

SUPPORTING DECISION MAKING IN HEALTH SCENARIOS WITH MACHINE LEARNING MODELS

Maria Gabriela Valeriano

Instituto Tecnológico de Aeronáutica - Universidade Federal de São Paulo

Brazil

mariagabrielavaleriano@gmail.com

Carlos Roberto Veiga Kiffer

Universidade Federal de São Paulo

Brazil

carlos.rv.kiffer@gmail.com

Ana Carolina Lorena

Instituto Tecnológico de Aeronáutica

Brazil

aclorena@gmail.com

ABSTRACT

Machine Learning (ML) techniques have been employed in data analysis for generating models able to support health care and management decisions, but in Brazil there is still a lack of adequate data collection procedures in hospital environments, preventing a proper data analysis. Here we discuss some approaches adopted to deal with these challenges and carry out analyzes of hospital data to support reliable decision-making. From data stored in the *COVID Data Sharing/BR* repository we set up two datasets for severity prediction of COVID-19 cases upon admission parameters. We explain decisions made in data preprocessing and discuss some results of the ML models, including the possibility of generalizing these models across different hospitals. The best AUC values achieved were around 0.75, a value lower than that reported in related literature, which can be explained by the absence of patient clinical information from the raw public repositories.

KEYWORDS. Machine Learning. COVID-19 prognosis models. Decision making.

Paper Topics: OR in health area

1. Introduction

Inside the Artificial Intelligence (AI) research field, Machine Learning (ML) techniques have been successfully used in several predictive and descriptive data analyses [Faceli et al., 2021]. Concerning the analysis of health data, ML systems are employed in model generation to support medical diagnosis and prognosis of patients in view of the evolution of their symptomatic conditions [Kononenko, 2001; Kourou et al., 2015]. Regarding COVID-19, since the beginning of the pandemic ML models have been produced to diagnose patients with this disease [de Moraes et al., 2020]. Some studies have also used ML techniques to build prognostic models capable of predicting the development of a severe condition [Fernandes et al., 2021].

The results Zhang et al. [2020] point out that the levels of erythrocyte, D-dimer, albumin and interleukin-6 sedimentation rate are closely related to the development of a severe condition. Jiang et al. [2020] have found as the most important clinical variables to predict the development of severe COVID-19 conditions the rate of alanine aminotransferase, the presence of body pain and the level of hemoglobin. In a study carried out in Brazil, Fernandes et al. [2021] have reported the rate of lymphocytes by C-reactive protein, C-reactive protein and Braden scale as the main predictors of COVID-19 severity. The different results found may be a reflection of little comprehension about COVID-19, as the disease is new and knowledge about its treatment is under continuous adaptation. In addition, this divergence of results can also be attributed to specific characteristics of the databases used in each study, and to characteristics of health data which carry out several challenges, such as the absence of values in some patient records [Rahman e Davis, 2013], class imbalance and input noise. As the ML analyzes are strongly data-driven, there is need a to guarantee data quality in order to obtain valid generalizations [Hand, 1999; Roh et al., 2019].

In Brazil, a large developing country which was very impaired by the COVID-19 pandemic, there are some initiatives on data collection, storage and public availability. This is the case of the *COVID Data Sharing/BR*¹ initiative, which contains laboratory and demographic information from COVID-19 patients attended at some large hospitals from the São Paulo metropolitan area. This is an initiative supported by the São Paulo Research Foundation (FAPESP). In this repository, it is possible to observe many differences related to the lab tests applied to each patient, resulting in a high rate of missing values. There are also different nomenclatures for similar lab tests and measurement units differences. Some databases do not have patient's outcome information either. There is a high variability on the type of data available per center too, which can be attributed to differences in the profiles of the patients attended at each health center as well as the lack of a standardized protocol for COVID-19 attendance and care in the Brazilian territory.

