Integration of Epidemiologic, Socioeconomic, and Sociodemographic Indicators to Predict Early COVID-19 In-Hospital Outcomes

Hetielle Matos[1], Artur Brenner Schmitt[1], Felipe André Zeiser[1], Cristiano André da Costa[1], Gabriel de Oliveira Ramos[1]

[1]Programa de Pós-Graduação em Computação Aplicada
Universidade do Vale do Rio dos Sinos
São Leopoldo – RS – Brasil

Abstract. *The COVID-19 pandemic is an unprecedented challenge for healthcare systems around the world. In Brazil, the COVID-19 pandemic affected the population differently. Sociodemographic and socioeconomic characteristics were important indicators of early access and quality of the health system. In this way, we combine epidemiological, socioeconomic, and sociodemographic data to predict in-hospital outcomes of COVID-19. The proposed approach utilizes models such as Random Forest, XGBoost, TabNet, and CatBoost, and employs Bayesian optimization for automatic hyperparameter selection. The results demonstrate that all models exhibit a relatively higher ability to correctly identify hospital discharge outcomes than mortality cases. However, XGBoost showed the best result, with a Precision of 0.72, Recall of 0.74, F1-score of 0.64, Accuracy of 0.74, and AUC of 0.83. The quantitative and qualitative results demonstrate that our method can effectively suggest high-quality in-hospital outcomes and demonstrate the possibility of using our methodology as a tool to assist healthcare professionals.*

1. Introduction

The COVID-19, caused by the novel coronavirus SARS-CoV-2, is believed to have originated in Wuhan, China, in late 2019 [Andersen et al. 2020]. It quickly spread across the globe, evolving into a pandemic that has affected millions of people. The zoonotic origin of the virus and its subsequent transmission to humans highlight the interconnectedness of our global society and the need for a comprehensive understanding of the pandemic's impact [Tsiotas and Tselios 2022]. The COVID-19 pandemic has presented an unprecedented challenge to healthcare systems worldwide, with varying degrees of severity and outcomes observed among infected individuals [Zeiser et al. 2022].

Understanding the factors contributing to the diverse outcomes experienced by COVID-19 patients is important to guiding effective public health interventions and improving healthcare resource allocation [Tsiotas and Tselios 2022]. While several studies have explored the clinical and biological factors associated with COVID-19 outcomes, the influence of socioeconomic and demographic indicators on in-hospitality outcomes still needs to be explored. Investigating the socioeconomic and demographic indicators that influence COVID-19 outcomes is crucial for identifying vulnerable populations, developing targeted interventions, and mitigating the effects of the virus [Zeiser et al. 2022].

In this way, we aim to investigate the impact of epidemiologic, socioeconomic, and sociodemographic indicators on machine learning models to predict early COVID-19 outcomes in hospitalized patients in Brazil from 2020 to 2023. Understanding the influence of socioeconomic and demographic indicators on hospital outcomes can provide important information for patient healthcare strategies for the COVID-19 disease. In addition, by investigating these indicators, we intend to clarify the underlying factors contributing to differential results, thus facilitating the development of targeted government interventions and strategies to contain pandemics such as COVID-19.

Importantly, while the COVID-19 pandemic may no longer be classified as a global pandemic, socioeconomic and demographic effects study remains crucial for effectively managing future crises or potential new pandemics [Zhang et al. 2023]. The knowledge gained from investigating these indicators can enhance preparedness and response efforts, ensuring that healthcare systems are better equipped to address the challenges posed by future public health emergencies [Betthäuser et al. 2023].

The remainder of this paper is organized as follows. Section 2 presents the most significant related works to define the present study. Next, section 3 presents the methodology of the work. The section 4 details the results and discussion. Finally, Section 5 presents the conclusions of the work.

2. Related Work

Different studies throughout the pandemic demonstrated the relationship between socioeconomic and demographic indicators and COVID-19 outcomes [Docherty et al. 2020, Zeiser et al. 2022, Barough et al. 2023, Cribari-Neto 2023]. These studies explored different indicators and aspects of the pandemic, providing valuable insights into the disparities and inequalities observed among affected populations. However, most machine learning models in the current literature do not integrate epidemiological, socioeconomic, and sociodemographic factors into a single analysis for Brazil. This gap is related to the difficulty in obtaining public data that qualifies the Brazilian cities' scenario.

