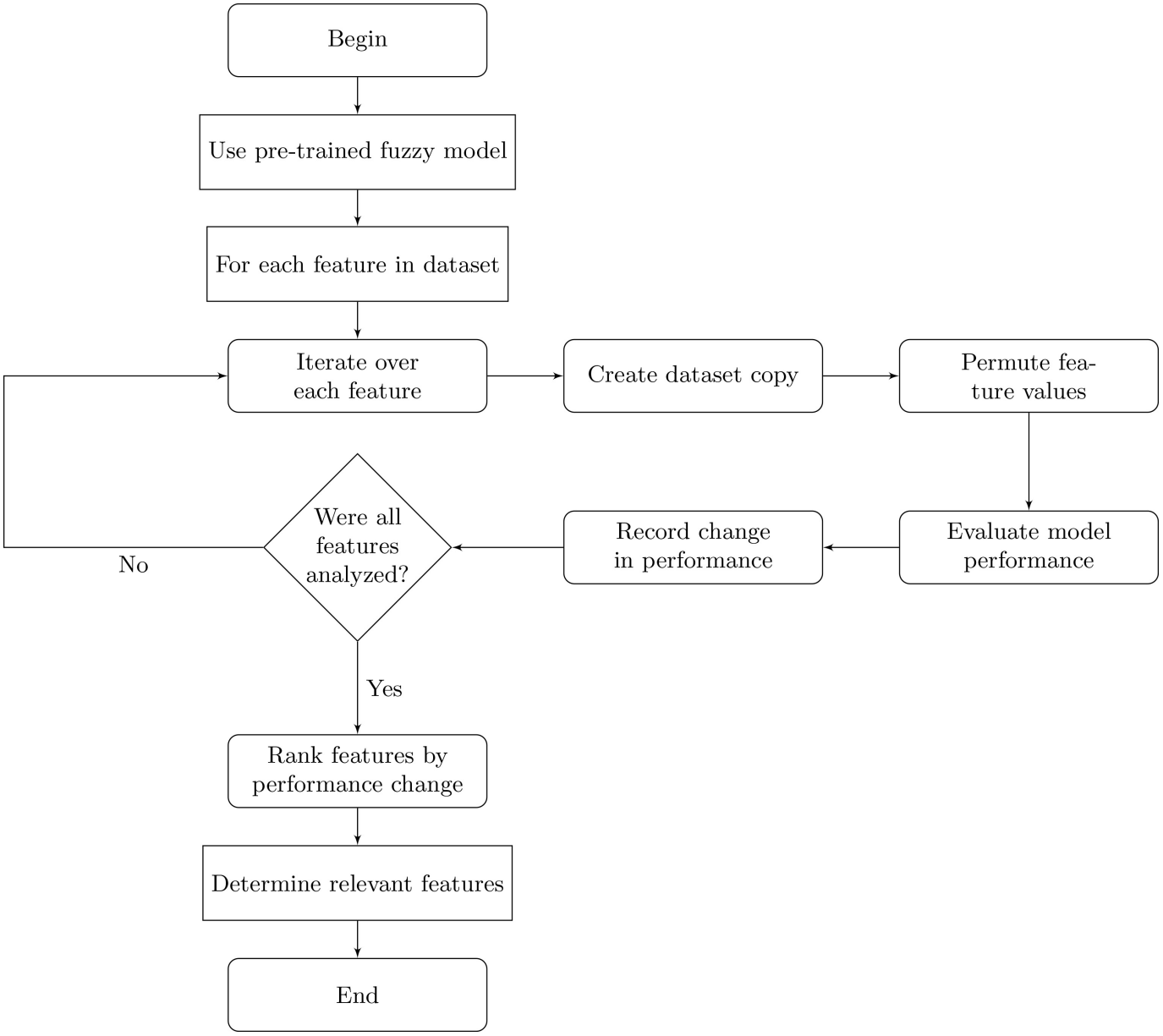

FIGURE 2. Schematic representation of the permutation feature importance algorithm.

3. Repeat Steps 1–2 for each feature in the dataset.

4. Rank the features according to the degree of change they caused after all features have been permuted and their effect on performance recorded.

5. State as more important those features causing a larger drop in performance.

6. Determine which features are relevant based on the previous steps.

3. Results and discussion on Portuguese cohort study

The present study evaluates and compares various methodologies to address uncertainties in Portuguese COVID-19 clinical data. In this section, we detail the models used, their applications, and the key predictor variables relevant to our analysis of patient outcomes.

3.1 Variable selection and data preprocessing

The predictor variables were chosen for their clinical relevance, ease of access for healthcare professionals, and non-invasive nature, which are defined as:

- Age—The patient's age in years, a basic metric for assessing risk factors associated with age-related vulnerabilities.
- Time_Adm_Sym—The duration from symptom onset to hospital admission, measured in days, reflecting the timeliness of treatment initiation.
- RDV_In_Sym—The interval from symptom onset to the start of RDV treatment, providing insights into the effectiveness of early antiviral interventions.
- Hospital_duration—The total duration of hospitalization in days. This variable is only available at discharge or death, but it is included to assess its potential impact on mortality predictions in retrospective analyses. By examining the length of stay, we can gain insights into the progression of severe cases

and identify factors that may contribute to longer stays, which are often associated with higher mortality. These insights could guide future research or influence decision-making in similar patient cohorts.

• FiO2Max—The maximum fraction of inspired oxygen required during hospitalization, expressed as a percentage, serving as a non-invasive proxy for respiratory compromise severity.

These variables contained no missing data, ensuring a consistent dataset, thereby eliminating the need for imputation. To ensure comparability across variables with different scales, all continuous predictor variables were normalized using min–max scaling, transforming their values to a 0–1 range. This transformation was crucial for maintaining consistency across all models, including the fuzzy systems.

The dataset was then divided into training and testing sets using an 80/20 split. This split maintained a representative distribution of survival and death outcomes in both sets. The same training and testing sets were applied across all models to maintain consistency and comparability.