Summing up, there is a lack of adequate data collection culture in Brazilian hospital environments, which can lead to a large number of data inconsistencies and impair their analysis. Therefore, although a large amount of data about patients is collected, there is a lack of standardization and, in general, great challenges for organizing suitable datasets to support a reliable predictive analysis. Inside this scenario we intend to contribute to the field by discussing and addressing some possibilities to deal with these challenges. From data stored in the COVID Data Sharing - BR repository we assembled two datasets for severity prediction of patients with COVID-19. In this work we will explain decisions made in data preprocessing and discuss some results of ML models built from these datasets, including the possibility of generalizing these models across different hospitals.

This paper is divided as follows: Section 2 describes in detail the steps to assemble the datasets used in this work. All the decisions made in data preprocessing were widely discussed with one or more data experts (physicians). Their description can contribute to future works using

¹<https://repositoriodatasharingfapesp.uspdigital.usp.br/>

the same raw database. Section 3 describes the methodology adopted to generate the ML models, and again we explain some strategies employed to deal with health data particularities. Section 4 presents and discusses our results. Finally, Section 5 concludes this work.

2. Data assembling and preprocessing

Using patient data from health centers, it is possible to use ML techniques to build COVID-19 diagnoses and/or prognosis models [de Moraes et al., 2020; Fernandes et al., 2021]. Some possible prognosis of positive COVID-19 patients include the future need for hospitalization, intensive care units, respirators, among others [Fernandes et al., 2021; Abuabara et al., 2021]. Based on patient data and on data from previous patients, ML models can give these expected prognosis of a patient during early stages of the disease, helping the health centers and professionals in care decisions and resource allocation.

The severity of COVID-19 cases can be defined in multiple ways. The evolution to death clearly indicates a higher severity of the case. But even if a case does not evolve to death, a long term hospitalization could leave sequels [Sagarra-Romero e Viñas-Barros, 2020] and also brings high demands of human and material resources. In this paper we have adopted as severity criteria 14 or more days of hospitalization or death. Therefore, the need for a long term hospitalization is also considered a proxy of the severity of the case.

Our ML predictive models use a supervised learning approach, which requires a labeled dataset for training. The idea is to use retrospective data from patients admitted in a health center, along with their outcomes, to build the labeled datasets. Two databases from the COVID-19 Data Sharing Repository contain information on the patients outcomes: data from Hospital Sírío Libanês (HSL) and data from Hospital Beneficência Portuguesa (HBP).

Each hospital's raw database is composed of three files. The first is a table with demographic patient information like birth year and gender, the second registers laboratory tests results performed at the hospital as well as reference values, and the third file contains some information about the patient's hospital stay and outcomes. Figure 1 represents all steps of data preprocessing, which are also numbered for reference.

Inside each hospital database we found some inconsistencies in lab tests labels. They had to be identified and renamed beforehand, with the aid of a domain expert (1). After that, the files were joined, aggregating the demographic information from each patient to his/her outcome and lab test results (2). At this point we had one table or dataframe per hospital, where each row refers to one lab test result.

The next preprocessing step consisted of filtering just the lab tests of patients who have confirmed COVID-19 diagnosis by CRP test up to the fifteen days after hospital attendance (3). After filtering the tests of positive cases we have discarded laboratory tests done after the fourth day of hospitalization (4). For each patient, when there is more than one result registered for a lab test, only the first value was kept. The idea is to capture the status of patient admission, in order to give early prognosis of severity (5). Next we kept data from only young and adult individuals, with age superior to sixteen, as children's lab tests results may be not comparable with those of the adults (6).