Studies to understand the dynamics and epidemiological and demographic factors began in the first months of the COVID-19 pandemic. The study conducted by [De Souza et al. 2021] focusing on COVID-19 patients in the state of Espirito Santo, Brazil, employed a range of machine learning algorithms to predict patient survival based on a dataset of 13,690. The authors achieved promising results, with the best-performing model exhibiting an area under the curve (AUC) of 0.92. However, the models were based on initial cases, which, due to a lack of knowledge about the virus, caused higher mortality rates, facilitating the learning process of the model.

Sociodemographic and socioeconomic factors are important factors in determining the outcome of COVID-19, especially in individuals with lower-income backgrounds and with lower educational attainment face higher risks of severe illness and worse COVID-19 outcomes [Baqui et al. 2021]. Similarly, a study by [Green et al. 2021] highlighted the association between limited access to healthcare resources, such as testing and treatment facilities, and adverse outcomes in underserved communities.

In addition, most studies are based on information that characterizes the behavior of the disease during hospitalization [De Souza et al. 2021, Figuerêdo et al. 2021]. This information involves the use of invasive respiration, hospitalization in the Intensive Care Unit (ICU), and X-ray findings [Figuerêdo et al. 2021]. This information helps identify more severe cases and facilitates the classification by machine learning models. However, in a real scenario, these data will only be fully available at the end of the patient's treatment. Furthermore, in this way, they do not add a supplemental decision-making measure for health teams.

While the existing literature has provided valuable insights into the influence of socioeconomic and demographic indicators on COVID-19 outcomes, there remains a need for further investigation. Many studies have primarily focused on individual factors in isolation without comprehensively exploring the combined effects of multiple indicators [De Souza et al. 2021, Figuerêdo et al. 2021]. Additionally, some studies have been conducted in specific regions, limiting the generalizability of the findings [De Souza et al. 2021]. By addressing these gaps, our study aims to provide a more comprehensive understanding of the complex relationship between socioeconomic and demographic indicators and COVID-19 patients' in-hospitality outcomes.

In summary, previous research has highlighted the significance of socioeconomic and demographic indicators in shaping COVID-19 outcomes. The influence of socioeconomic indicators, including income, education, and healthcare access, and demographic indicators, such as age, gender, and ethnicity, has been extensively investigated. However, there is a need for further exploration of these factors in combination and across diverse populations. By building upon the existing literature, our study aims to contribute to understanding how socioeconomic and demographic indicators collectively impact the in-hospitality outcomes of COVID-19 patients, providing crucial insights for healthcare interventions and policy development.

3. Material and Methods

An overview of the methodology used in this work is presented in Fig. 1. Our methodology can be divided into four main steps: pre-processing, models' optimization, training, and testing. Pre-processing consists of treating missing data, data format and normalization (Section 3.2). Next, we optimized the models' hyperparameters with Bayesian Optimization (Section 3.3). With the best hyperparameters configuration, we describe in Section 3.4 the model's training process.

3.1. Datasets

We carried out a retrospective collection of publicly recorded data correlated with the COVID-19 pandemic that could complement the Brazilian population's epidemiological, socioeconomic, and sociodemographic characteristics. We used four datasets: Severe Acute Respiratory Syndrome Database (SARS), National Immunization Program Information System (SI-PNI), TABNET Hospital Beds, and the 2010 Census. PNI and TABNET Hospital Beds the collection period comprises February 25, 2020 (first case of COVID-19 reported in Brazil [Zeiser et al. 2022]) and May 3, 2023 (date of start of the study). The SRAG records any case requiring hospitalization or death related to COVID-19. The SI-PNI registers all vaccines applied in Brazil. Given the variations in recommended dose amounts during the COVID-19 pandemic, we considered that a person was fully vaccinated with the minimum vaccination course, two doses for AstraZeneca, Pfizer, and Coronavac, and one dose for Janssen.

