J. Health Inform. *2024 Janeiro-Dezembro; 16*

Artigo Original

DOI: https://doi.org/10.59681/2175-4411.v16.2024.1027

Prediction of COVID-19 patients' admission to the Intensive Care Unit based on the precision nursing framework

Predição da admissão de pacientes COVID-19 em Unidade de Terapia Intensiva baseada na estrutura de Enfermagem de Precisão

Predicción del ingreso de pacientes COVID-19 en la Unidad de Cuidados Intensivos basada en el marco de la Enfermería de Precisión

Greici Capellari Fabrizzio[1], Alacoque Lorenzini Erdmann[2], Lincoln Moura de Oliveira[3], Elisiane Lorenzini[4], Rodrigo Jensen[5], José Luís Guedes dos Santos[6]

Keywords: Forecasting; Precision medicine; COVID-19.

ABSTRACT

Objective: apply and compare machine learning algorithms to predict COVID-19 patients' admission to Intensive Care Unit from the Precision Nursing theoretical framework. Methods: Retrospective study with 180 patients reported in the city of Florianopolis. The performance of the following algorithms was evaluated: multilayer perceptron - artificial neural network, AdaBoost, logistic regression, random forest, kNN, Naive Bayes, SVM and decision tree. Results: The predictor variables that most influenced the model were hospital admission, race and throat pain. The multilayer perceptron model achieved better prediction for AUC (0.917), sensitivity (0.861), and specificity (0.825). Conclusion: This application proved to be a viable method for predicting the admission of COVID-19 infected patients to ICU and the clinical biomarkers prove to be relevant for the clinical practice of Nursing because they are easily observable and can be quickly implemented.

Descritores: Modelos de predição; Medicina de Precisão; COVID-19

RESUMO

Objetivo: aplicar e comparar algoritmos de aprendizagem de máquinas para prever a admissão de pacientes COVID-19 na Unidade de Terapia Intensiva a partir do referencial teórico de Enfermagem de Precisão. Métodos: Estudo retrospectivo com 180 pacientes notificados na cidade de Florianópolis. Foi avaliado o desempenho dos seguintes algoritmos: perceptron multicamadas - rede neural artificial, AdaBoost, regressão logística, floresta aleatória, kNN, Naive Bayes, SVM e árvore de decisão. Resultados: As variáveis preditoras que mais influenciaram o modelo foram a admissão hospitalar, raça e a dor de garganta. O perceptron multicamadas conseguiu uma melhor performance para AUC (0,917), sensibilidade (0,861), e especificidade (0,825). Conclusão: Esta aplicação provou ser um método viável para prever a admissão de pacientes infectados com COVID-19 na UCI e os biomarcadores clínicos provaram ser relevantes para a prática clínica de Enfermagem porque são facilmente observáveis e podem ser rapidamente implementados.

Descriptores: Predicción; Medicina de Precisión; COVID-19.

RESUMEN

Objetivo: aplicar y comparar algoritmos de aprendizaje automático para predecir el ingreso de pacientes COVID-19 en la Unidad de Cuidados Intensivos desde el marco teórico de la Enfermería de Precisión. Métodos: Estudio retrospectivo con 180 pacientes notificados en la ciudad de Florianópolis. Se evaluó el desempeño de los siguientes algoritmos: perceptrón multicapa - red neuronal, AdaBoost, regresión logística, random forest, kNN, Naive Bayes, SVM y árbol de decisión. Resultados: Las variables predictoras que más influyeron en el modelo fueron el ingreso hospitalario, la raza y el dolor de garganta. El modelo perceptrón multicapa obtuvo mejores resultados en cuanto a AUC (0,917), sensibilidad (0,861) y especificidad (0,825). Conclusión: Esta aplicación demostró ser un método factible para predecir el ingreso de pacientes infectados con COVID-19 en la UCI y los biomarcadores clínicos demostraron ser relevantes para la práctica clínica de Enfermería porque son fácilmente observables y pueden implementarse rápidamente.