In these hospital databases there are patients received at the emergency room, in the clinic and hospitalized patients. Since our goal is to predict severity outcomes from hospitalization days and death, we then kept only the hospitalized patients (7). A patient was considered hospitalized only when the length of hospital stay was greater than one day. For several reasons such as hospital overload, or for personal reasons, a patient could be discharged without being restored or cured. For

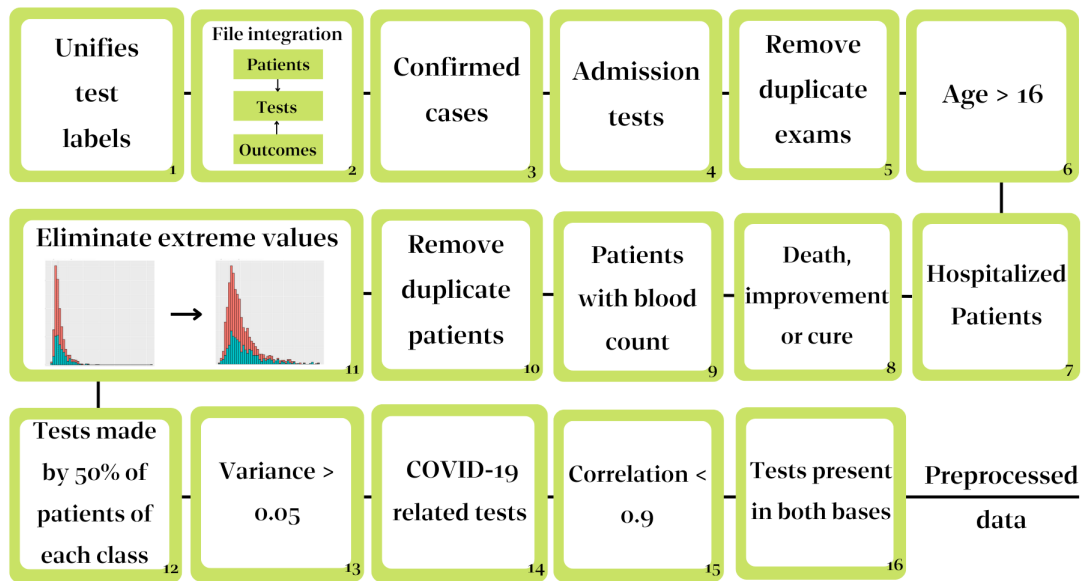


Figure 1: Steps included in data preprocessing.

this reason, we kept only patients whose outcome was *death, medical discharge cured or improved medical discharge* (8).

At this point we created a severity column to store each patient's label, either as severe or not. There is no agreement in the literature on how to classify a COVID-19 case as severe. In our study we have adopted as severity criteria a hospitalization stay of 14 days or more, or death before the 14th day of hospitalization. Regarding the number of 14 days, there is also no definition of how long a hospitalization stay should be to be considered a *long term* hospitalization. An study about long term hospitalization in China has adopted 14 days since this is the average number of days of hospitalization in their cohort [Wu et al., 2020]. In one of our databases (HSL), 14 was also the average number of hospitalization days, while in the other database (HBP) this number was 12. For standardization purposes, we adopted 14 days for both hospitals as a long term stay.

The data was then transformed, with each patient becoming a row of the dataset and the lab test results being stored in the columns. The next filter step consisted of removing patients without blood counts results (9). With this step we eliminated patients with few recorded parameters and consequently reduced the percentage of missing values. Blood counts were adopted as inclusion criterion because it is a lab test largely employed in hospital's admissions in Brazil, due to its low cost. Some patients were hospitalized more than one time. In these cases only the last hospitalization record was kept in order to remove duplicates and dependency between instances (10).

After that we plotted histograms to visualize data and observed the presence of outliers. In health scenarios sometimes it can be difficult to automatically deal with outliers. While some outliers are results of human or experimental errors, they can also be correct and anomalous when the patient presents some health disturbances. At this point the datasets contained hundreds of columns, making it impossible to perform a test-by-test expert evaluation. Facing this challenge, we decided to eliminate the four most extreme values for each lab test (11). If the histogram was

Table 1: Summary of the datasets used in this study, including number of patients, number of admission tests considered, period of data collection, percentage of severe cases, percentage of male patients and age statistics

Hosp.	# Pat.	# Fea.	Period	% Sev.	% Male	Age [min-max]
HSL	1432	20	03/2020 - 05/2021	36.7	65.6	62.2 [19 - 89]
BP	185	20	03/2020 - 01/2021	36.2	55.1	59.4 [24 - 86]

Table 2: Acronyms for the groups of models built

Group	Description
I	Internal cross-validation in HSL
II	Internal cross-validation in HBP
III	External validation with training at HSL and testing at HBP

displaced to the left, the four most extreme major values were removed. The opposite reasoning applies in the case of histograms displaced towards the right. When the histogram was centered, the two most extreme values from both tails of the histogram were removed. The histogram plots showed more equalized results after this step.