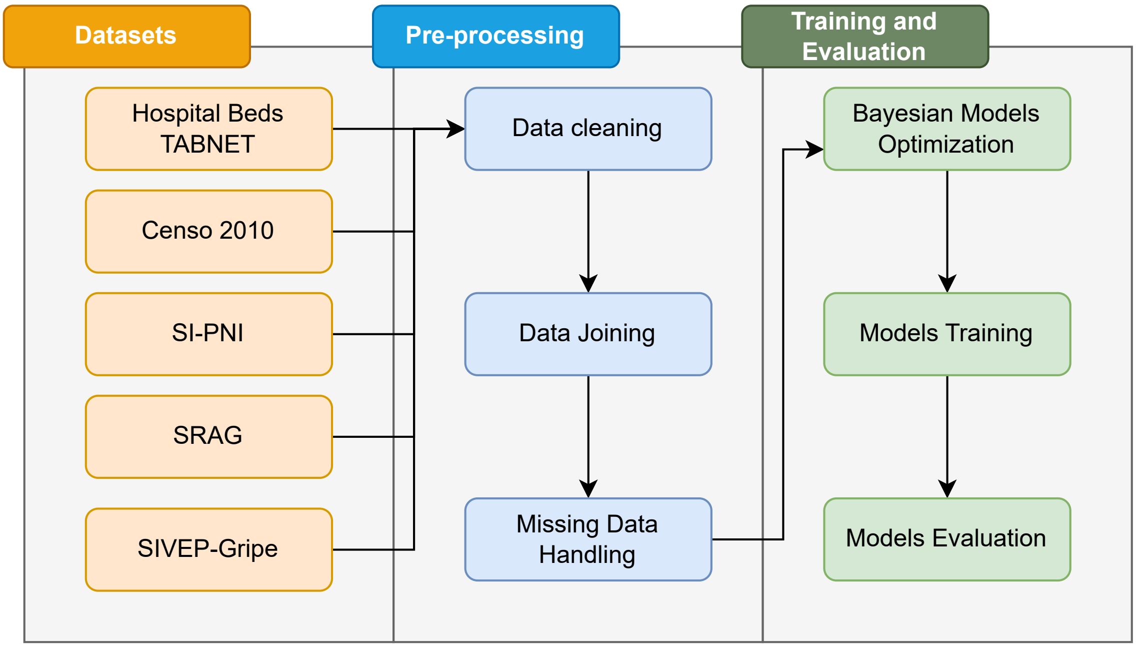

Figure 1. Summary of the proposed methodology.

3.2. Pre-processing

The SRAG bases are divided by annual files. In this way, we concatenated the files for the four years of analysis. At first, we removed records of patients not diagnosed with COVID-19, who were not hospitalized and had no known outcome. After the concatenation, we process dates and categorize string data such as cities and gender. To reduce the dimensionality of the model, we counted the number of symptoms and comorbidities of each patient registered in the SARS. Patients were also stratified into seven age groups: 0 to 19, 20 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, and +80. These age groups align with similar studies in the literature [Zeiser et al. 2022]. In addition, we created a column representing the count of epidemiological weeks across our study range.

Given the nature of the COVID-19 pandemic, contextual information is needed to characterize epidemiological severity over time. One of the pieces of information that significantly reduces the probability of death for patients is vaccination. As the current Brazilian public datasets do not allow for interoperability between them, it is impossible to know whether a patient was vaccinated. Therefore, we calculated a proportional vaccination rate per municipality in Brazil per epidemiological week. This calculation is performed by accumulating the number of people with a complete vaccination cycle per municipality and dividing it by the city's population according to the 2010 Census.

This same calculation is performed for the proportion of hospital beds available in each municipality in Brazil over the epidemiological weeks. Another important factor in measuring early access to the health system is the distance from the patient to the hospital unit. In this way, we calculated the distance from the patient's city of origin to the city of the hospital institution. We performed this calculation using the distance between the geographic coordinates of the municipalities.

Finally, to compose the socioeconomic and sociodemographic data, we considered the information from the 2010 Census. The columns, descriptions, and origins of the data used in our study are presented in Table 1. After pre-processing each dataset, we performed a left join of the datasets, considering the IBGE code of the municipality and the epidemiological week as the SRAG connection keys for the vaccination and bed data. For the 2010 Census data, only the municipality's IBGE code was considered. We filled in the missing data for sex, race, education, and area of residence by a default value defined empirically as 9. Then, we removed records that did not have a city of residence. The total set of cases was 1,180,158 records. Finally, the final dataset was divided into two random sets of training (80%) and test (20%).

