

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

## COVID-19 Time Series Forecasting – Twenty Days Ahead

Kathleen C. M. de Carvalho <sup>a,b</sup>, João Paulo Vicente <sup>a,b</sup>, João Paulo Teixeira <sup>a \*</sup>

<sup>a</sup> Research Centre in Digitalization and Intelligent Robotics (CEDRI), Instituto Politecnico de Bragança, Bragança, Portugal

<sup>b</sup> Federal University of Technology - Paraná (UTFPR), Cornélio Procópio, 86300-000, Brazil

---

### Abstract

The new Coronavirus, responsible for the COVID-19 disease, is the most discussed topic in the current days, and the forecast numbers of new cases and deaths are the most important source of data in governmental decision-making. The present work presents a prediction model with two different approaches concerning the input data, by using Artificial Neural Networks (ANN). The use of a substantial mitigation procedure adopted (mandatory use of masks) was experimented as an input to the network, in order to evaluate the improvement in the results. The ANN forecasting model was demonstrated to predict with higher accuracy within the next twenty days using the information about the mandatory use of face masks. The final results showed that the twenty days ahead forecasting was made with an error of 24,7% and 1,6% for the number of cumulative cases of infection and deaths for Brazil, and 37,9% and 33,8% for Portuguese time series, respectively.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the CENTERIS –International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

**Keywords:** Brazil COVID-19 forecasting; Portugal COVID-19 forecasting; COVID-19 time series; Mitigation procedures; Use of face masks.

---

---

\* Corresponding author. Tel.: +351 273 30 3129.

E-mail address: [joaopt@ipb.pt](mailto:joaopt@ipb.pt)

## 1. Introduction

One of the most issues addressed in 2020 and 2021 is the prevention and control measures that can mitigate the spread of the SARS-COV-2, Coronavirus disease (COVID-19) [1]. Therefore, the knowledge of contamination behavior turns an important focus on researchers around the world. Through this goal, this article proposes an estimation of the number of new cases of infection and deaths in twenty days ahead, in Brazil and Portugal.

The Multi-Layer Perceptron Artificial Neural Network (ANN) model was used to forecast the time series of infections and deaths in both countries. That architecture of the ANN has a set of neurons in the input layer, one or more in the hidden layer, and only one on the output layer [2,3]. Having only one neuron at the output layer, the estimation of cases and deaths was done in separated networks. The feed-forward architecture of the ANN was used. In this architecture, the data flow from the input to the output without any feedback [4].

The ANNs has been widely used as a method for time series forecasting because this tool allows modeling and forecasting based on data behavior [5–9]. The authors in [10] predicted data of new cases and deaths using an ANN multilayer perceptron and the accumulative number of infections and deaths, new infection cases, and deaths. However, the [11] approach uses either ANN MLP, but the behavior of time series is analyzed according to accumulative cases and deaths. This approach was used to forecast new cases of infections in Brazil, Portugal, and USA. In [12], authors forecasted the daily numbers of cases, deaths and ICU.

The present work used the ANN tool to time series forecasting of new cases of infection and deaths of COVID-19 in Brazil and Portugal. Two approaches were defined by the combination of information about the numerical data of infections and deaths, and the use of face mask. The goal of using different approaches is to train the network with different inputs to compare the efficacy of the methods.

The prevention of transmission methods was based on the World Health Organization recommendation. The use of face masks was presented at Interim guidance, published on June 5<sup>th</sup>, 2020 [1] that shows relevant scientific evidence to the use of masks for the prevention of COVID-19 transmission. An additional study, accomplished in Switzerland, presents the efficacy of social distance to decrease the number of infections caused by the Coronavirus [1,13]. However, only the mandatory use of masks was used as qualitative data in that work, since the information about social distance is not clearly presented by all the countries.

The COVID-19 numbers of infection cases and deaths of eleven countries were used to train the ANN for the two approaches. The data of two countries was used to the validation process and the other two countries were chosen for the prediction. The combination between infections and deaths numbers with the prevention action were used as the input data of the ANN. This consists of a combination of numerical and qualitative data. An interpolation process was used to the daily numbers of infections and deaths in order to reduce the effect of lower number of tests during weekend.