3.2 Non-fuzzy model adjustments

To optimize the performance of the non-fuzzy models, including GBM, glmnet, RF and SVM, we conducted hyperparameter tuning, which involves selecting the best values for parameters that govern the model training process. This process enhances the model ability to learn from the data. Unlike these models, logistic regression does not require hyperparameter tuning, as it directly estimates relationships between predictors and outcomes through maximum likelihood estimation, relying on the data rather than tunable settings. In SVM, sigma controls the width of the Gaussian kernel, determining the influence range of a single training example, while *c* represents the penalty for misclassification. For GBM, n.trees is the number of trees, interaction.depth sets the tree depth, shrinkage controls the learning rate, and n.minobsinnode specifies the minimum observations per terminal node. For RF, mtry is the number of variables randomly sampled as candidates at each split in a tree. In glmnet, alpha determines the balance between LASSO and ridge penalties in glmnet, while lambda controls the strength of the regularization applied. The most frequently used hyperparameter settings for each model are summarized in Table 3.

To ensure the robustness of our results, a 5-fold cross-validation was employed on the training set, where the data were split into five subsets: four for training and one for testing. This process was repeated five times, ensuring each subset was tested once. Such a cross-validation provides stable estimation, making a separate validation set redundant. The validation was consistently applied across all models, ensuring comparability and consistency in the results.

3.3 Logistic regression

Logistic regression was implemented using the same cross-validation framework described above, although it did not require hyperparameter tuning. This regression was directly fitted to the training dataset, and the statistically significant predictors identified are presented in Table 4, reporting the impact of each predictor on the probability of death, assuming all other variables remain constant.

The OR quantifies the change in odds of death associated with a one-unit increase in each predictor variable, holding other factors constant. The interpretation of the coefficients reported in Table 4 is as follows:

• For every additional year of age, the odds of death increase by 6.5% (OR = 1.065), holding all other variables constant.

• For every additional day increase in Time_Adm_Sym, the odds of death decrease by 18.5% (OR = 0.815), assuming all other variables remain constant.

• An increase in RDV_In_Sym by one day raises the odds of death by 11.8% (OR = 1.118), provided all other predictors are held constant.

• For every additional day of Hospital_duration, the odds of death decrease by 4.9% (OR = 0.951), holding all else constant.

• For every unit increase in FiO2Max, the odds of death increase by 7% (OR = 1.070), assuming all other variables remain constant.

After fitting the model to the training data, we applied the logistic regression model to the test set to evaluate its predictive performance. The resulting confusion matrix is shown in Table 5.

Table 6 presents the performance metrics with 95% CIs, showing strong predictive performance. The high sensitivity indicates the model ability to accurately identify survival cases, while the AUC reflects its overall discriminative ability. However, the specificity, while lower, suggests a moderate ability to identify death cases accurately.

Together, Tables 5 and 6 illustrate the effectiveness of the logistic regression model in predicting patient outcomes, with a high accuracy and robust sensitivity, making it a valuable tool for clinical decision-making.

3.4 GLMNet model and coefficient estimation

The glmnet model is guided by two crucial parameters: alpha and lambda. The alpha parameter determines the balance between LASSO and ridge regressions within the model, with values ranging from 0 (pure ridge regression) to 1 (pure LASSO regression). In our study, the alpha and lambda parameters were determined through cross-validation to optimize the balance between model complexity and predictive accuracy. Specifically, alpha was set to 0.3, which introduces a bias toward ridge regression while retaining some feature selection capabilities of LASSO regression. The value of lambda was set to 0.0136, controlling the penalty on the model coefficients to prevent overfitting (see Table 3). Table 7 shows the estimated coefficients and their corresponding 95% CIs, obtained using bootstrapping to provide good estimates.

The interpretation of the coefficients and their 95% CIs is as follows:

• The coefficient for Age (estimated at 0.0493) has a 95% CI of [0.0316, 0.0667], suggesting an association between increased age and higher mortality risk.

• The coefficient for Time_Adm_Sym (estimated at −0.0905) has a 95% CI of [−0.1734, −0.0213], indicating a protective effect of earlier hospital admission.

TABLE 3. Summary of most frequently used hyperparameter settings for non-fuzzy models with Portuguese COVID-19 data.

Model	Hyperparameter	Description and value
SVM		
	sigma	0.3156, controlling the width of the Gaussian kernel.
	c	0.25, representing the penalty parameter for misclassification.
GBM		
	n.trees	150, the number of trees in the model.
	interaction.depth	1, the maximum depth of each tree.
	shrinkage	0.1, the learning rate that controls the contribution of each tree.
	n.minobsinnode	10, the minimum number of observations in each terminal node.
RF	mtry	2, the number of variables randomly sampled at each tree split.
glmnet		
	alpha	0.3, a value balancing between LASSO and ridge regressions.
	lambda	0.0136, the regularization strength applied to the model coefficients.

LASSO: least absolute shrinkage and selection operator; SVM: support vector machines; GBM: gradient boosting machines; RF: random forests.

TABLE 4. Logistic regression results with Portuguese COVID-19 data.

Variable	Estimate	Standard error	*z*-value	*P*-value	OR	95% CI for OR
Intercept	−9.848	1.251	−7.873	<0.001		
Age	0.063	0.014	4.693	<0.001	1.065	[1.037, 1.094]
Time_Adm_Sym	−0.205	0.064	−3.208	0.001	0.815	[0.717, 0.925]
RDV_In_Sym	0.111	0.048	2.307	0.021	1.118	[1.017, 1.228]
Hospital_duration	−0.050	0.017	−3.179	0.001	0.951	[0.918, 0.985]
FiO2Max	0.068	0.006	10.698	<0.001	1.070	[1.057, 1.083]

CI: confidence interval; FiO2Max: maximum fraction of inspired oxygen during hospitalization; Hospital_duration: total hospital stay in days; OR: odds ratio; RDV_In_Sym: days from symptom onset to RDV initiation; Time_Adm_Sym: days from symptom onset to hospital admission.

TABLE 5. Confusion matrix of the logistic regression model applied to the test set of Portuguese COVID-19 data.

Observed	Predicted	
	Survival	Death
Survival	127	8
Death	4	19

TABLE 6. Performance metrics for logistic regression model with 95% CIs using Portuguese COVID-19 data.

Metric	Estimate	Lower limit of the 95% CI	Upper limit of the 95% CI
Accuracy	0.924	0.880	0.962
Sensitivity	0.969	0.935	0.993
Specificity	0.704	0.519	0.870
AUC	0.908	0.816	0.984

CI: confidence interval; AUC: area under the curve.

TABLE 7. Coefficients of the glmnet model with 95% CIs using Portuguese COVID-19 data.