[1] *Federal University of Santa Catarina (UFSC), SC, Brazil. https://orcid.org/0000-0002-3848-5694*
[2] *Federal University of Santa Catarina (UFSC), SC, Brazil. https://orcid.org/0000-0003-4845-8515*
[3] *Federal University of Ceará (UFCE), CE, Brazil. https://orcid.org/0000-0001-6016-745X*
[4] *Federal University of Santa Catarina (UFSC), SC, Brazil. https://orcid.org/0000-0001-8426-2080*
[5] *São Paulo State University (UNESP), SP, Brasil. https://orcid.org/0000-0001-6191-2001*
[6] *Federal University of Santa Catarina (UFSC), SC, Brazil. https://orcid.org/0000-0003-3186-8286*

Autor Correspondente: **Greici Capellari Fabrizzio**
e-mail: **greicicapellari@hotmail.com**

Artigo recebido: 28/03/2023
Aprovado: 21/11/2024
https://jhi.sbis.org.br/

INTRODUCTION

SARS-CoV-2 has changed its behavior; new variants have emerged and drugs and new vaccines have been discovered. Patients infected before the vaccination required 5 to 25% Intensive Care Unit (ICU), which increased the demand for ICU beds, mechanical ventilators, and material and human resources(1–3).

The predictive analysis of precision can aid in complex clinical decisions that converge to the early identification of patients with risk of worsening in the clinical picture. Decision models created from machine learning algorithms prove to be useful in a world sanitary crisis context, in that they can analyze a large amount of data, learn about the relationship between those data, and extract insights. Moreover, the knowledge of patterns of relationship between biomarkers relative to a given outcome, in this case hospitalization in ICU, aids managers and health professionals in their decision-making(4-5-6).

Machine learning algorithms predict which patients are more vulnerable or more resistant to COVID-19 infection and help increase the diagnosis precision and quickness(4–7).When classifying the disease's risk, we should consider physiological and genetic aspects specific to the individual, in that it is assumed that the genetic predisposition to infection by COVID-19 is possible(7).

This search for healthcare personalization, aimed at more assertive interventions, forms the precision nursing precepts basis. The patient's history, including clinical, environmental, and behavioral variables, is collected in the Nursing Process (NP). Patient's clinical condition at that moment is also assessed by nurses, via vital signs and physical examination. The integration of these data associated with large-scale analysis capacity, provides intelligence that supports decision-making in health(8).

This study aimed to apply and compare machine learning algorithms to predict COVID-19 patients' admission to Intensive Care Units from the Precision Nursing theoretical framework.

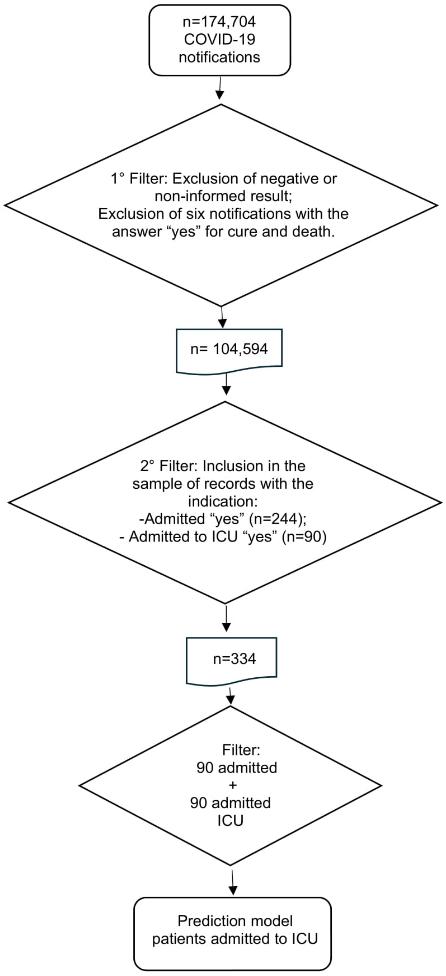
METHODS

A retrospective study was carried out, of the notifications of COVID-19 infected patients in the municipality of Florianópolis, located in the south region of Brazil. Data were collected at the website "covidômetro" (https://sites.google.com/view/gerve) in April 2021 and included COVID-19 notifications from 8 March 2020 to 20 April 2021.

The population of the research was composed of 174,704 patients notified with COVID-19. Although the patients were notified in the municipality of Florianópolis, it is worth stressing that not all the infected lived in the city. The databank deals with COVID-19 notifications; therefore, exclusion criteria were applied for removing from the sample cases negative for COVID-19 or that did not inform the confirmation of the notification, that is, the cases were discarded after the result of examinations or when the notification had not been confirmed. Furthermore, we excluded from the sample six notifications with the answer "yes" for cure and death in the same investigation form. The information contained in the notification form identifies the patient's clinical situation at that moment and we opted to consider situations in which the patients had already obtained an outcome. Therefore, we considered the primary outcomes, or target of interest variable (admission to ICU compared to general admission to hospital).