Column	**Description**	**Source**
SEM_EPI	Case notification epidemiological week	SRAG
MUN_NOT	Patient's city of hospitalization	SRAG
MUN_RES	Patient's city of residence	SRAG
CS_RACA	Patient's race	SRAG
CS_ESCOLA	Patient's education	SRAG
CS_ZONA	Patient's area of residence	SRAG
QNT_COMORBIDADES	Number of patient comorbidities	SRAG
QNT_SINTOMAS	Number of patient symptoms	SRAG
UNI_NOT	Hospital code	SRAG
FAIXA_ETARIA	Patient's age group	SRAG
DISTANCIA	Distance between city of residence and city of hospitalization	SRAG
CASOS	Number of cases in the epidemiological week	SRAG
ESP_VIDA	Life Expectancy in the City of Residence	CENSO2010
SOBRE60	Probability of survival up to 60 years	CENSO2010
GINI	Gini index	CENSO2010
PIND	Proportion of extremely poor	CENSO2010
RDPC	Average per capita income	CENSO2010
THEIL	Theil index	CENSO2010
IDHM	Municipal Human Development Index	CENSO2010
PESORUR	Rural population in the city of residence	CENSO2010
PESOURB	Urban population in the city of residence	CENSO2010
LEITOS	Number of hospital beds in the hospitalization municipality	LEITOS
TIPO_UNI	Public or private hospital	LEITOS
LOTACAO	Proportion of cases per available hospital beds	SRAG/LEITOS
PERC_VAC	Percentage of the population fully vaccinated	SI-PNI

Table 1. Data description and source used for training machine learning models.

3.3. Bayesian optimization

We considered the following models for developing the in-hospital outcome prediction model: Random Forest, XGBoost, TabNet [Arik and Pfister 2021] and CatBoost [Dorogush et al. 2018]. We chose to optimize the hyperparameters of each model using the Bayesian optimization strategy. In Table 2, we present the search spaces for optimizing the algorithms. We used only the training set to search for the best hyperparameters of the models.

3.4. Training and testing

We used the hyperparameters obtained by Bayesian optimization to train the models. After training, we generate the confusion matrix of the models based on the test set. From the confusion matrix, we generated the metrics of Precision, Recall, F1-score, Accuracy, and Area Under the Receiver Operating Characteristic Curve (AUC).

Model	Parameters
Random Forest	’criterion’: hp.choice(’criterion’, [’entropy’, ’gini’]), ’max_depth’: scope.int(hp.quniform(’max_depth’,10,40,1)), ’max_features’: hp.choice(’max_features’, [’auto’, ’sqrt’,’log2’, None]), ’min_samples_leaf’: hp.uniform(’min_samples_leaf’, 0, 0.5), ’min_samples_split’ : hp.uniform (’min_samples_split’, 0, 1), ’n_estimators’ : hp.choice(’n_estimators’, [10, 50, 300, 750, 1200,1300,1500]), ’min_weight_fraction_leaf’: hp.uniform(’min_weight_fraction_leaf’, 0, 0.5),
XGBoost	’n_estimators’:scope.int(hp.quniform(’n_estimators’,200,1000,1)), ’learning_rate’:hp.loguniform(’learning_rate’,-7,0), ’max_depth’: scope.int(hp.quniform(’max_depth’,10,20,1)), ’subsample’: hp.uniform(’subsample’,0.2,1), ’colsample_bytree’: hp.uniform(’colsample_bytree’, 0.2,1), ’colsample_bylevel’: hp.uniform(’colsample_bylevel’, 0.2,1), ’min_child_weight’: hp.loguniform(’min_child_weight’,-1,7), ’alpha’: hp.choice(’alpha’, [0.0, hp.loguniform(’c1’,-10, 10)]), ’lambda’: hp.choice(’lambda’, [0.0, hp.loguniform(’c2’,-10, 10)]), ’gamma’: hp.choice(’gamma’, [0.0, hp.loguniform(’c3’,-10, 10)]),
TabNet	’optimizer_params’: hp.loguniform(’optimizer_params’,-7,0), ’output_dim’: scope.int(hp.quniform(’max_size’,20,60,1)), ’n_steps’: scope.int(hp.quniform(’n_steps’,1,8,1)), ’epsilon’: hp.uniform(’epsilon’,-5,0), ’batch_size’: hp.choice(’batch_size’, [512,1024,2048,4096,8192]
CatBoost	’optimizer_params’: hp.loguniform(’optimizer_params’,-7,0), ’output_dim’: scope.int(hp.quniform(’max_size’,20,60,1)), ’n_steps’: scope.int(hp.quniform(’n_steps’,1,8,1)), ’epsilon’: hp.uniform(’epsilon’,-5,0), ’batch_size’: hp.choice(’batch_size’, [512,1024,2048,4096,8192]

Table 2. Hyperparameter search space for each of the machine learning models.