Variable	Estimate	Lower limit of the 95% CI	Upper limit of the 95% CI
Age	0.0493	0.0316	0.0667
Time_Adm_Sym	−0.0905	−0.1734	−0.0213
RDV_In_Sym	0.0320	−0.0059	0.0884
Hospital_duration	−0.0279	−0.0521	−0.0065
FiO2Max	0.0531	0.0468	0.0599

CI: confidence intervals; FiO2Max: maximum fraction of inspired oxygen required during hospitalization; Hospital_duration: length of stay in days; OR: odds ratio; RDV_In_Sym: time from symptom onset to RDV start in days; Time_Adm_Sym: time from symptom onset to hospital admission in days.

- The coefficient for RDV_In_Sym (estimated at 0.0320) has a 95% CI of [−0.0059, 0.0884], which includes zero, indicating that the effect of RDV initiation timing on mortality is statistically nonsignificant at 5%.
- The coefficient for Hospital_duration (estimated at −0.0279) has a 95% CI of [−0.0521, −0.0065], indicating that longer hospital stays are associated with reduced mortality risk.
- The coefficient for FiO2Max (estimated at 0.0531) has a 95% CI of [0.0468, 0.0599], indicating a positive association between higher oxygen requirements and increased mortality risk.

The performance of the glmnet model is presented in Table 8. Both the glmnet and logistic regression models demonstrate strong performance across various metrics, although each has its strengths in different areas. The performance of the other models is shown in Table 9.

To further illustrate the trade-off between sensitivity and specificity across all models, we provide the ROC curves in Fig. 3. To compare the performance of the models in the context of the imbalanced dataset, we also computed the AUPRC. The glmnet model achieved an AUPRC of 0.803, while the logistic regression model showed a higher AUPRC of 0.841. For the other models, we obtained AUPRC values of 0.678, 0.570 and 0.534 for SVM, GBM and RF, respectively. These results further emphasize the superior reliability of logistic regression model in predicting the positive class, followed by the glmnet model.

Based on all the given information, each model showed substantial predictive power. However, the GLMs particularly distinguished themselves due to their interpretability. In GLMs, the coefficients directly reflect the impact of each predictor on the outcome, assuming all other variables remain constant, offering a balance of interpretability and flexibility. For example, a coefficient for Age indicates how each additional year increases or decreases the odds of mortality, which clinicians can directly use to assess risk. Therefore, even though the SVM with a radial basis function kernel displayed a high AUC, GLMs—specifically the logistic regression model—strike a balance between interpretability and flexibility, making them an appealing choice for statistical modeling where understanding the relationships between predictors and outcomes is vital. This interpretability is particularly crucial in clinical settings, where decisions must be transparent and easily understood by healthcare professionals.

3.5 Comparison of fuzzy modeling methods

Now, we implement and compared different fuzzy modeling methods, including FRBCS.W, FRBCS.CHI, FH.GBML and GFS.GCCL, using the same training dataset of Portuguese COVID-19 cases that was employed for non-fuzzy models. We evaluate the effectiveness of these fuzzy methods in predicting mortality among COVID-19 patients and we compare their performance against non-fuzzy models in terms of accuracy, sensitivity, specificity, and the interpretability of the results.

Each fuzzy modeling method was implemented using the frbs.learn function, with specific control parameters that are crucial for defining the behavior of the models. The control parameters include num.labels (number of fuzzy labels), type.mf (membership function type), type.tnorm (t-norm), type.snorm (s-norm), and type.implication.func (implication function). An example in algorithmic form detailing the procedure for one of these methods is presented in Algorithm 2 and illustrated in the flowchart of Fig. 4.

Algorithm 2: Fuzzy model generation and evaluation.

Input: Chosen method, training data, input data range, control parameters;

Output: Evaluation of the fuzzy model performance and identification of relevant variables.

1. Initialize the fuzzy model using the chosen method, training data, input data range, and control parameters.
2. Use the code: control = list (num.labels = 7, type.mf = “GAUSSIAN”, type.tnorm = “MIN”, type.snorm = “MAX”, type.implication.func = “ZADEH”).
3. Train the fuzzy model on the training data.
4. Predict outcomes for the test data utilizing the trained fuzzy model.
5. Display a confusion matrix to assess the model performance.
6. Summarize the results of the FRBCS model, including relevant variables.
7. Use the summary to determine which variables most impact the model accuracy.

The control parameters used in the fuzzy models, including the example shown in Algorithm 2, play a crucial role in defining the model behavior. Specifically, num.labels = 7 divides the data into seven fuzzy categories, providing a finer granularity compared to simpler models. For instance, the fuzzy sets might represent levels such as “very low”, “low”, “slightly below average”, “average”, “slightly above average”, “high”, and “very high”. Choice of the membership

TABLE 8. Performance metrics for glmnet model with 95% CIs using Portuguese COVID-19 data.

Metric	Estimate	Lower limit of the 95% CI	Upper limit of the 95% CI
Accuracy	0.924	0.880	0.962
Sensitivity	0.857	0.946	1.000
Specificity	0.938	0.480	0.848
AUC	0.892	0.793	0.974

CI: confidence interval; AUC: area under the curve.

TABLE 9. Performance metrics for SVM, GBM and RF models with 95% CIs using Portuguese COVID-19 data.

Model	Metric	Estimate	Lower limit of the 95% CI	Upper limit of the 95% CI
SVM				
	Accuracy	0.924	0.880	0.968
	Sensitivity	0.985	0.958	1.000
	Specificity	0.630	0.455	0.815
	AUC	0.952	0.885	0.993
GBM				
	Accuracy	0.905	0.842	0.943
	Sensitivity	0.947	0.875	0.967
	Specificity	0.704	0.550	0.900
	AUC	0.891	0.818	0.966
RF				
	Accuracy	0.886	0.835	0.930
	Sensitivity	0.939	0.894	0.977
	Specificity	0.630	0.438	0.818
	AUC	0.922	0.861	0.961

CI: confidence interval; SVM: support vector machine; GBM: gradient boosting machine; RF: random forest; AUC: area under the curve.