After the inclusion and exclusion criteria were applied, there remained 90 patients admitted to Intensive Care Units (ICUs), and 90 more patients that had passed through hospital admission were selected randomly to maintain the data balance for modeling. Thus, 180 patients formed the study's sample, as presented in Figure 1.

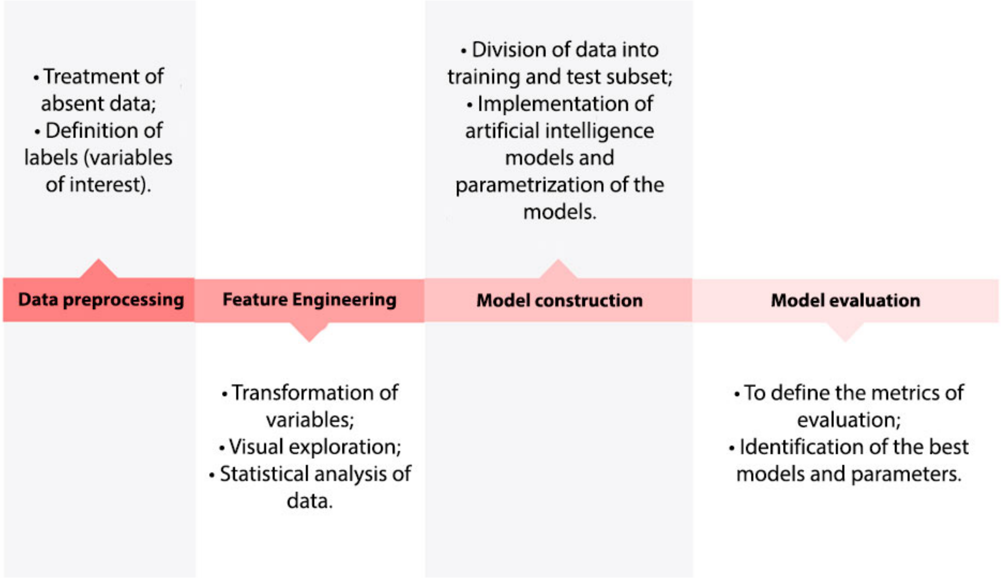
Figure 1 - Flowchart of patient selection and data mining



Construction of the machine learning model

We followed four steps in constructing the model, namely, data preprocessing, feature engineering, construction of the model, and evaluation of the model, as demonstrated in Figure 2.

Figure 2 - Data pipeline



In the phase of data preprocessing, absent values were observed, in the databank, for each of the variables considered in the notification form. The variables with greatest number of absent data were as follows: closure date of the notification form, date of death, health center, test's application date, and neighborhood. For the analysis, the variables with absent data were not relevant for modeling and, therefore, were removed from the sample.

When the predictive model was built, there was no consensus in the scientific literature as to which predictor variables (attributes) were most important for predicting hospitalization due to COVID-19 in the Intensive Care Unit. Nor was it described in the scientific literature whether comorbidities are variables that have a greater influence on mortality or the need for intensive care, when compared to clinical variables such as biomarkers and phenotype(2).

Therefore, attributes that had already been used in previous studies were considered(1–3) and which appeared on the COVID-19 notification form. However, even between studies, some variables differ due to the availability and cost of collection. From the selection of input variables and association with elements of precision nursing, we included the variables corresponding to phenotype, genotype, and biomarkers, which are described below:

Among the variables corresponding to phenotype (demographic information), we considered the following variables: age group in years, distributed into 10 categories (up to 9 years, 10 to 19 years, 20 to 29 years, 30 to 39 years, 40 to 49 years, 50 to 59 years, 60 to 69 years, 70 to 79 years, 80 to 89, and above 90); sex (female and male); and race (yellow, white, brown/pardo, and black). As far as phenotype is concerned (comorbidities and pregnancy): uncompensated respiratory disease; chronic heart disease; diabetes; advanced-stage renal disease; immunosuppression; and high-risk pregnancies. As regards biomarkers, signs and symptoms, such as throat pain, dyspnea, fever, and cough were considered. We did not evaluate variables corresponding to epigenetics.