4. Results and discussion

This section presents the results of the proposed method and the comparison with the current literature for predicting in-hospital outcomes. The hyperparameters were chosen automatically based on the Bayesian optimization. In Figure 2, we present the confusion matrix obtained for each model.

Analyzing the confusion matrices, we can see a common behavior among all models, which is greater ease in identifying hospital discharge outcomes. The Random Forest model showed a greater tendency to misclassify death cases as hospital discharge. This behavior may be related to the higher number of hospital discharge cases in the dataset. The highest rate of false negatives was also found in the [De Souza et al. 2021] literature. The opposite behavior was presented by the TabNet model, which presented a higher rate of false positives. The model with the lowest false positive and false negative rates was XGBoost. This behavior may be related to the operating characteristic of using ensemble learning, where it combines several weaker models to create a stronger model. XGBoost trains the models in sequence, trying to correct the mistakes made by the previous mod-

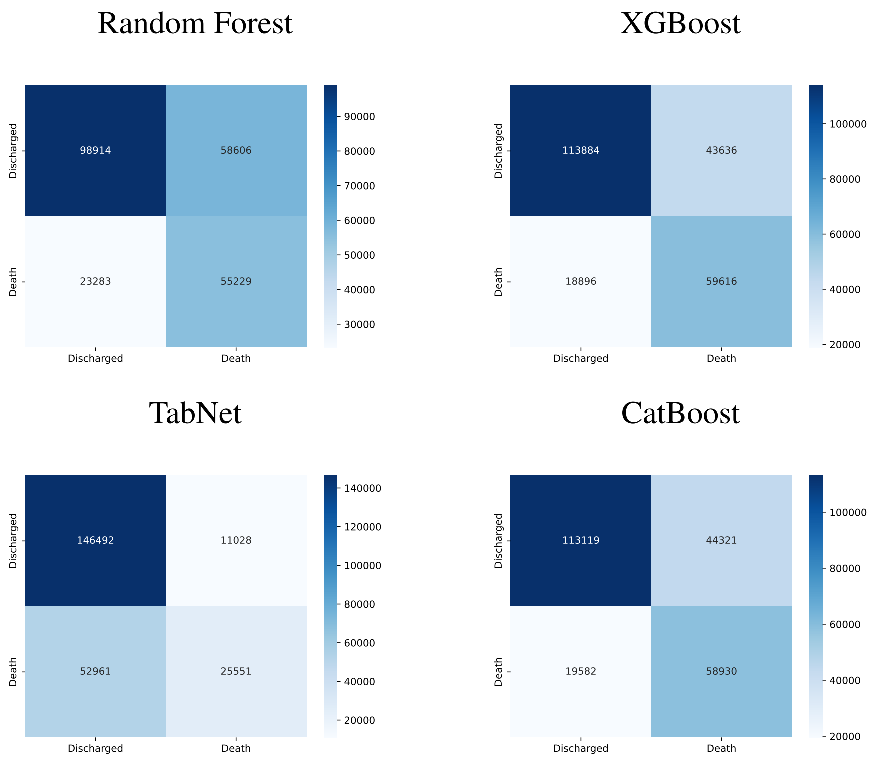

Figure 2. Confusion matrix for each model.

els. This can lead to a better fit of the training data and, consequently, a reduction in false positives and negatives.

In the Table 3 presents the performance obtained for the evaluation metrics in the test set.

Model/Study	Precision	Recall	F1-score	Accuracy	AUC
Random Forest	0.65	0.67	0.64	0.65	0.72
XGBoost	**0.72**	**0.74**	**0.72**	**0.74**	**0.83**
TabNet	**0.72**	0.63	0.63	0.73	0.78
CatBoost	**0.72**	0.73	0.71	0.73	0.81

Table 3. Results for performance metrics.

Analyzing the results of Table 3, we can see that the Random Forest model shows relatively lower precision, recall, F1-score, accuracy, and AUC than the other models in the table. One of the main strengths of Random Forest is its ability to handle high-dimensional datasets with a large number of features while also providing an estimate of feature importance. However, in this particular scenario, it seems that Random Forest may struggle with achieving high precision and recall, possibly due to the inherent complexity and variability of the COVID-19 data. Random Forest models tend to have a higher tendency for overfitting, which can result in reduced generalization performance and lower predictive accuracy.