At this point there are still many lab tests with few values recorded. We decided to keep the lab tests for which at least 50% of patients from each class (severe and non severe) are filled (12). All remaining missing values are imputed using the filled data, which had to be representative enough. Next we eliminated features with a variance lower than 0.05 (13). The features with low variance across all patients have a low contribution to the ML models, since they are unable to discriminate between patients of the two classes of interest. The remaining lab tests were then examined by a domain expert, in order to eliminate those that are not COVID-19 related (14). Besides that, for features pairs with a correlation index greater than 0.9, one of them was randomly removed to eliminate data redundancies (15). Finally, we eliminated the lab tests that are not present in one of the datasets, so that we can generate models that generalize across the two hospitals considered. Table 1 summarizes the datasets main characteristics after all previous steps described are performed, with number of patients, input features, period of data collection, percentage of severe cases, percentage of male individuals and range of age. The two datasets, and the codes used to preprocess them are available on github².

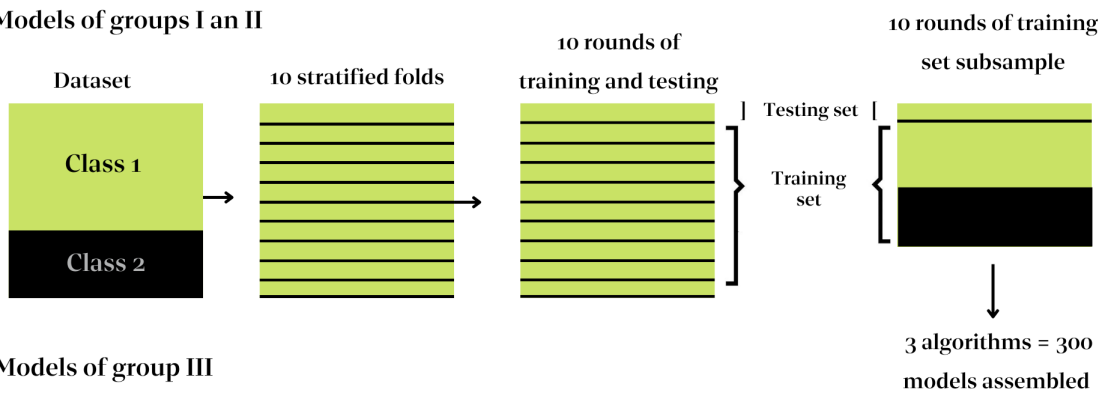
3. Methodology

Three Machine Learning (ML) techniques with distinct biases were employed in the analysis of the datasets assembled in the previous section: Support Vector Machines (SVM), Gradient Boosting (GB) and Random Forests (RF). SVM is based on the Statistical Learning Theory and builds decision models separating data from different classes with a maximum margin. GB and RF generate ensembles of Decision Tree models from data and have achieved state-of-the-art performance in the analysis of tabular datasets such as ours. All implementations used are from the scikit-learn Python package [Pedregosa et al., 2011].