The different approach methods are explained in the methodology section. The results of the two approaches are presented in section 3 for the two countries in the test set. Finally, section 4 presents the conclusions.

## 2. Methodology

The proposed methodology used to forecast new infection cases and deaths related to COVID-19 is represented on Fig. 1. The daily numbers of each country are part of a time series that will be pre-processed to normalize the different dimensions between countries. Therefore, a moving average with a short window is applied before the data to be used as an input of the ANN.

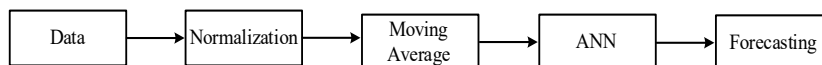


Fig. 1. Diagram of the proposed method

## 2.1. Data collection

The cumulative infection cases and deaths' daily numbers by country were collected from the website Worldmeter [14]. Data of fifteen countries were collected during the period between February 15<sup>th</sup> of 2020 for all countries except for China, where the first data is from January 22<sup>th</sup> of 2020, until January 21<sup>th</sup> of 2021. The data of the last twenty days were used to comparative purposes, and to analyze the network accuracy.

The countries with longer dataset were selected, plus Brazil and Portugal because of the author's interests. These countries are: China, USA, Spain, Germany, France, Iran, Turkey, UK, Russia, Italy, Belgium, Canada, South Korea, Brazil, and Portugal. The data arises from official reports directly by the communication channel through governmental agencies. Other communication resources that also use the same source are Johns Hopkins CSSE, Financial Times, The New York Times, and Business Insider [14]. In addition to the numerical information about infections and deaths numbers, data about the mitigation procedure has been used, specifically the use of face masks. All the data related to the use of masks arose from 'masks4all' [15], which is shown, by country, where the use of face masks in public is required. The time series has the values of 1 or 0, for mandatory or not, respectively. In the case of Brazil, the defined data to mandatory use of masks were based on the governmental decree of the São Paulo state, where locates the pandemic epicenter since the first cases.

## 2.2. Normalization

The absolute numbers of infection cases and deaths are related to the population size of each country, and the infection cases and deaths for each county consist of two time series. The time series of all the fifteen countries were used in the same ANN. Therefore, it was necessary to normalize the time series to the same scale. In the methodical normalization, it was applied the division of all the values by its maximum. Thus, all data are kept in the same range, between 0 and 1, as shown on Fig. 2.

Due to underreporting issues, mainly on the weekends, the accumulated values of cases and deaths may present weekly oscillations. The implementation of a moving average served to mitigate those oscillations. The moving average has a window length of three days in the cumulative numbers.

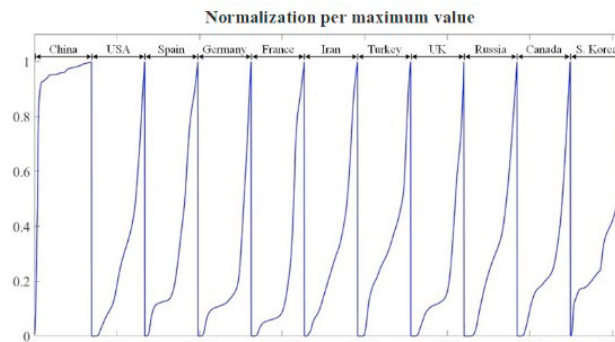


Fig. 2. Normalized curves of infected cases for countries, used for training.

## 2.3. Artificial Neural Network

An ANN was the tool chosen for the prediction of the daily infection cases and deaths with the COVID-19 time series. It was necessary to regroup the data in specific quantity and sequence so that the networks could predict the next day [16], particularly in this work the previous seven days were used to predict the next day, and so on. A longer period of days ahead was forecasted. In this case, the previous days of the ANN entrance were recursively predicted.

The dataset of countries under study was divided into three groups: training, validation, and prediction groups, for training validation and test sets of the ANN. The training phase of the nets holds eleven countries to the ANN learn the behavior of the virus contamination. The validation set hold two countries. Those set were used during the training stage just to stop train iterations avoiding overfitting. The test set were not used during the training process and used

just to the evaluation of the prediction accuracy. Data from China, USA, Spain, Germany, France, Iran, Turkey, UK, Russia, Canada, and South Korea were used in the training set. Italy and Belgium in the validation set, and Brazil and Portugal in the test set.

