A Simple Artificial Neural Networks to Determine Vaccination Priority

Wisang Sugiarta[a], Mike Benigeri[a] and Ahmed Ghachem[a]
[a] Institut national d'excellence en santé et services sociaux, Montreal, Canada

Keywords: Machine Learning | Vaccination | Hospitalizations

Abstract

Background: The 2019 novel coronavirus (COVID-19) has created countless logistic problems for governments to solve. However, in light of the fastest vaccination development and approval process in history, people are looking forward towards the end of the pandemic. Yet, the virus still presents a large risk as Quebec's hospitals are fighting to keep their heads above water. Moreover, most governments are assigning vaccination priority based almost solely on age, even though research suggests there are a numerous other important factors.

Objectives: The objective of this study is to develop a machine learning model that can assign vaccination priority better than the baseline of only age in Quebec in order reduce the number of hospitalizations. We aim to use simple features including only a medical profile and history as well as some socio-demographic features.

Methods: We develop a very simple 12-layer artificial neural network to complete the classification assignment of identifying people who will be hospitalized if they contract COVID-19. The training data we used was all compiled from diagnosed patients from September 15th to January 4th and whether those patients were hospitalized or not. We then run the model on the entire population and the model will determine the cases where the outcome is a hospitalization and thus determine the people who should be prioritized for vaccinations.

Results: We show it is possible to create models that are able to target individuals whose vaccination should be prioritized. The results of this study show that we are able to classify 78% of all cases leading to a hospitalization as well with a precision of 28%. On average, this allows us to prevent one hospitalization for every 4.6 people we vaccinate on average.

1 Introduction

In 2019, the novel coronavirus (COVID-19) transformed the daily life for nearly every person. It marked the first pandemic in nearly a century and has amassed hundreds of millions of confirmed cases and millions of deaths (5; 9; 4). However, this sprung researchers from various institutions and disciplines to start the fastest vaccination creation process in history (5). While, at the time of writing, numerous vaccines has been shown to be safe, effective and have been approved by governments, the roll out of these vaccines is proving to be a challenging feat (5). The question of what groups of people should be prioritized is being posed by governments. It has been shown that age along with some highlighted illnesses and other factors can predict serious outcomes in patients (6; 8). There have even been machine learning (ML) models that are able, to a certain degree, predict the outcome after a positive diagnosis. Thus, we aim to see if an ML model can be developed such that we can target specific people in aim to significantly reduce hospitalizations in Quebec.

2 Materials and Methods

Dataset: This study used a merged dataset combining information from the *Régie de l'assurance maladie du Québec*. This database includes information pertaining to medical conditions, hospital admissions and sociodemographic information of all the beneficiaries of provincial health insurance in Quebec. Furthermore, it contains relevant COVID-19 diagnoses, hospitalizations and mortality information. This data is not publicly available.

Study population: The dataset was compiled using all positive COVID-19 positive diagnoses in the time frame ranging from September 14, 2020 to January 4, 2021. We discarded all the positive tests of anyone under 18 years old as they often represented anomalies and skewed results as infants and younger children were often hospitalized out of an abundance of precaution. We also discarded long-term care facility (CHSLD) residents as they are not included in hospitalization or ICU admission counts in Quebec.

Statistical analyses and class imbalance: Since only a small amount of the dataset's were hospitalized 4.9%, it is important to note that a naive classifier, which always classifies points to the majority class, can achieve accuracy greater than 0.95. Thus, accuracy is an inappropriate metric. To counter the heavy imbalance in the datasets, we use a combination of class weighting and oversampling the minority class.

Training data/Testing data: The two datasets were split into 4:1 training and test sets randomly. We did not include the undetermined cases in the training and test sets. We hyper-tuned hyper-parameters using a gird-search for all parameters, including a parameter for class weights and random sampling ratios, in each given model to maximize the combination of sensitivity, PPV and AUC scores using 5-fold cross validation across the training and test data. The models were then trained on full training data with optimized hyper-parameters and tested on the unseen test data.

Machine learning algorithms: The ML algorithm used in this study was an 12-layer artificial neural network (ANN).

3 Results

We first present some characteristics of the dataset. With the exclusions outlined in the methods section, Table 1 outlines simple baseline characteristics as well as some significant illnesses that have been featured in literature. We can see that the mean age of people hospitalized of 71.8 ± 18.3 is far older than the population of people who have contracted COVID, 39.4 ± 22.5. Also note that while most diagnosis are in Montreal, only half of the hospitalizations occur in the same area. The five other variables show a significant increase in the hospitalization category from the more general group, thus we expect them to play a large role in prediction, as literature suggests.