We developed three groups of models, referred next as I, II and III, as summarized in Table 2. In groups I and II we trained and tested the ML models using data from a same hospital, in an internal 10-fold cross-validation setup. Herewith, each dataset was divided into ten folds of approximately equal size while maintaining the proportion of observations per class. Next, ten train-test rounds of the previous ML models were performed. At each round, one data fold was left out for testing while the remaining folds composed the training set. As the proportion of non-severe cases is quite high compared to the proportion of severe cases, a random subsampling of the

²<https://github.com/gabivaleriano/defend-severity>

Models of groups I and II



Models of group III

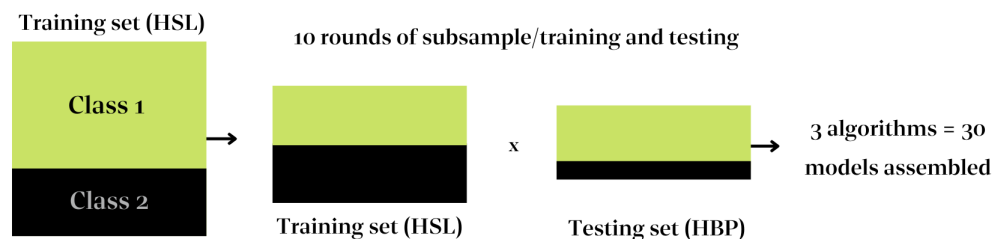


Figure 2: Methodology employed to assemble models of groups I, II and III.

majority class samples was performed in the training datasets. This is necessary because when faced with highly imbalanced datasets, the ML techniques tend to favor the majority class in detriment to the minority class. This random subsampling step was repeated ten times to avoid any sample bias. Summarizing, 100 models are built per dataset and ML technique, as schematized in the top of Figure 2.

In the models of group III we employed data from HSL in the training step and data from HBP in the testing step. With these results we can obtain insights about the ability to generalize models across hospitals. As data used for training were also imbalanced, the random subsampling strategy was employed again, 10 times, so that 10 models were trained in this experiment. We adopted only HSL data for training because this dataset has a large amount of data and consequently a more representative sample which increases the reliability of the models trained. We can also interpret the results from these experiments as an external validation of the results achieved from models trained and tested with internal cross-validated HSL data.

Missing data in the datasets were input using a three-nearest neighbor strategy based on training data. For each observation with a missing value, the three-nearest neighbors in the training dataset are sought. The missing values were imputed as the average of the filled feature values of these neighbors.

Each of the ML techniques has a hyper-parameter that must be defined before the training step. Different choices for their values can result in different accuracy performances. In this work we adopted a Bayesian optimization procedure for hyper-parameter tuning available at the Python hyper-opt package [Bergstra et al., 2015]. This approach employs past evaluations to build a probabilistic model, which is used to choose the best hyper-parameter values.

Finally, the evaluation of the predictive results of the ML models in the test sets was performed according to three representative metrics: AUC (Area Under ROC Curve), Sensitivity (accuracy in the severity class) and Specificity (accuracy in the non-severe class). These metrics

Table 3: Predictive results achieved for severity prognosis of COVID-19 cases in models of groups I, II, and III. SVM refers to the Support Vector Machine classifier, GB denotes the Gradient Boosting classifier and RF refers to the Random Forest classifier. Best results per metric and dataset are in bold.

Models	SVM			GB			RF		
	AUC	Sensit.	Specif.	AUC	Sensit.	Specif.	AUC	Sensit.	Specif.
I	0.684	0.636	0.634	0.756	0.749	0.695	0.756	0.693	0.683
II	0.731	0.704	0.683	0.755	0.789	0.733	0.724	0.619	0.703
III	0.513	0.973	0.019	0.747	0.663	0.754	0.748	0.640	0.736

vary in the $[0,1]$ range and, the higher the value, the better the results achieved. For AUC, results around 0.5 are bad and correspond to a random prediction. Average and standard deviation values across the multiple training-test runs were recorded.