The original data consist of categorical variables, the age group and race variables, categorized into more than two values and the other variables categorized into yes or no. These categories were transformed into numerical values by means of the One-hot Encoding where each variable is transformed into a binary variable (zero or one). The modeling of data was implemented using the following algorithms: multilayer perceptron (MLP) - neural network, AdaBoost, logistic regression, random forest, kNN, Naive Bayes, SVM, and decision tree. The parametrization of models occurred in an interactive process and selected the best performing algorithms using to the evaluation metrics.

Data analysis and validation of the model

Data analysis was performed in the open-source software Orange Data Mining and Python. Models were validated by means of data cross-validation techniques, observing the area under the curve, sensitivity, specificity, and confusion matrix metrics.

Ethical aspects

The databank used in the study refers to public domain data, anonymous, with no information making the identification of patients possible. In this way, the research did not require approval from the human research ethics committee.

RESULTS AND DISCUSSION

Of the patients admitted to hospital and those admitted to ICU, 56.11% were males, 158 patients (87.77%) were white, and most patients, 45 (25%), were between 50 and 59 years old. Among the most common symptoms, 95 patients (52%) had cough, 93 patients (52%) had fever, and 76 patients (42%) showed dyspnea. The most frequent comorbidities among patients were heart disease (n= 20, 11.11%), diabetes (n= 19, 10.56%), and respiratory disease (n= 8, 4.44%). The other variables are described in Table 1.

Table 1 – Baseline demographic data and clinical data of the study patients at the time of completion of the notification form

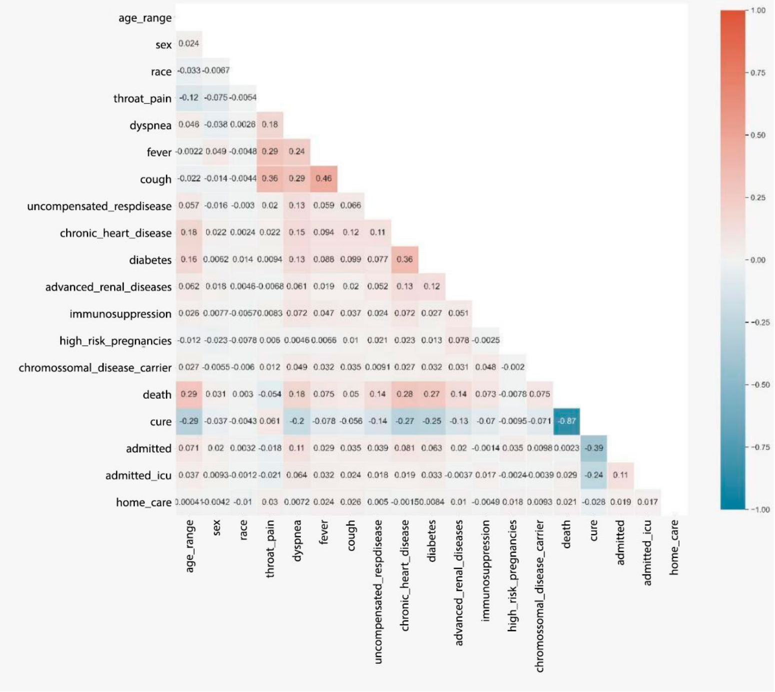
Variables (n=180)	n (%)
Sex	
Male	101 (56.11)
Female	79 (43.89)
Age group	
50 to 59 years	45 (25)
60 to 69 years	32 (17.7)
40 to 49 years	29 (16.11)
70 to 79 years	25 (13.88)
30 to 39 years	20 (11.11)
80 to 89 years	14 (7.77)
20 to 29 years	8 (4.44)
Below 19 years	4 (2.22)
Above 90 years	3 (1.66)
Race	
White	158 (87.77)
Black	13 (7.22)
Yellow	7 (3.88)
Brown/Pardo	2 (1.11)
Signs and Symptoms	
Cough	95 (52)
Fever	93 (52)
Dyspnea	76 (42)
Throat pain	24 (13)
Comorbidities	
Heart disease	20 (11.11)
Diabetes	19 (10.56)
Respiratory disease	8 (4.44)
Immunosuppression	2 (1.11)
Renal disease	1 (0.56%)
Chromosomal disease	1 (0.56%)
High-risk pregnancy	1 (0.56%)

An earlier study carried out with 10,504 patients positive for COVID-19 and admitted to ICU, also identified the cough (n= 5181, 49.7%), fever (n= 4849, 46.5%), and dyspnea (n= 3246, 31.1%) symptoms as the most frequent for these patients. As far as comorbidity is concerned, diabetes (n=1613, 15.5%) was the most recurrent, followed by obesity (n= 917, 8.8%) and chronic respiratory disease (n= 883, 8.5%)(9). Diabetes and respiratory disease are also among the most prevalent comorbidities in the present study; however, while heart disease appears as the most prevalent comorbidity in this study, in the literature it is shown in fourth place, after diabetes, obesity, and chronic respiratory disease. The obesity variable mentioned in the literature, was not evaluated in the present study.