On the other hand, the XGBoost model demonstrates the highest performance across most metrics, including precision, recall, F1-score, accuracy, and AUC. XGBoost is an ensemble learning algorithm that combines the predictions of multiple weak models to create a strong predictive model. It leverages gradient-boosting techniques to train weak learners, minimizing the overall loss function iteratively. The strengths of XGBoost lie in its ability to handle complex interactions between features, handle missing data, and effectively handle imbalanced datasets. This allows XGBoost to capture subtle patterns and make accurate predictions. However, one potential weakness of XGBoost is its increased computational complexity and longer training time, especially when dealing with large datasets.

TabNet demonstrates relatively high precision but lower recall compared to XGBoost and CatBoost. TabNet is a neural network-based model specifically designed for tabular data. It incorporates attention mechanisms to focus on important features during the training process. TabNet's strengths lie in its interpretability, as it provides feature importance scores and its ability to handle high-dimensional datasets and capture complex feature interactions. However, the lower recall suggests that TabNet may have difficulty correctly identifying all positive cases, limiting its usefulness in scenarios where recall is crucial.

CatBoost performs well across most metrics, including precision, recall, F1-score, accuracy, and AUC. CatBoost is another gradient-boosting algorithm that can handle categorical features and missing data without extensive data preprocessing. It uses ordered boosting, which allows it to handle categorical variables with a large number of categories effectively. CatBoost's strengths lie in its ability to handle complex data with categorical features, robustness against overfitting, and efficient training speed. However, CatBoost may perform better than XGBoost in scenarios where handling high-dimensional data or capturing complex feature interactions is crucial.

XGBoost and CatBoost demonstrate strong performance across multiple metrics predicting the target variable. XGBoost excels in capturing complex patterns and achieving high precision, recall, and overall accuracy, while CatBoost performs well with categorical features and maintains efficient training speed. TabNet, although it shows higher precision, may have limitations in recall. While it has strengths in handling high-dimensional data, Random Forest appears to have lower performance in this particular scenario. Desta forma, considerando as particularidades de cada modelo e os resultados apresentados na Tabela 3 podemos inferir que os melhores resultados foram obtidos pelo XGBoost.

Regarding the ROC curve (Fig. 3), the proposed method obtained a value of 0.83. Compared with other classification methods using similar strategies [Baqui et al. 2021], our approach achieved superior performance in terms of AUC-ROC. Furthermore, this demonstrates the effectiveness of the proposed method in capturing relevant patterns and making accurate predictions for COVID-19 outcomes. The robust performance of the model further supports its potential utility as a valuable tool for risk stratification and clinical decision-making in managing COVID-19 patients.

When we analyze the importance of the features for the XGBoost model (Figure 4), we can see that the model considers epidemiological factors (number of cases, total vaccinated, and current epidemiological week) as the most important. Then we can see health quality and socioeconomic factors (number of beds, probability of survival

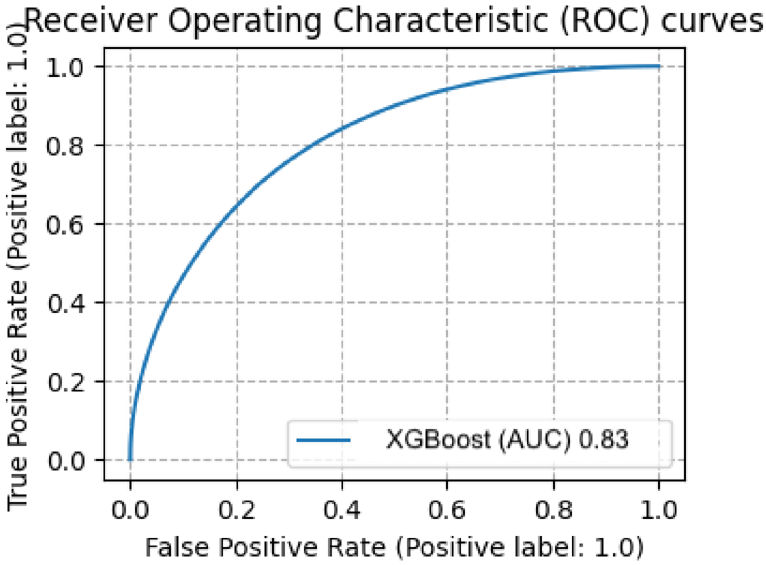

Figure 3. ROC curve for the XGBoost model.

after 60 years, life expectancy, rural population, and per capita income). These factors indicate the possibility of an association of economic characteristics in the outcomes of hospitalized patients with COVID-19 in Brazil.