4. Experimental Results and Discussion

Table 3 presents the average predictive performance for models of groups I, II and III, induced for COVID-19 severity prognosis. The best performance values per metric and ML model are highlighted in bold in the table. In the models assembled for groups I and II we have adopted a ten fold cross-validation approach to train and test the models with data from the same hospital (HSL in group I and HBP in group II). As described previously, we also trained a third group (group III), in order to verify the possibility to generalize the results across different health centers. Data from HSL was used for training the ML models, which were tested with data from HBP. For each train-test round the model performance was registered, and the average and standard deviation (SD) were calculated. The SD for models of group I were under 0.08. For models of group II SD values were between 0.1 and 0.2, an expected result since the cohort of patients was smaller in HBP compared to HSL. For models of group III the SD values were under 0.07.

In the three scenarios, it was possible to achieve AUC results close to 0.75. The sensitivity and sensibility values varied more. GB was the best performing technique in all metrics for scenarios I and II. That is, the cross-validation results within each hospital are better for this technique, both in terms of AUC, sensitivity and specificity.

Taking the group III and SVM results, we notice a prominence in the correct classification of severe cases (high sensitivity), while the performance is very deteriorated for the non-severe class (low specificity). Therefore, the SVM model shows an atypical behavior classifying all patients as severe. This reflects in the AUC value, which is close to 0.5, representing a random predictive behavior. For GB and RF, the results of the external validation in group III were more similar to those of the internal validation (model I).

In a previous work [Valeriano et al., 2022] we employed the same raw databases to predict severity, but including also patients received at the hospitals' clinic and emergency rooms. Predicting severity for hospitalized patients only is more challenging since they already have a more severe condition than the non-hospitalized patients, making it more difficult to distinguish severe from non-severe cases. Because of that, we inserted here some extra steps to improve data quality (steps 6, 7, 8, 13, 14 and 16) compared to our previous work and also inserted a Bayesian optimization for a better hyper-parameter tuning [Putatunda e Rama, 2018].

Similar studies to predict severity for hospitalized patients showed better predictive performances than our models, with AUC scores of up to 0.92 in [Kim et al., 2020] and around 0.885-0.897 in [Fernandes et al., 2021]. But these work included more clinic variables unavailable in our cohort, such as BRADEN scale and the initial symptoms (dyspnea and initial body temperature). The clinic information has a great prediction power of COVID-19 clinical aggravation. In a previous work using other database containing only symptoms and comorbidities of a set of Brazilian

citizens performing a COVID-19 test, we were able to classify severe patients with an AUC of 0.954 [Abuabara et al., 2021]. Therefore, we expect that by aggregating clinic information to the lab tests results the predictive performance metrics can be improved. In the absence of this information the results are not satisfactory, but they show at least the correlation between severity and the results of routine lab tests performed in the first days of hospital admission.

The results also show evidence that models can be generalized between the two hospitals considered. We believe that this was possible because these hospitals attend a similar population, from the same city and with similar social and financial conditions. Except for the SVM classifier, the results presented for models of group III are similar to those presented for models of groups I and II.

An interesting point for future studies is to investigate the influence of data distribution. In a study published by Yoon et al. [2018], a novel approach is presented to deal with related but different databases. The authors propose a translation from one dataset to another, using a multiple GAN (Generative Artificial Network) architecture. This translation between datasets resolve problems of different features and different distributions. Adopting such an approach in the datasets of this work could improve the results for models of group III, since it would eliminate differences in datasets distributions and also would allow to include more features in the models that were not included here because they are absent from one of the datasets. Another relevant matter for future studies is the adoption of a more robust method to data imputation [Beretta e Santaniello, 2016].

We would like to point also the benefits of public health repositories such as COVID Data Sharing - BR. From data available in this initiative different works have been produced [Pimentel et al., 2021; Thimoteo et al., 2020; dos Santos e Berton, 2021], allowing to enlarge the comprehension of the COVID-19 disease diagnosis and prognosis in our country. Unfortunately, the repository has not been updated for over a year and the evolution of variants and the disease dynamic in these hospitals could not be accessed. The difficulty to access health data discourage researchers from dedicating their efforts to data analysis, and also delays the benefits that society could receive from such research.