Two separate ANN were used to predict the number of infected cases and deaths. The ANN was tested using different architectures and training functions separately to find the best results for both cases. The final architecture consists of seven neurons in the hidden layer, one neuron in the output layer, the linear activation function in both layers, and Levenberg-Marquardt as the training function algorithm.

For comparative analyses, it was necessary to de-normalize the predicted data, which means multiplying predicted values with the maximum factor used in each case (country). The results and the comparative analyses are presented in the next section.

In order to compare the two proposed approaches, the relative percentage error for the total number of cases and deaths within the period of twenty predicted days was used. The expression that represents this calculation is given by the Eq. 1.

$$E_{\%} = \frac{|V_{\text{real}} - V_{\text{pred}}|}{V_{\text{real}}} \times 100\% \quad (1)$$

### 3. Results and Discussion

To analyze the proposed ANN, two approaches were implemented. In the first, the network's input data were only numerical, with the number of cumulative infected cases and death. The second approach carried out using the numerical data, utilized in the first approach, was combined with the information about the mitigation procedure, the use of face masks.

#### 3.1. First approach: numerical data only

Fig. 3 shows the behavior of the curve generated by the ANN models for the cumulative number of infected cases and deaths, respectively, in Brazil, considering only numerical data in the input. The zoom presented in both right-side figures indicates the range where the ANN predicted new data (twenty days head).

In Fig. 3, the yellow curve represents the real data, of infection cases and deaths. The blue circles are the one-day ahead prediction, that uses the real seven days previous data in the input. On the other hand, the black points represent the prediction for one-to-twenty days ahead, using real previous seven days (n-7 to n-1), and then successive predicted previous days. The red vertical line points out the beginning date of the mandatory use of face masks.

It is possible to observe the cumulative daily error of right-side plots since the forecasting methodology uses the seven days prior to the output day to predict the values of infected cases or deaths. This methodology is assumed for the two approaches and the two predictions, cases, and deaths. It is important to highlight that the day n represents January 2<sup>th</sup> followed by the next days, until January 21<sup>th</sup> of 2021.

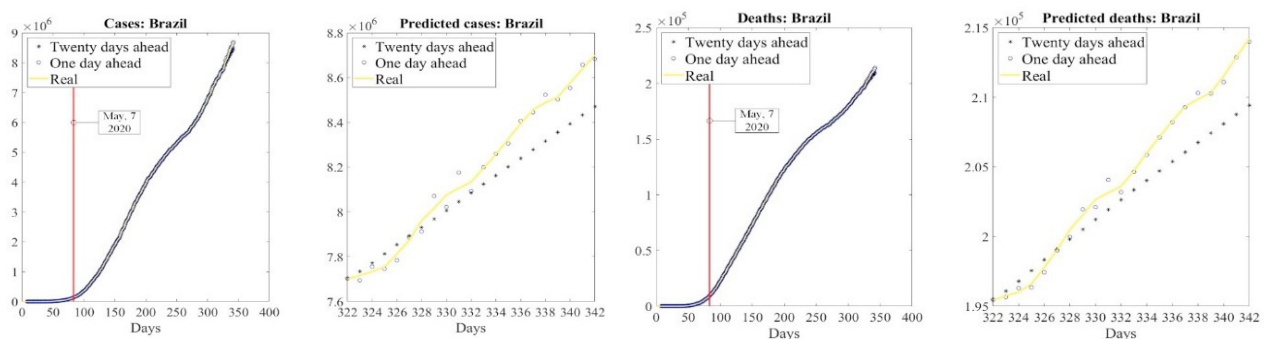


Fig. 3. First approach - Brazil infected cases and deaths

The same methodology was used for Portugal's time series. As follows, Fig. 4 shows the outputs of ANN used to forecast cases and deaths, respectively, for the same period of days.