5. Conclusion

This article compared four machine learning models for predicting early COVID-19 in-hospital outcomes. To improve the generalizability of the results, we combined several datasets focusing on epidemiologic, socioeconomic, and sociodemographic indicators. The main scientific contribution of this work is the proposal of a method for predicting hospital outcomes in patients with COVID-19. The models can serve as a basis for future studies and provide a second opinion to healthcare professionals during COVID-19 outbreaks.

The present study has some limitations. First, although the study provides information about the relative importance of the characteristics considered by the models, it is important to interpret these conclusions with caution. The importance attributed to a characteristic can be influenced by the interaction with other variables and by the specificity of the dataset used. Therefore, additional care is needed in interpreting these results. Furthermore, the current results are based on a Brazilian dataset. Differences in health systems, disease control measures, and demographic characteristics can influence hospital outcomes. The selection of machine learning models used in the study may introduce a selection bias. Although the models were developed based on Bayesian optimization, other models or approaches could have been considered, which could lead to different results.

Despite limitations, this study contributes to the healthcare field by providing an effective method to predict hospital outcomes in patients with COVID-19. These predictions can help health professionals to identify patients at higher risk of complications or death, allowing for appropriate interventions and resource planning. In addition, the

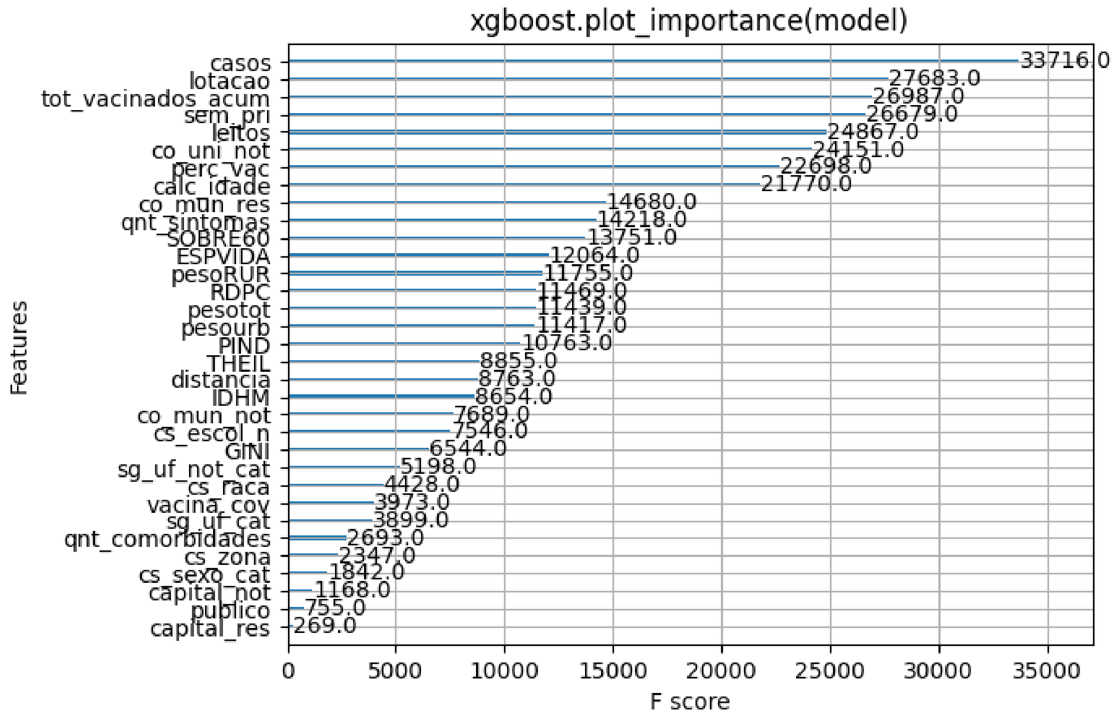

Figure 4. Feature importance for the XGBoost model.

analysis of the importance of characteristics highlighted the relevance of epidemiological, health, and socioeconomic factors in predicting hospital outcomes.