As an example of how ML can support decision making in a health center one can mention the COVID-19 Capacity Planning and Analysis System (CPAS) developed in the United Kingdom (UK) [Qian et al., 2021]. Working with the National Health System (NHS) the authors implemented a system for supporting decisions in different levels. The system is able to give insights from different perspectives, for different groups of decision makers and gives an accurate forecast of ICU demands for several time windows and possible policies implemented to face the pandemic. CPAS is trained using data that reflects the complete patient journey from hospital admission, ICU admission, mechanical ventilation and finally hospital release or death. The authors worked with data from COVID related ICU admissions from 94 hospitals, comprising general health conditions of patients, specific risks of COVID, basic demographic information and what treatments they received. For the severe patients the authors also accessed details of measurements of these patients, physiological status of the first 24 hours, organ support status records, database containing all the details of all admissions and all outpatient consultation information.

The authors cite an illustrative case to show the potential of CPAS to aid resources management of one specific hospital. In one of the pandemic waves, with growing cases and hospitalizations, the hospital management was planning to convert some existing general wards into ICU wards, and maybe transfer some patients to another hospital. The result of CPAS simulating predicted the ICU admission of between 20 - 50 patients. With this estimation, the hospital management decided to convert some wards into ICU wards and that there was no need for transferring

patients. The hospital actually admitted 42 patients to the ICU in this period, showing the CPAS capacity of providing accurate and useful predictions. A major effort on data collection in hospitals would be needed in Brazil in order to take advantage of similar decision support systems.

5. Conclusions

Brazil has faced the proliferation of several contagious diseases in recent years, the most recent case being the COVID-19 pandemic. Reference hospitals in large Brazilian cities face many difficulties in managing their resources and professionals during these periods. Some of the major issues observed in health data in Brazil, one of the countries most affected by the COVID-19 pandemic worldwide, are: getting data access; lack of standardization in procedures adopted during patient care and on laboratory and clinical exams requested; changes in data collection procedures through time; large amounts of inconsistent and missing data; different population mixes and characteristics per health center; a high class imbalance; among others. All of them make it difficult to assemble ML predictive models which are reliable and general enough for practical usage.

In this work we assembled two datasets from the COVID Data Sharing BR repository to predict which hospitalized patients would evolve to a severe condition, defined as an hospitalization longer than 14 days or death. We discussed all decisions taken in data preprocessing as well as in ML model training using such data. Two approaches were used for building and evaluating the prognosis models: an internal 10 fold cross-validation using data from each hospital; and a cross-hospital training and testing. In both cases the ML models reached AUC up to 0.75, which we believe can be raised if clinical information currently unavailable in the raw databases are used.

Acknowledgments

The authors acknowledge the Brazilian research agency FAPESP for providing data from Hospitals Beneficência Portuguesa and Sírío Libanês under the COVID-19 Data Sharing/BR initiative. The financial support of the research agencies CAPES (grant 88887.507037/2020-00) and CNPq is also acknowledged.

References

- Abuabara, L., Valeriano, M. G., Kiffer, C. R. V., Yanasse, H. H., e Lorena, A. C. (2021). Using machine learning to support health system planning during the covid-19 pandemic: a case study using data from são josé dos campos (brazil). *CLEI Electronic Journal*, 24(3). URL <https://doi.org/10.19153/cleiej.24.3.5>.
- Beretta, L. e Santaniello, A. (2016). Nearest neighbor imputation algorithms: a critical evaluation. *BMC medical informatics and decision making*, 16(3):197–208.
- Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., e Cox, D. D. (2015). Hyperopt: a python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, 8(1): 014008.
- de Moraes, B. A. F., Miraglia, J., Donato, T., e Filho, A. (2020). Covid-19 diagnosis prediction in emergency care patients: a machine learning approach. <https://www.medrxiv.org/content/medrxiv/early/2020/04/07/2020.04.04.20052092.full.pdf>.
- dos Santos, J. A. e Berton, L. (2021). Applying machine learning to assist the diagnosis of covid-19 from blood and urine exams. In *Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional*, p. 256–267. SBC.