Table 1: Basline Characteristics

Category	Male	Age	Montreal	NPV	Hypertension	Diabetes	Pneumonia	Cancer
Positive Diagnosis	0.48	39.4	62.3%	0.5	6.7%	5.4%	3.5%	3.9%
Hospitalization	0.52	71.8	49.6%	2.3	26.6%	25.7%	18.1%	17.6%

*NPV = Number of Previous Hospitalizations

We next present the final models used for predicting hospitalizations in three different age groups, 18+, 18-69 and 18-59. Clearly, the performance drops considerably as we eliminate older people from the training and test datasets.

Table 2: ANN Models for Predicting Hospitalizations in Different Age Categories.

Age Category	Sensitivity	Specificity	PPV	NPV
18+	0.78	0.89	0.28	0.99
18-69	0.66	0.84	0.14	0.99
18-59	45	79	4	99

*PPV = Positive Predictive Value, NPV = Negative Predictive Value

Finally, Table 3 shows how the model performs versus prioritizing only age. This table contains information on the entire population of Quebec and its statistics throughout the pandemic. This outlines, for a given age group, the number of people, the rate of hospitalizations, the number of hospitalizations and the proportions of population and hospitalizations. Again, note that this study does not include people under 18 years of age. The vaccination by age simulated the process of vaccinating everyone in a certain age group in descending order, while by target model would be the ANN models we created.

Table 3: Vaccination Priority by Age in Quebec

Age Group	Population		Hospitalizations			Vaccination Priority	
	Number	Percent	Rate	Number	Percent	By Age	By Target Model
18-59	4 662 098	67 %	1.5 %	69 799	14 %	66.8	55.6
18-69	5 810 503	83 %	3.0 %	177 099	36 %	32.8	13.8
18+	6 970 607	100 %	7.1%	492 221	100%	14.2	4.6

* The number in the Vaccination Priority column denotes the average number of people needed to be vaccinated in order to prevent one hospitalization by a method.

4 Discussion

Given the results of this study, it is possible to create models that assign individual vaccination priority with a simple ANN. We saw that for 3 different categories, our models were able to prevent more hospitalizations on average than simply vaccinating people based on older age. This process can be used for targeted mass vaccination in complement with vaccination by age. This is significant in the fact that if we consider the model trained and tested on the 18+ age category, we are capable of reducing 78%of all hospitalizations by vaccinating 14% the population. However, it's important to point out that literature shows that above 65, vaccination by age group is effective as well. As an example, in Quebec, we can reduce 64% of hospitalizations by simply vaccinating everyone 70+, who represent 17% of the population. Thus, complementing vaccination by age with out targeted approach could yield significant improvements.

However, it is important to highlight one limitation being that the training and testing of this model was done solely on patients who have already been diagnosed with COVID-19. It is entirely possible that these people are not entirely representative of the actual population. But, for this study we will assume uniformity among the population having been tested positive and the general population. It is also important to highlight that the vaccination's effectiveness is still being studied, so it is cautious to assume some error and that vaccines will not prevent hospitalizations for everyone.

References

[1]

[2] Government of quebec: Data on covid-19 in québec. `https://www.quebec.ca/en/health/health-issues/a-z/2019-coronavirus/situation-coronavirus-in-quebec/`, 2020.

[3] A learning-based model to evaluate hospitalization priority in covid-19 pandemics. *Patterns 1*, 6 (2020), 100092.

[4] World health organization. cumulative number of reported probable cases of sars. `https://www.who.int/csr/sars/country/2003_07_11`, 2020.

[5] World health organization: Worldometer. covid-19 coronavirus pandemic. `https://www.worldometers.info/coronavirus`, 2020.

[6] An, C., L. H. K. D. e. a. Machine learning prediction for mortality of patients diagnosed with covid-19: a nationwide korean cohort study. *Sci Rep 18716* (2020).

[7] Casiraghi, E., Malchiodi, D., Trucco, G., Frasca, M., Cappelletti, L., Fontana, T., Esposito, A. A., Avola, E., Jachetti, A., Reese, J., Rizzi, A., Robinson, P. N., and Valentini, G. Explainable machine learning for early assessment of covid-19 risk prediction in emergency departments. *IEEE Access 8* (2020), 196299–196325.

[8] da Silva, R. G., Ribeiro, M. H. D. M., Mariani, V. C., and dos Santos Coelho, L. Forecasting brazilian and american covid-19 cases based on artificial intelligence coupled with climatic exogenous variables. *Chaos, Solitons Fractals 139* (2020), 110027.

[9] Iftimie, S., López-Azcona, A. F., Vallverdú, I., Hernàndez-Flix, S., de Febrer, G., Parra, S., Hernández-Aguilera, A., Riu, F., Joven, J., Camps, J., Castro, A., and . First and second waves of coronavirus disease-19: A comparative study in hospitalized patients in reus, spain.

[10] Wang P, Zheng X, L. J. Z. B. Predictions of epidemic trends in covid-19 with logistic models and machine learning technics. *Chaos Solitons Fractals 139*, 110058 (2020).