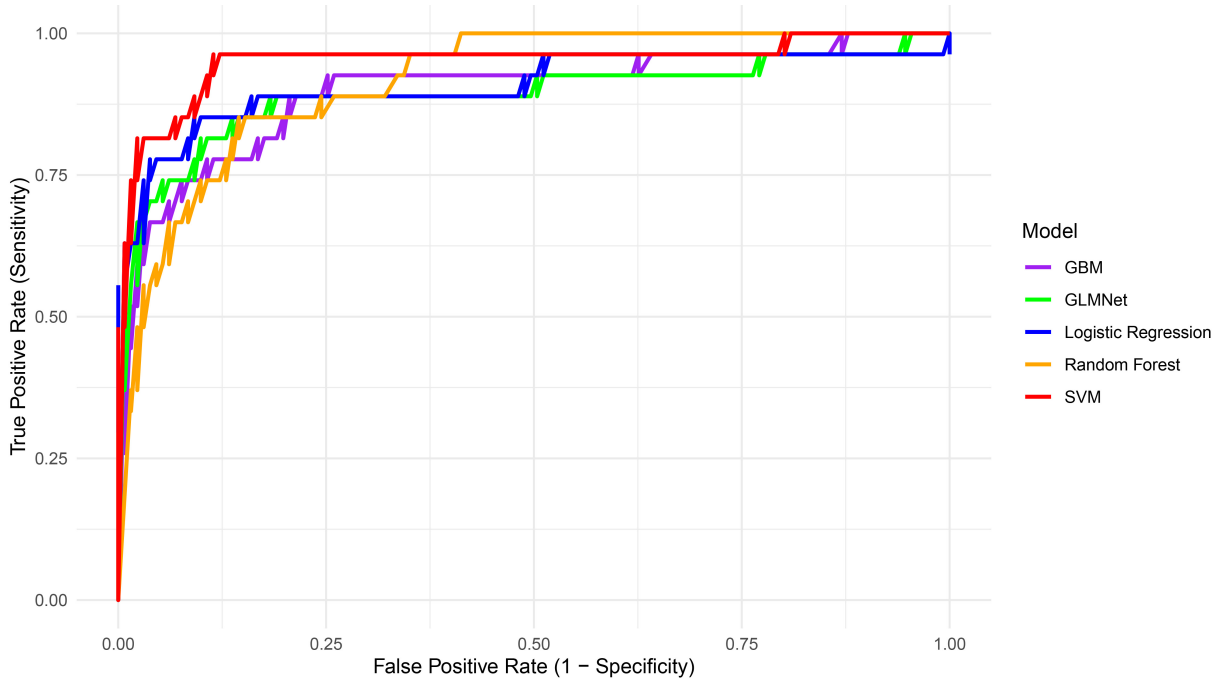

FIGURE 3. ROC curves for the GBM, GLMNet, logistic regression, RF and SVM models using Portuguese COVID-19 data. GBM: gradient boosting machine; SVM: support vector machines.

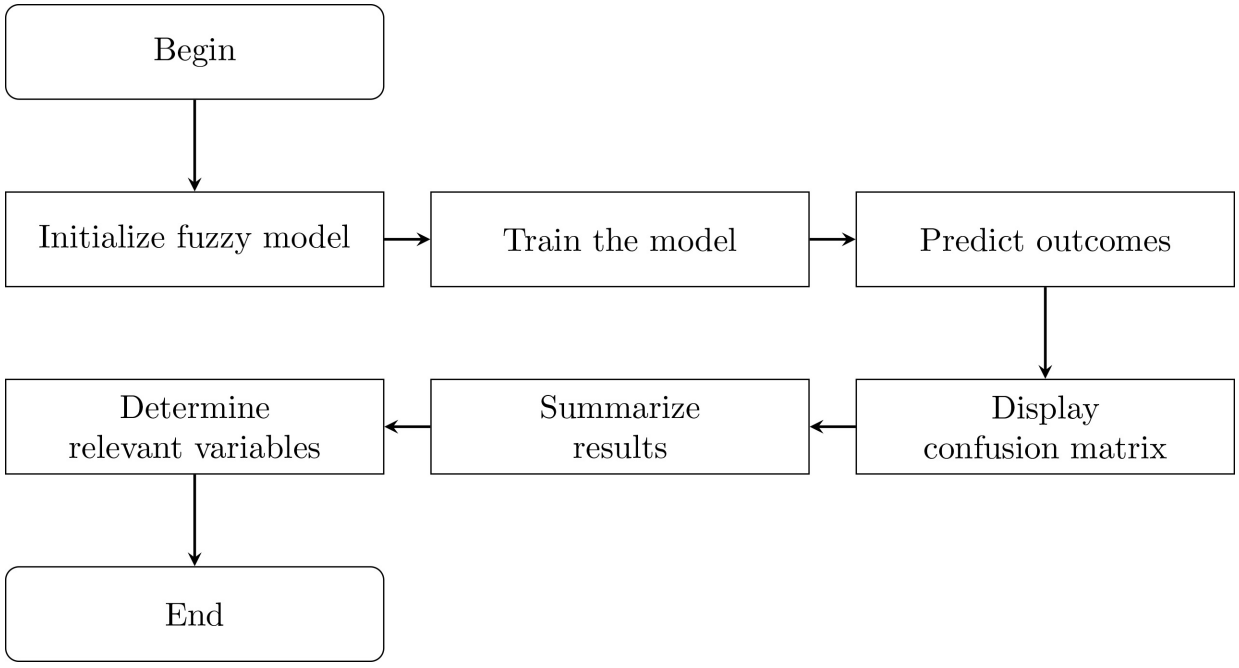

FIGURE 4. Flowchart of the fuzzy model generation and evaluation process.

function (type.mf = "GAUSSIAN") ensures smooth transitions between these levels, while the *t*-norm and s-norm (type.tnorm = "MIN", type.snorm = "MAX") define the logic operations for intersection and union within the fuzzy system. Lastly, type.implication.func = "ZADEH" specifies how implications are handled within the model, following Zadeh fuzzy logic principles. These control parameters are essential in tailoring the fuzzy models to reflect the complexity of the input data, especially when dealing with subtle differences in the dataset.

It is important to note that the control parameters may vary across different fuzzy methods to best suit each method characteristics and the specific application within the context of COVID-19 data. When predicting outcomes for new and test data using the generated fuzzy logic, the model applies a set of fuzzy rules to each input case. For example, a rule might be: IF Age is "high" AND FiO2Max is "high", THEN the outcome is 80% "death" and 20% "alive". Here, the fuzzy output reflects the degrees of membership to each category, representing the uncertainty in the classification. The set of fuzzy rules generated by the model is then used to classify the test data, where each input is evaluated according to these rules, resulting in a fuzzy output that indicates the degrees of membership to categories such as "death" and "alive".