The coefficient of correlation between the variables is highlighted in the bidimensional correlation matrix demonstrated in Figure 3, where it is possible to observe the positive and negative correlations between variables, in which cough and fever showed the highest positive correlation (0.46), followed by diabetes and chronic heart disease (0.36); cough and throat pain (0.36); death and age range (0.29); fever and throat pain (0.29); cough and dyspnea (0.29); death and chronic heart disease (0.28); death and diabetes (0.27); fever and dyspnea (0.24); death and dyspnea (0.18); chronic heart disease and age group (0.18); dyspnea and throat pain (0.18), and so on.

Among the variables that showed a negative correlation are cure and death (-0.87), followed by admission and cure (-0.39); cure and age group (-0.29); cure and chronic heart disease (-0.27); cure and diabetes (-0.25); admission to ICU and cure (-0.24); cure and dyspnea (-0.20); cure and uncompensated respiratory disease (-0.14); cure and advanced-stage renal diseases (-0.13); cure and immunosuppression (-0.07), and so forth.

Figure 3 - Bidimensional correlation matrix



The variables predicting admission to ICU with highest impact on the model predictions according to the presentation in Table 2, included hospital admission, race, and throat pain, ranked by the Gini impurity index(10). We opted for hospital admission as a variable predictive of admission to ICU because not only patients already admitted to ICU require care at these units. Many patients not admitted to a clinical facility arrived for hospital care having as a doorway the Emergency Medical Services (SAMU) and/or the emergency department of hospitals, already showing clinical conditions of need for care in an ICU bed.

Table 2 – Ranking of variables

Variable	Gini decrease
Admission	0.263
Race	0.019
Throat pain	0.010
Fever	0.005
Age group	0.005
Immunosuppression	0.004
Chromosomal disease	0.003
High-risk pregnancies	0.003
Advanced-stage renal diseases	0.003
Chronic heart disease	0.003
Dyspnea	0.001

As regards the race predictive variable, the obtained result cannot be generalized, the sample was not representative of the population; there were 158 (87.77%) patients of the white race. The sample size shows some limitations related to the high-risk pregnancy variable that was not influential on the model.

The predictive variables identified in the present study differ from previous findings identified in the literature. In one study, age, temperature, and tachypnea were the most parsimonious predictive variables for patients' admission to ICU(9). Another study described respiratory insufficiency, shock, inflammation, and renal insufficiency as highest-risk predictive variables for patients transfer to ICU. These studies evaluate different biomarkers which confers unique specificities upon the research(1).

An interesting biomarker to be analyzed considering the advance of vaccination against COVID-19 would be the vaccination status of patients as to its capacity of predicting admission to ICU. A study indicated a marked reduction in severe cases of COVID-19 after vaccination, while this reduction arrived at 65.6% in vaccinated patients admitted to ICU. These numbers may vary according to immunological medicinal product used and time interval between applications; vaccination coverage rates; and performance of complete vaccination scheme and booster dose(11).

Patients with more than one comorbidity fully vaccinated were 48.8% less likely to be admitted to the ICU, 55.4% less probability to use mechanical ventilators, and 22.6% less likely to die(12). In addition, they were hospitalized for significantly shorter periods while unvaccinated patients were absent from work for longer periods compared to partially or fully vaccinated patients(12).

The variables predicting admission of COVID-19 infected patients to ICU were related to signs and symptoms biomarkers and given comorbidities, which make up the patients' previous history. This identification of the subjacent biological mechanisms preceding severe cases of COVID-19 is nursing's scope of practice. Thus, the knowledge of predictive variables may aid in the clinical practice of nurses by aiding in clinical reasoning and decision-making, in a context that calls for a quick and at the same time reasoned decision(13).