- Faceli, K., Lorena, A., Gama, J., Almeida, T. A., e Carvalho, A. (2021). Inteligência artificial—uma abordagem de aprendizado de máquina. *Rio de Janeiro: LTC*.
- Fernandes, F. T., de Oliveira, T. A., Teixeira, C. E., Batista, A. F. d. M., Dalla Costa, G., e Chivegatto Filho, A. D. P. (2021). A multipurpose machine learning approach to predict covid-19 negative prognosis in são paulo, brazil. *Scientific reports*, 11(1):1–7.
- Hand, D. J. (1999). Statistics and data mining: intersecting disciplines. *Acm Sigkdd Explorations Newsletter*, 1(1):16–19.
- Jiang, X., Coffee, M., Bari, A., Wang, J., Jiang, X., Huang, J., Shi, J., Dai, J., Cai, J., Zhang, T., et al. (2020). Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua*, 63(1):537–551.
- Kim, H.-J., Han, D., Kim, J.-H., Kim, D., Ha, B., Seog, W., Lee, Y.-K., Lim, D., Hong, S. O., Park, M.-J., et al. (2020). An easy-to-use machine learning model to predict the prognosis of patients with covid-19: Retrospective cohort study. *Journal of medical Internet research*, 22(11):e24225.
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1):89–109.
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., e Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, 13:8–17.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.
- Pimentel, G. D., Vega, M. C. D., e Pichard, C. (2021). Low vitamin d levels and increased neutrophil in patients admitted at icu with covid-19. *Clinical Nutrition ESPEN*, 44:466–468.
- Putatunda, S. e Rama, K. (2018). A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of xgboost. In *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, p. 6–10.
- Qian, Z., Alaa, A. M., e van der Schaar, M. (2021). Cpas: the uk’s national machine learning-based hospital capacity planning system for covid-19. *Machine Learning*, 110(1):15–35.
- Rahman, M. M. e Davis, D. N. (2013). Machine learning-based missing value imputation method for clinical datasets. In *IAENG transactions on engineering technologies*, p. 245–257. Springer.
- Roh, Y., Heo, G., e Whang, S. E. (2019). A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33(4): 1328–1347.
- Sagarra-Romero, L. e Viñas-Barros, A. (2020). Covid-19: Short and long-term effects of hospitalization on muscular weakness in the elderly. *International journal of environmental research and public health*, 17(23):8715.

- Thimoteo, L. M., Vellasco, M. M., Amaral, J. M. D., Figueiredo, K., Yokoyama, C. L., e Marques, E. (2020). Interpretable machine learning for covid-19 diagnosis through clinical variables. *Anais do Congresso Brasileiro de Automática 2020*.
- Valeriano, M. G., Kiffer, C. R. V., Higino, G., Zanao, P., Barbosa, D. A., Moreira, P. A., Santos, P. C. J. L., Grinbaum, R., e Lorena, A. C. (2022). Let the data speak: analysing data from multiple health centers of the s ao paulo metropolitan area for COVID-19 clinical deterioration prediction. In *2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*. IEEE. URL <https://doi.org/10.1109/ccgrid54584.2022.00115>.
- Wu, Y., Hou, B., Liu, J., Chen, Y., e Zhong, P. (2020). Risk factors associated with long-term hospitalization in patients with covid-19: a single-centered, retrospective study. *Frontiers in Medicine*, p. 315.
- Yoon, J., Jordon, J., e Schaar, M. (2018). Radialgan: Leveraging multiple datasets to improve target-specific predictive models using generative adversarial networks. In *International Conference on Machine Learning*, p. 5699–5707. PMLR.
- Zhang, H., Wang, X., Fu, Z., Luo, M., Zhang, Z., Zhang, K., He, Y., Wan, D., Zhang, L., Wang, J., et al. (2020). Potential factors for prediction of disease severity of covid-19 patients. *MedRxiv*.