To illustrate how the model classifies a new case, consider the following example. A new patient has the characteristics: Age classified as "very high" FiO2Max as "high", Time_Adm_Sym as "moderate", RDV_In_Sym as "early", and Hospital_duration as "short". The fuzzy model might apply a combination of rules, such as: IF Age is "very high" AND FiO2Max is "high" AND Time_Adm_Sym is "moderate" AND RDV_In_Sym is "early" AND Hospital_duration is "short", THEN the outcome is 85% "death" and 15% "alive".

In practice, the model may not have an exact rule that perfectly matches the new case. Instead, it evaluates the case based on the available rules that most closely align with the patient characteristics. The model aggregates the results of these rules, producing a fuzzy output that represents the degrees of membership to each category, such as "death" and "alive". For this example, the model might assign an 85% membership to the category "death" and 15% to the category "alive", reflecting the combined uncertainty and assessment of the patient risk based on the five variables.

If we need to decide, such as determining a crisp classification, the fuzzy output can be defuzzified. For instance, using the centroid method, we calculate a score based on the weighted average of the membership degrees, which then lead to a decision threshold (if the result is above 50%, classify as "death"). This allows the fuzzy model to handle uncertainty and provide predictions, which are then translated into actionable classifications.

After applying the fuzzy rules and obtaining a fuzzy output, the output can be defuzzified to a crisp value, such as "death" or "alive". In our example, defuzzification is performed using the centroid method, where the weighted average of the membership degrees is utilized to determine the definitive category. Defuzzification is necessary in our study to enable a direct comparison with non-fuzzy models in terms of accuracy, sensitivity, and specificity.

The accuracy, sensitivity, and specificity for each fuzzy method were obtained as follows:

- FRBCS.W—Accuracy = 0.818, sensitivity = 0.829, and specificity = 0.650.
- FRBCS.CHI—Accuracy = 0.857, sensitivity = 0.828, and specificity = 1.00.
- FH.GBML—Accuracy = 0.837, sensitivity = 0.829, and specificity = 0.692.
- GFS.GCCL—Accuracy = 0.745, sensitivity = 0.829, and specificity = 0.525.

From the comparison of the fuzzy methods used, it is clear that FRBCS.CHI had the highest accuracy (0.857), specificity

(1.0), and AUC (0.929). Both FRBCS.W and FH.GBML had similar accuracy (0.818 and 0.837, respectively) and sensitivity (0.829 for both), while GFS.GCCL had the lowest accuracy (0.745) and specificity (0.525). Based on these results, FRBCS.CHI shows the best overall performance, with high accuracy and specificity.

Regarding the results of the relevance of variables in the FRBCS.CHI model, FiO2Max emerges as the most relevant variable, accounting for approximately 44.42% of the predictive power of the model. This means that FiO2Max contributes the most to the model ability to make accurate predictions. Following FiO2Max, Age contributes 19.47%, RDV_In_Sym a 13.32%, Time_Adm_Sym a 13.20%, and Hospital_duration a 9.58%. These percentages reflect the relative contribution of each variable to the overall predictive accuracy of the model, as determined by the permutation feature importance analysis, which measures how much the model prediction performance decreases when each variable values are randomly permuted. The feature importance of predictor variables based on permutation method with Portuguese COVID-19 data is plotted in Fig. 5.

When comparing the fuzzy results and the logistic regression model, the consistency of the selected predictors in both models stands out. In the logistic regression, all the five variables are statistically significant at 5%. While fuzzy models provide insights into which variables are important, they do not inherently indicate the direction of the relationship (as whether an increase in a variable increases or decreases the probability of an outcome). This contrasts with GLMs, where coefficients directly indicate how changes in predictors influence outcomes, both in direction and magnitude. For example, a positive estimated coefficient for Age in a GLM indicates that as age increases, the estimated probability of death also increases.

3.6 Discussion on methodological choices

We next explain our choice of analytical methods and their alignment with the goal of precisely predicting COVID-19 mortality.

In our dataset, the relationships between predictor variables and clinical outcomes may not be strictly linear or binary. While some relationships could be approximately linear or binary under certain conditions, others might involve more complex, non-linear, or gradual transitions. Non-linear relationships can be effectively modeled using various machine learning techniques, such as GBM, GLMs, RF and SVM, with non-linear terms. However, conventional models may struggle with data that lack clear boundaries. For example, a slight change in FiO2Max or Age may not immediately alter the estimated survival probability but could gradually shift the associated risk. Fuzzy logic models are well-suited for such cases, as they allow for the modeling of such uncertainties by assigning degrees of membership to different categories, such as “high chance of survival” or “moderate risk of death”. The fuzzy models are particularly valuable in clinical decision-making, where it is more informative to express that a patient has, for example, a 70% chance of recovery and a 30% risk