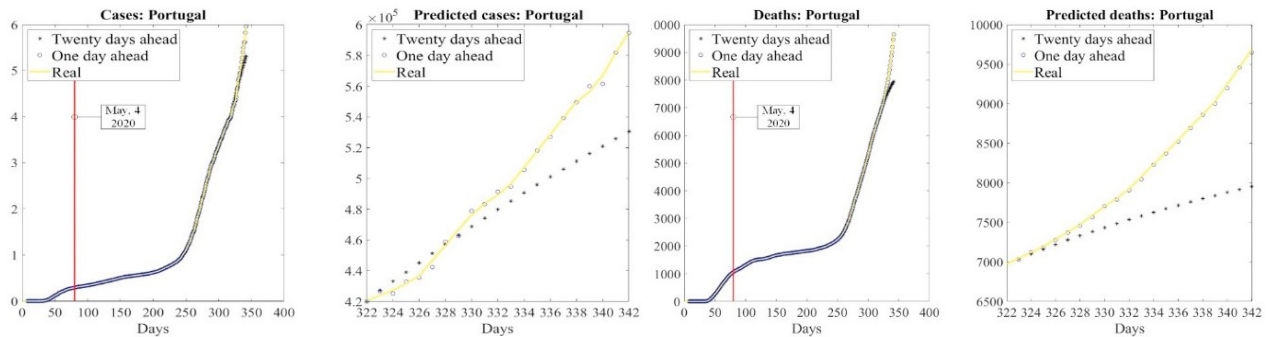


Fig. 4. First approach - Portugal cases and deaths

### 3.2. Second approach: numerical data and use of face masks

For the second approach, the input data includes the previous seven days of cumulative numbers of infected cases and deaths plus the information about the use of face masks in public buildings. Fig. 5 and Fig. 6 show the prediction to Brazil and Portugal, respectively, according to the inputs of the second approach.