The association of biomarkers with disease enables identifying alterations in the organism's homeostasis. Precision Nursing involves precisely identifying the best predictors for COVID-19, a task associated with the artificial-intelligence-centered data analysis capacity, and from this planning nursing interventions that can benefit the patient(13-14).

Precision Nursing can be defined as the broad understanding of the environment where the individual is, his life habits, previous diseases, signs and symptoms, and from this identification, the planning of interventions aligned with each patient's individual needs(13-8). Therefore, Precision Nursing is aligned with the Nursing Process and may be applied to each of its stages: Nursing Data Collection, Nursing Diagnosis, Nursing Planning, Nursing Implementation and Evaluation.

The performance of the admission to ICU prediction model is presented in Table 3 and ROC can be observed in Figure 4 and the confusion matrix in Figure 5.

Table 3 – Performance of algorithms

Algorithms	AUC	Sensitivity	Specificity
Multilayer perceptron	0.917	0.861	0.825
AdaBoost	0.911	0.822	0.800
Logistic regression	0.911	0.806	0.862
Random forest	0.896	0.828	0.850
kNN	0.893	0.806	0.750
Naive Bayes	0.886	0.833	0.962
SVM	0.875	0.817	0.850
Decision tree	0.857	0.811	0.925

Figure 4 - Receiver Operating Characteristics (ROC) for the multilayer perceptron model

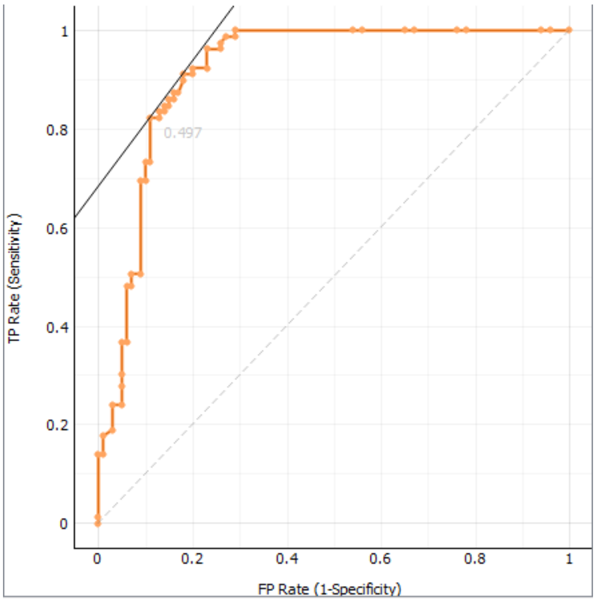
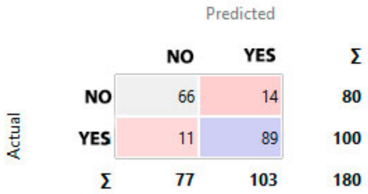


Figure 5 - Confusion matrix for the multilayer perceptron model



		Predicted		
		NO	YES	Σ
Actual	NO	66	14	80
	YES	11	89	100
	Σ	77	103	180

The multilayer perceptron artificial neural network was the model that obtained the best performance in predicting the COVID-19 patients' admission to ICU. Parameterized at 50, 200, and 350 neurons in the hidden layers, the activation function used was tanh, optimization method was Adam, L2 regularization was applied with $\alpha = 0.6$, and the maximum number of admissions was set at 200 times.

CONCLUSION

This application proved to be a viable method for predicting the admission of COVID-19 infected patients to ICU. The predictor variables that most influenced the model were hospital admission, race and throat pain. The neural network model achieved better prediction for AUC (0.917), sensitivity (0.861), and specificity (0.825) and it was accurate in predicting, 155 cases out of 180 were correctly predicted in their outcomes.

The emergence of new variants and the inclusion of the vaccination variable make data generation nonstationary. Thus, the associations that hold the data together change over time so we cannot match all data with more recent data without careful evaluation. These models require constant evaluation as to their performance. This model was applied in a specific population, with given health characteristics, which limits the generalization potential of the model. We suggest the reapplication of the study in other populations.

These clinical biomarkers prove to be relevant for the clinical practice of Nursing in patients with COVID-19 because they are easily observable and can be quickly implemented to the identification of patients with a probability of decline in their clinical status. Thus, the model positively influences the practice of care, as it provides information for the decision-making of health professionals and managers, in situations involving need for ICU beds and quick mobilization of materials and human resources.