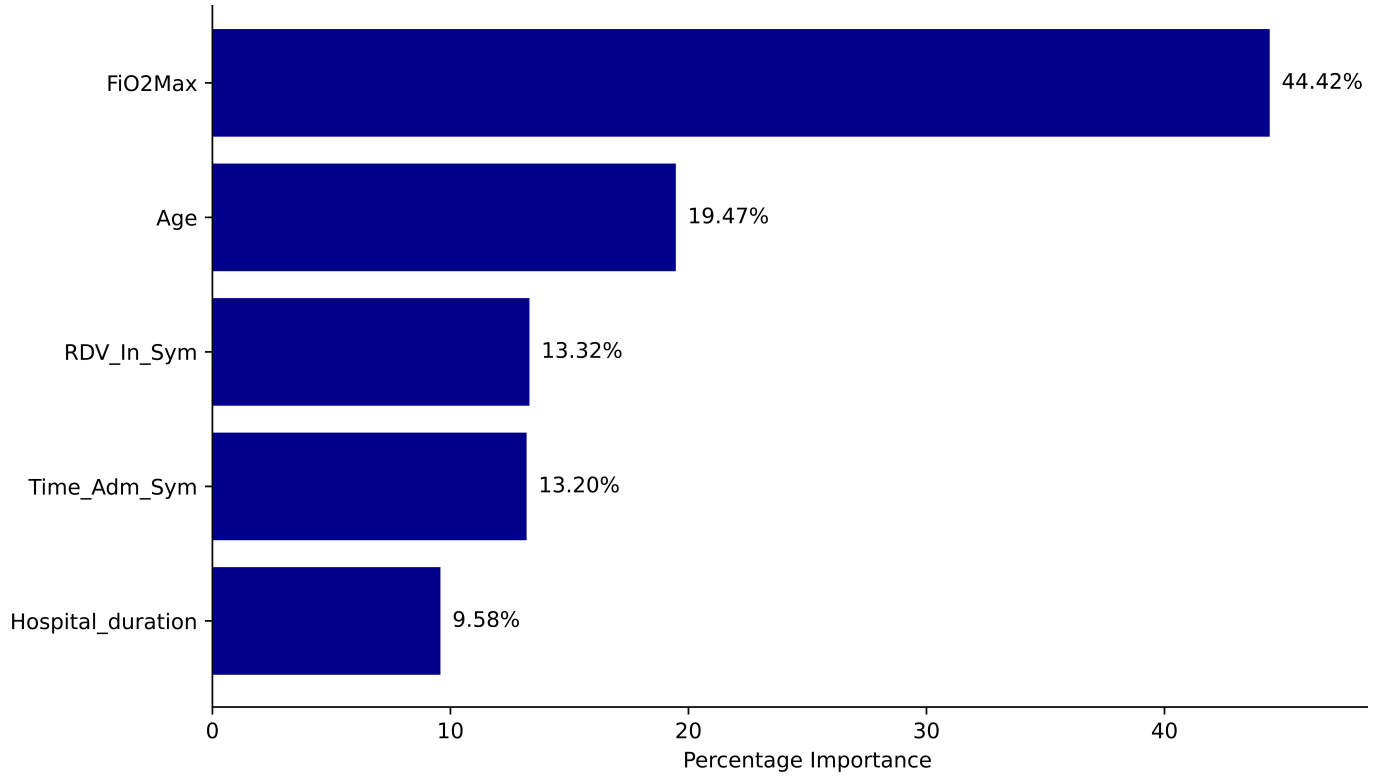

FIGURE 5. Bar plot of the feature importance of predictor variables based on permutation method with Portuguese COVID-19 data. FiO2Max: maximum fraction of inspired oxygen required during hospitalization; Hospital_duration: length of stay in days; OR: odds ratio; RDV_In_Sym: time from symptom onset to RDV start in days; Time_Adm_Sym: time from symptom onset to hospital admission in days.

of deterioration, rather than categorizing the patient as simply "survival" or "non-survival".

The fuzzy models used in this study—FRBCS.W, FRBCS.CHI, FH.GBML and GFS.GCCL—are specifically designed to handle these inherent uncertainties in COVID-19 clinical data. Such models generate fuzzy rules based on the input data, where each rule encapsulates a relationship between predictor variables and the outcome. Additionally, each of these fuzzy models applies different strategies for rule generation, optimization, and handling of uncertainties, which distinguishes them from one another. For instance, FRBCS.W emphasizes rule weighting, FRBCS.CHI focuses on generating rules using the chi-square method, FH.GBML applies genetic-based optimization techniques to enhance rule quality, and GFS.GCCL uses genetic algorithms with cooperative coevolutionary learning to optimize the fuzzy rules. These distinct models allow each of them to handle uncertainties and capture complex relationships in COVID-19 data differently, providing a comprehensive analysis of the factors influencing patient outcomes.

4. Conclusions

This study explored several predictive models for COVID-19 mortality, focusing on non-invasive predictors and the impact of remdesivir treatment. Logistic regression demonstrated high predictive performance (AUC = 0.908), with the glmnet model following closely (AUC = 0.892). Ensemble methods, including random forests (AUC = 0.922) and gradient boosting (AUC = 0.891), offered comparable results but did not surpass generalized linear models, while support vector machines reached the highest AUC (0.952) but with lower specificity. Fuzzy rule-based models provided valuable insights by managing uncertainties, although they did not surpass traditional methods in accuracy. A key finding of our study was the effect of timely remdesivir administration, with logistic regression indicating that delayed treatment increases mortality risk. However, this effect was not confirmed by the glmnet model, highlighting the complexity of interpreting remdesivir role in patient outcomes. Despite our promising results, the study has limitations. The retrospective nature of the dataset may introduce biases, and while the models offered valuable insights, they may not fully capture all nonlinear relationships within the data. Further validation with diverse datasets is required to improve the generalizability of our findings. Additionally, future work should explore interactions between remdesivir and other clinical variables to better understand its role in mortality outcomes.

Overall, this study showed the value of combining non-invasive clinical predictors with statistical, machine learning, and fuzzy techniques to predict mortality of COVID-19 patients. This multidisciplinary approach provided a robust framework for supporting clinical decision-making that can address broader healthcare challenges [55].

AVAILABILITY OF DATA AND MATERIALS

The data used to support the findings of this study are available upon request.

AUTHOR CONTRIBUTIONS

CC and VL—designed the study, managed the data, and created the models and codes, performed the data analysis. CC, VL, PC and MAA—drafted and wrote the final version the article, revised the article critically for important intellectual contents. CC, VL and PC—performed a clinical development and validated the analysis and results. All authors contributed to editorial changes in the article. All authors read and approved the final article.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was conducted under international ethical standards for medical research involving human subjects, as outlined in the Declaration of Helsinki. The study protocol (Ref: 0912021) received approval from the local ethics committee (Hospital da Senhora da Oliveira—Guimarães, EPE, Portugal). Given the retrospective nature of the data collection, a waiver for informed consent was requested and granted, in accordance with the applicable legal provisions (Article 6, number 2 of the Clinical Research Law—21/2014). Data collection and handling were carried out under strict confidentiality measures. An independent entity was engaged for data cleaning to ensure the accuracy, completeness, and consistency of the data, which were pseudonymized, with the key kept only by the principal investigator. The data were further processed in accordance with the General Data Protection Regulation and the hospital privacy policies, ensuring minimal risk to the participants. All protocols were performed in compliance with the relevant guidelines and regulations to uphold the highest ethical standards.

ACKNOWLEDGMENT

The authors acknowledge the support provided by their respective institutions. The authors also thank the reviewers for their constructive feedback, which contributed to improving the quality of the article.