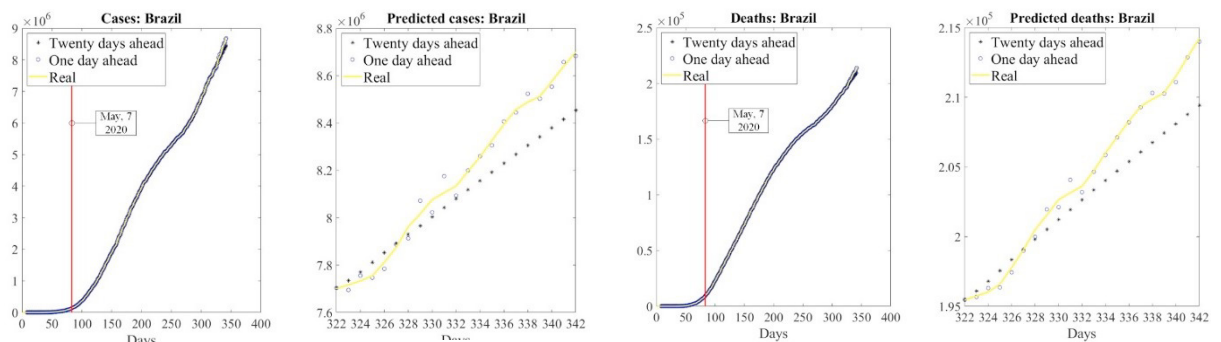


Fig. 5. Second approach - Brazil infected cases and deaths

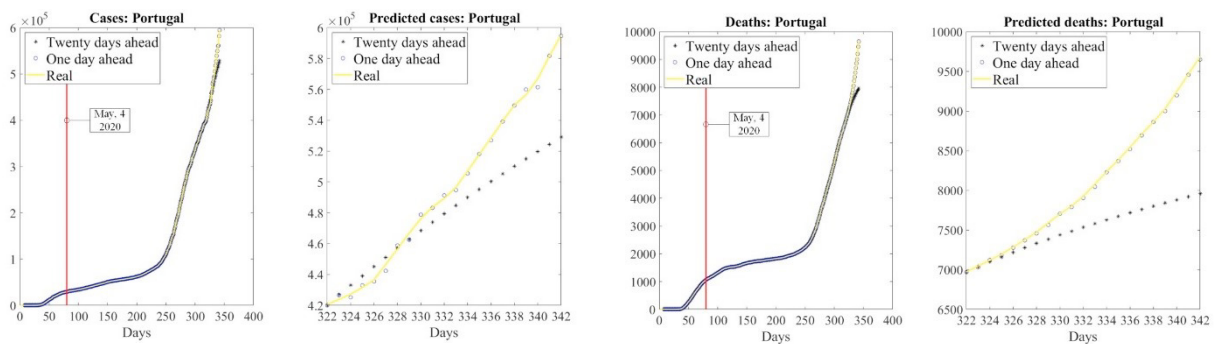


Fig. 6. Second approach - Portugal infected cases and deaths

The use of the qualitative data about the mandatory use of masks results in an improvement of relative error in both countries for deaths forecasting, for twenty days ahead.

### 3.3. Discussion

In order to verify the accuracy of ANN to predict cases and deaths twenty days ahead, the total values of real and forecasted infections and deaths (for the twenty days ahead values of day  $(n+19)$  – day  $(n-1)$ ) were compared by calculation of relative percentage error, according to Eq. 1. On Table 1 a comparison of relative errors for the two proposed approaches is presented.

Table 1. Comparative table

<b>First Approach: totals during twenty days predicted</b>						
	<b>Cases</b>			<b>Deaths</b>		
	<b>Real</b>	<b>Predicted</b>	<b>Relative error</b>	<b>Real</b>	<b>Predicted</b>	<b>Relative error</b>
<b>Brazil</b>	999.236	770.637	22,9%	18.787	13.994	25,5%
<b>Portugal</b>	174.520	109.838	37,1%	2.714	981	63,9%
<b>Second Approach: totals during twenty days predicted</b>						
	<b>Cases</b>			<b>Deaths</b>		
	<b>Real</b>	<b>Predicted</b>	<b>Relative error</b>	<b>Real</b>	<b>Predicted</b>	<b>Relative error</b>
<b>Brazil</b>	999.236	752.266	24,7%	18.787	19.093	1,6%
<b>Portugal</b>	174.520	108.415	37,9%	2.714	1.797	33,8%

The analysis using relative error was chosen only in the analyses of the prediction for twenty days ahead, presented in Table 1, because the high value of the cumulative time series induces an unreal low daily relative error.

Table 1 shows a relatively higher error for twenty days ahead for Portugal than for Brazil. This can be understood because Portugal suffers an unexpected increase of infected cases and deaths during the period of the twenty days of prediction that exceeded all the number of infected cases by 100.000 inhabitants in all Europe [17]. This increase may be due to the absence of social restrictions for the Christmas season in addition to the surprisingly high number of cases with the COVID-19 UK variant (VOC-202012/01Alpha), assumed as much more contagious. This variant is also 64% more deadly as published by Robert Challen, et al. [18] the Alpha SARS-CoV-2 variant causes 4,1 deaths against 2,5 deaths of the classic SARS-CoV-2 over 1000 infected cases. The authors stated, “The mortality hazard ratio associated with infection with VOC-202012/1 compared with infection with previously circulating variants was 1.64”.

The relative error reduced severely in the forecasting of the number of deaths with the usage of face masks mandatory information. For the Brazilian case, it reduces from 25,5% to 1,6%, and for the Portuguese case, the error reduced from 63,9% to 33,8%.

## 4. Conclusions

This paper presented the developments and results of two ANN based models to predict twenty days ahead of the time series of the number of infected people and deaths caused by the Coronavirus disease (COVID-19). The models were used to predict the time series for two countries (Brazil and Portugal) with very different perspectives to control the pandemic transmission. It was aimed to experiment with the combination of quantitative and qualitative data in order to compare the results of each mitigation procedure, and the combination between the mitigation procedure and the mandatory use of face masks.

The first approach, only related to numerical data, presented a satisfactory result, in forecasting Brazil’s tendency since the behavior of the infections is well determined. However, in Portugal’s outputs, the ANN had an increase of errors, which can be explained by the unexpected growth of records of infections in the country after Christmas holidays. It was partially due to the new Alpha variant of COVID-19, which represented, at that time, 40% of total cases in the country and 60% of cases in *Lisboa e Vale do Tejo* region [18].

By the second approach, in which was implemented qualitative data about the use of masks, an important reduction of errors in death prediction was noticed.