In the ever-evolving context of the COVID-19 pandemic, additional research is needed to validate and improve the proposed approaches, considering different populations and contexts. With this, it will be possible to develop more effective prevention, management, and clinical decision-making strategies to improve the health outcomes of patients with COVID-19. Future work will involve evaluating other machine learning methods to explore their effectiveness. Additionally, more sophisticated techniques for model interpretation will be employed to enhance understanding of the relationship between attributes and prediction accuracy. Feature selection techniques will also be applied to identify the most relevant and non-redundant subset of features for addressing this problem.

Acknowledgement

We thank the anonymous reviewers for their valuable suggestions. The authors would like to thank the Coordination for the Improvement of Higher Education Personnel - CAPES (Financial Code 001), and the National Council for Scientific and Technological Development - CNPq (Grant numbers 309537/2020-7 e 305263/2021–8) for your support in this work.

References

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of sars-cov-2. *Nature medicine*, 26(4):450–452.

Arik, S. Ö. and Pfister, T. (2021). Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6679–6687.

Baqui, P., Marra, V., Alaa, A. M., Bica, I., Ercole, A., and van der Schaar, M. (2021). Comparing covid-19 risk factors in brazil using machine learning: the importance of socioeconomic, demographic and structural factors. *Scientific reports*, 11(1):15591.

Barough, S. S., Safavi-Naini, S. A. A., Siavoshi, F., Tamimi, A., Ilkhani, S., Akbari, S., Ezzati, S., Hatamabadi, H., and Pourhoseingholi, M. A. (2023). Generalizable machine learning approach for covid-19 mortality risk prediction using on-admission clinical and laboratory features. *Scientific Reports*, 13(1):2399.

Betthäuser, B. A., Bach-Mortensen, A. M., and Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the covid-19 pandemic. *Nature Human Behaviour*, 7(3):375–385.

Cribari-Neto, F. (2023). A beta regression analysis of covid-19 mortality in brazil. *Infectious Disease Modelling*, 8(2):309–317.

De Souza, F. S. H., Hojo-Souza, N. S., Dos Santos, E. B., Da Silva, C. M., and Guidoni, D. L. (2021). Predicting the disease outcome in covid-19 positive patients through machine learning: a retrospective cohort study with brazilian data. *Frontiers in Artificial Intelligence*, 4:579931.

Docherty, A. B., Harrison, E. M., Green, C. A., Hardwick, H. E., Pius, R., Norman, L., Holden, K. A., Read, J. M., Dondelinger, F., Carson, G., et al. (2020). Features of 20 133 uk patients in hospital with covid-19 using the isaric who clinical characterisation protocol: prospective observational cohort study. *bmj*, 369.

Dorogush, A. V., Ershov, V., and Gulin, A. (2018). Catboost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.

Figuerêdo, J. S. L., Araújo-Calumby, R. F., and Calumby, R. T. (2021). Machine learning for prognosis of patients with covid-19: An early days analysis. In *Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional*, pages 59–70. SBC.

Green, H., Fernandez, R., and MacPhail, C. (2021). The social determinants of health and health outcomes among adults during the covid-19 pandemic: A systematic review. *Public Health Nursing*, 38(6):942–952.

Tsiotas, D. and Tselios, V. (2022). Understanding the uneven spread of covid-19 in the context of the global interconnected economy. *Scientific reports*, 12(1):666.

Zeiser, F. A., Donida, B., da Costa, C. A., de Oliveira Ramos, G., Scherer, J. N., Barcellos, N. T., Alegretti, A. P., Ikeda, M. L. R., Müller, A. P. W. C., Bohn, H. C., et al. (2022). First and second covid-19 waves in brazil: A cross-sectional study of patients' characteristics related to hospitalization and in-hospital mortality. *The Lancet Regional Health-Americas*, 6:100107.

Zhang, B., Kang, W., Xiong, S., Huang, X., Chen, P., Huang, J., Hou, Y., Ma, L., and Xiang, T. (2023). Changes in the epidemiological characteristics of prehospital emergency services before and during the covid-19 pandemic, chengdu, 2016–2021. *Scientific Reports*, 13(1):7796.