FUNDING

This work was supported by Portuguese funds through the CMAT—Research Centre of Mathematics of the University of Minho, Portugal, within projects UIDB/00013/2020 (https://doi.org/10.54499/UIDB/00013/2020) and UIDP/00013/2020 (https://doi.org/10.54499/UIDP/00013/2020) (Cecilia Castro). This research was also partially supported by the Vice-rectorate for Research, Creation, and Innovation (VINCI) of the Pontificia Universidad Católica de Valparaíso (PUCV), Chile, under grants VINCI 039.470/2024 (regular research), VINCI 039.493/2024 (interdisciplinary associative research), VINCI 039.309/2024 (PUCV centenary), and FONDECYT 1200525 (Víctor Leiva), from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge and Innovation.

CONFLICT OF INTEREST

The authors declare no conflict of interest. Víctor Leiva is serving as one of the Editorial Board members of this journal. We declare that Víctor Leiva had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to ASA.

REFERENCES

[1] Leão J, Leiva V, Saulo H, Tomazella V. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. Statistics in Medicine. 2018; 37: 4421–4440.

[2] Liu Y, Wang J, Leiva V, Tapia A, Tan W, Liu S. Robust autoregressive modeling and its diagnostic analytics with a COVID-19-related application. Journal of Applied Statistics. 2024; 51: 1318–1343.

[3] Akkilic AN, Sabir Z, Raja MAZ, Bulut H. Numerical treatment on the new fractional-order SIDARTHE COVID-19 pandemic differential model via neural networks. The European Physical Journal Plus. 2022; 137: 334.

[4] Sabir Z, Raja MAZ, Alhazmi SE, Gupta M, Arbi A, Baba IA. Applications of artificial neural network to solve the nonlinear COVID-19 mathematical model based on the dynamics of SIQ. Journal of Taibah University for Science. 2022; 16: 874–884.

[5] Sabir Z, Alnahdi AS, Jeelani MB, Abdelkawy MA, Raja MAZ, Baleanu D, *et al*. Numerical computational heuristic through morlet wavelet neural network for solving the dynamics of nonlinear SITR COVID-19. Computer Modeling in Engineering and Sciences. 2022; 131: 763–785.

[6] Rahman MZU, Akbar MA, Leiva V, Martin-Barreiro C, Imran M, Riaz MT, *et al*. An IoT-fuzzy intelligent approach for holistic management of COVID-19 patients. Heliyon. 2024; 10: e22454.

[7] Leiva V, Alcudia E, Montano A, Castro C. An epidemiological analysis for assessing and evaluating COVID-19 based on data analytics in Latin American countries. Biology. 2023; 12: 887.

[8] Ospina R, Gondim JAM, Leiva V, Castro C. An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: methodology and case study in Brazil. Mathematics. 2023; 11: 3069.

[9] Ospina R, Ferreira AGO, de Oliveira HM, Leiva V, Castro C. On the use of machine learning techniques and non-invasive indicators for classifying and predicting cardiac disorders. Biomedicines. 2023; 11: 2604.

[10] Botmart T, Sabir Z, Javeed S, Sandoval Núñez RA, Wajaree Weera, Ali MR, *et al*. Artificial neural network-based heuristic to solve COVID-19 model including government strategies and individual responses. Informatics in Medicine Unlocked. 2022; 32: 101028.

[11] Redruello-Guerrero P, Jimenez-Gutierrez C, Ramos-Bossini AL, Jiménez-Gutiérrez PM, Rivera-Izquierdo M, Sánchez JB. Artificial intelligence for the triage of COVID-19 patients at the emergency department: a systematic review. Signa Vitae. 2022; 18: 17–26.

[12] Al Aseri Z, Al-Tawfiq JA, Alnakhli M, AlNooh A, Alnassar A, Alkhalid S, *et al*. Epidemiological characteristics and initial clinical presentation of patients with laboratory-confirmed MERS-CoV infection in an emergency department. Signa Vitae. 2022; 18: 68–74.

[13] Perez-Lillo N, Lagos-Alvarez B, Muñoz-Gutierrez J, Figueroa-Zúñiga J, Leiva V. A statistical analysis for the epidemiological surveillance of COVID-19 in Chile. Signa Vitae. 2022; 18: 19–30.

[14] Ospina R, Leite A, Ferraz C, Magalhaes A, Leiva V. Data-driven tools for assessing and combating COVID-19 outbreaks based on analytics and statistical methods in Brazil. Signa Vitae. 2022; 18: 18–32.

[15] Sardar I, Akbar MA, Leiva V, Alsanad A, Mishra P. Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: methodology, evaluation, and case study in SAARC countries. Stochastic Environmental Research and Risk Assessment. 2023; 37: 345–359.

[16] Martin-Barreiro C, Cabezas X, Leiva V, Ramos de Santis P, Ramirez-Figueroa JA, Delgado E. Statistical characterization of vaccinated cases and deaths due to COVID-19: methodology and case study in South America. AIMS Mathematics. 2023; 8: 22693–22713.

[17] Groves-Kirkby N, Wakeman E, Patel S, Hinch R, Poot T, Pearson J, *et al*. Large-scale calibration and simulation of COVID-19 epidemiologic scenarios to support healthcare planning. Epidemics. 2023; 42: 100662.

[18] Soulsby CR, Hutchison C, Gardner J, Hart R, Sim MAB, Millar JE. Socio-economic deprivation and the risk of death after ICU admission with COVID-19: the poor relation. Journal of the Intensive Care Society. 2023; 24: 44–45.

[19] Zabor EC, Reddy CA, Tendulkar RD, Patil S. Logistic regression in clinical studies. International Journal of Radiation Oncology, Biology, Physics. 2022; 112: 271–277.

[20] Wallisch C, Bach P, Hafermann L, Klein N, Sauerbrei W, Steyerberg EW, *et al*. Review of guidance papers on regression modeling in statistical series of medical journals. PLOS ONE. 2022; 17: e0262918.

[21] Nusinovici S, Tham YC, Yan MY, Ting DS, Li J, Sabanayagam C, *et al*. Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of Clinical Epidemiology. 2020; 122: 56–69.

[22] D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, *et al*. General cardiovascular risk profile for use in primary care: the Framingham heart study. Circulation. 2008; 117: 743–753.