The present study shows some limitations regarding the quality of data, many patients did not yet have a positive or negative result for the disease at the time of form closure, or the forms had not yet been completed which influences the accuracy of machine learning models as they rely on consistent data analysis. Another limitation of the research is that it only considered variables that appear on the COVID-19 notification form. Obesity, for example, was considered an important comorbidity in the scientific literature, but was not assessed in the research.

ACKNOWLEDGEMENTS

We thank the National Council for Scientific and Technological Development (CNPq), which provided a doctoral scholarship that contributed to the development of this research.

REFERÊNCIAS

1. Cheng FY, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. Journal of Clinical Medicine 2020, Vol 9, Page 1668 [Internet]. 2020 Jun 1 [cited 2022 Jan 20];9(6):1668. Available from: https://www.mdpi.com/2077-0383/9/6/1668/htm
2. Hou W, Zhao Z, Chen A, Li H, Duong TQ. Machining learning predicts the need for escalated care and mortality in COVID-19 patients from clinical variables. Int J Med Sci [Internet]. 2021 [cited 2022 Jan 20];18(8):1739. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7976594/
3. Covino M, Sandroni C, Santoro M, Sabia L, Simeoni B, Bocci MG, et al. Predicting intensive care unit admission and death for COVID-19 patients in the emergency department using early warning scores. Resuscitation [Internet]. 2020 Nov 1 [cited 2022 Mar 20];156:84–91. Available from: www.sciencedirect.com
4. Cheng J, Cui J, Yu W, Kang H, Tian Y, Jiang X. Factors influencing nurses' behavioral intention toward caring for COVID-19 patients on mechanical ventilation: A cross-sectional study. PLoS One [Internet]. 2021 Nov 1 [cited 2022 Jan 20];16(11):e0259658.

Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0259658

5. Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine Learning in oncology: A clinical appraisal. Cancer Lett. 2020 Jul 1; 481:55–62.
6. Sá GCB, Silva AV, Lima NCA, Nascimento SM. Predicting bed occupancy rate of ICU's in brazilian regions considering vaccination. J. Health Inform [Internet]. 2022 Apr-Sep [cited 2023 Mar 27];14(2):89-95. Available from: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/919/508
7. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight covid-19. Physiol Genomics [Internet]. 2020 Apr 1 [cited 2022 Jan 20];52(4):200-2. Available from: https://journals.physiology.org/doi/abs/10.1152/physiolgenomics.00029.2020
8. Yuan C. Precision Nursing: New Era of Cancer Care. Cancer Nurs. 2015;38(5):333–4.
9. Izquierdo JL, Ancochea J, Soriano JB. Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. J Med Internet Res [Internet]. 2020 Oct 1 [cited 2022 Jan 20];22(10). Available from: /pmc/articles/PMC7595750/
10. Kwong JC, Khondker A, Kim JK, Chua M, Keefe DT, dos Santos J, et al. Posterior Urethral Valves Outcomes Prediction (PUVOP): a machine learning tool to predict clinically relevant outcomes in boys with posterior urethral valves. Pediatric Nephrology [Internet]. 2021 Oct 22 [cited 2023 Fev 5];1:1–8. Available from: https://link.springer.com/article/10.1007/s00467-021-05321-3
11. Moghadas SM, Vilches TN, Zhang K, Wells CR, Shoukat A, Singer BH, et al. The impact of vaccination on COVID-19 outbreaks in the United States. medRxiv [Internet]. 2020 Nov 30 [cited 2022 Jan 20]; Available from: /pmc/articles/PMC7709178/
12. Maltezou HC, Basoulis D, Bonelis K, Gamaletsou MN, Giannouchos T v., Karantoni E, et al. Effectiveness of full (booster) COVID-19 vaccination against severe outcomes and work absenteeism in hospitalized patients with COVID-19 during the Delta and Omicron waves in Greece. Vaccine [Internet]. 2023 [cited 2023 Feb 19]; Available from: https://pubmed.ncbi.nlm.nih.gov/36740558/
13. Corwin EJ, Ferranti EP. Integration of biomarkers to advance precision nursing interventions for family research across the life span. Nurs Outlook. 2016;64(4):292–8.
14. Fu MR, Kurnat-Thoma E, Starkweather A, Henderson WA, Cashion AK, Williams JK, et al. Precision health: A nursing perspective. Int J Nurs Sci [Internet]. 2020 Jan 10 [cited 2022 Jan 26];7(1):5. Available from: /pmc/articles/PMC7031154/