[23] Alzahrani T, Nguyen T, Ryan A, Dwairy A, McCaffrey J, Yunus R, *et al*. Cardiovascular disease risk factors and myocardial infarction in the transgender population. Circulation. 2019; 12: e005597.

[24] Zadeh LA. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man, and Cybernetics. 1973; 3: 28–44.

[25] Tserkovny A. A t-norm fuzzy logic for approximate reasoning. Journal of Software Engineering and Applications. 2017; 10: 639–662.

[26] Czmil A. Comparative study of fuzzy rule-based classifiers for medical applications. Sensors. 2023; 23: 992–1012.

[27] Mohammadpour RA, Abedi SM, Bagheri S, Ghaemian A. Fuzzy rule-based classification system for assessing coronary artery disease. Computational and Mathematical Methods in Medicine. 2015; 2015: 64867.

[28] Gu X, Zhang C, Ni T. Feature selection and rule generation integrated learning for Takagi-Sugeno-Kang fuzzy system and its application in medical data classification. IEEE Access. 2019; 7: 1501–1508.

[29] Karthikeyan R, Geetha P, Ramaraj E. 'Rule-based system for better prediction of diabetes', 2019 3rd International Conference on Computing and Communications Technologies, 21–22 February 2019, Chennai, India. IEEE: New York, USA. 2019.

[30] Singh N, Singh P. 'Medical diagnosis of coronary artery disease using fuzzy rule-based classification approach', Advances in Biomedical Engineering and Technology: Select Proceedings of ICBEST 2018. 09–10 February 2018, Raipur, Chhattisgarh, India. Springer: Singapore. 2021.

[31] Hossain S, Sarma D, Chakma RJ, Alam W, Hoque MM, Sarker IH. 'A rule-based expert system to assess coronary artery disease under uncertainty', Proceedings of the First International Conference on Computing Science, Communication and Security. COMS2 2020. Gujarat, India, 26–27 March 2020. Springer, Singapore. 2020.

[32] Steimann F. On the use and usefulness of fuzzy sets in medical AI. Artificial Intelligence in Medicine. 2001; 21: 131–137.

[33] Mousavi SM, Abdullah S, Niaki STA, Banihashemi S. An intelligent hybrid classification algorithm integrating fuzzy rule-based extraction and harmony search optimization: medical diagnosis applications. Knowledge-Based Systems. 2021; 220: 106943.

[34] Rahman MZ, Akbar MA, Leiva V, Tahir A, Riaz MT, Martin-Barreiro C. An intelligent health monitoring and diagnosis system based on the internet of things and fuzzy logic for cardiac arrhythmia COVID-19 patients. Computers in Biology and Medicine. 2023; 154: 106583–106597.

[35] Saranya A, Rajeswari J. Enhanced prediction of student dropouts using fuzzy inference system and logistic regression. ICTACT Journal on Soft Computing. 2016; 6: 1157–1162.

[36] Zeinalnezhad M, Chofreh AG, Goni FA, Klemes JJ. Air pollution prediction using semi-experimental regression model and adaptive neuro-fuzzy inference system. Journal of Cleaner Production. 2020; 261: 121218.

[37] Aggarwal A, Chakradar M, Bhatia MS, Kumar M, Stephan T, Gupta SK, *et al*. COVID-19 risk prediction for diabetic patients using fuzzy inference system and machine learning approaches. Journal of Healthcare

Engineering. 2022; 2022: 4096950.
[38] Vivekanandhan V, Sakthivel S, Manikandan M. Adaptive neuro fuzzy inference system to enhance the classification performance in smart irrigation system. Computational Intelligence. 2022; 38: 308–322.
[39] Tanaka H, Uejima S, Sai K. Linear regression analysis with fuzzy model. IEEE Transactions on Systems, Man, and Cybernetics. 1982; 12: 903–907.
[40] Palacios CA, Reyes-Suarez JA, Bearzotti LA, Leiva V, Marchant C. Knowledge discovery for higher education student retention based on data mining: machine learning algorithms and case study in Chile. Entropy. 2021; 23: 485.
[41] Butt NA, Awais MM, Abbas Q. Improved diagnostic accuracy in dependent personality disorders: a comparative study of neural architectures and hybrid approaches on functional magnetic resonance imaging data. Journal of Medical Imaging and Health Informatics. 2019; 9: 697–705.
[42] Singh D, Upadhyay R, Pannu HS, Leray D. Development of an adaptive neuro fuzzy inference system based vehicular traffic noise prediction model. Journal of Ambient Intelligence and Humanized Computing. 2020; 12: 2685–2701.
[43] Vapnik V. The nature of statistical learning theory. 2nd edn. Springer: New York, NY, USA. 2000.
[44] Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20: 273–297.
[45] Breiman L. Random forests. Machine Learning. 2001; 45: 5–32.
[46] Breiman L, Cutler A. Random forests—machine learning. 2004. Available at: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (Accessed: 28 September 2024).
[47] Efron B, Tibshirani RJ. An introduction to the Bootstrap. 1st edn. CRC Press: Boca Raton, FL, USA. 1994.
[48] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of Statistics. 2001; 29: 1189–1232.
[49] Zadeh LA. Fuzzy sets. Information and Control. 1965; 8: 338–353.
[50] Sugeno M, Kang GT. Structure identification of fuzzy model. Fuzzy Sets and Systems. 1988; 28: 15–33.
[51] Lam HK. A review on stability analysis of continuous-time fuzzy-model-based control systems: From membership-function-independent to membership-function-dependent analysis. Engineering Applications of Artificial Intelligence. 2018; 67: 390–408.
[52] Riza LS, Bergmeir C, Herrera F, Benítez JM. frbs: fuzzy rule-based systems for classification and regression in R. Journal of Statistical Software. 2015; 65: 1–30.
[53] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 1st edn. Springer: New York, NY, USA. 2009.
[54] Kuhn M. Building predictive models in R using the caret package. Journal of Statistical Software. 2008; 28: 1–26.
[55] Sánchez L, Leiva V, Galea M, Saulo H. Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. Applied Stochastic Models in Business and Industry. 2021; 37: 53–73.

How to cite this article: Cecilia Castro, Víctor Leiva, Pedro Cunha, Muhammad Azeem Akbar. Predicting COVID-19 mortality using statistical, machine learning and fuzzy classification methods: insights from a Portuguese cohort study. Signa Vitae. 2024; 20(12): 10-27. doi: 10.22514/sv.2024.152.