The final results showed that the twenty days ahead forecasting was made with an error of 24,7% and 1,6% for the number of cumulative cases of infection and deaths for Brazil, and 37,9% and 33,8% for Portuguese time series, respectively.

Finally, given the results, it is highly recommended to use information about the implemented mitigation procedures to predict the evolution of the pandemic contamination.

As future work, the study of infections considering the progress of vaccination as another input data is suggested.

## Acknowledgements

This work was supported by Fundação para a Ciência e Tecnologia within the Project Scope: UIDB/05757/2020.

## References

- [1] World Health Organization, Advice on the use of masks in the context of COVID-19, World Heal. Organ. (2020) 1–16. <https://doi.org/10.1093/jiaa077>.
- [2] C. Felisoni De Angelo, R. Zwicker, N.M.M. Dias Fouto, M.R. Luppe, Séries temporais e redes neurais: uma análise comparativa de técnicas na previsão de vendas do varejo brasileiro. (Portuguese), Brazilian Bus. Rev. (Portuguese Ed. 8 (2011) 1–21.
- [3] J.P. Teixeira, P.O. Fernandes, Forecasting of a Non-Seasonal Tourism Time Series with ANN, Proc. 14th Int. Work. Time Ser. – ITISE. (2014) 724–734. [http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2&SrcAuth=ORCID&SrcApp=OrcidOrg&DestLinkType=FullRecord&DestApp=WOS\\_CPL&KeyUT=WOS:000359136600081&KeyUID=WOS:000359136600081](http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2&SrcAuth=ORCID&SrcApp=OrcidOrg&DestLinkType=FullRecord&DestApp=WOS_CPL&KeyUT=WOS:000359136600081&KeyUID=WOS:000359136600081).
- [4] A. Azadeh, S.F. Ghaderi, S. Sohrabkhani, Forecasting electrical consumption by integration of Neural Network, time series and ANOVA, Appl. Math. Comput. 186 (2007) 1753–1761. <https://doi.org/10.1016/j.amc.2006.08.094>.
- [5] M. Qi, G.P. Zhang, An investigation of model selection criteria for neural network time series forecasting, Eur. J. Oper. Res. 132 (2001) 666–680. [https://doi.org/10.1016/S0377-2217\(00\)00171-5](https://doi.org/10.1016/S0377-2217(00)00171-5).
- [6] G.P. Zhang, M. Qi, Neural network forecasting for seasonal and trend time series, Eur. J. Oper. Res. 160 (2005) 501–514. <https://doi.org/10.1016/j.ejor.2003.08.037>.
- [7] J.P. Teixeira, D. Freitas, Segmental durations predicted with a neural network, EUROSPEECH 2003 - 8th Eur. Conf. Speech Commun. Technol. (2003) 169–172.
- [8] M. Wiecek, J. Silka, M. Wozniak, Neural network powered COVID-19 spread forecasting model, (2020). <https://doi.org/10.1016/j.chaos.2020.110203>.
- [9] S.K. Tamang, P.D. Singh, B. Datta, Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique, Glob. J. Environ. Sci. Manag. 6 (2020) 53–64. <https://doi.org/10.22034/GJESM.2019.06.S1.06>.
- [10] P.H. Borghi, O. Zakordonets, J.P. Teixeira, A COVID-19 time series forecasting model based on MLP ANN, Procedia Comput. Sci. 181 (2021) 940–947. <https://doi.org/10.1016/j.procs.2021.01.250>.
- [11] L.S. de Oliveira, S.B. Gruetzmacher, J.P. Teixeira, COVID-19 Time Series Prediction, Procedia Comput. Sci. 181 (2021) 973–980. <https://doi.org/10.1016/j.procs.2021.01.254>.
- [12] K. Carvalho, J.P. Vicente, M. Jakovljevic, J.P. Teixeira, Forecasted Incidence, Intensive Care Unit Admissions and Projected Mortality attributable to Covid-19 in Portugal, UK, Germany, Italy and France – 4 weeks Ahead, (2021). <https://doi.org/10.20944/PREPRINTS202105.0116.V1>.
- [13] Michel Bielecki, R. Züst, D. Siegrist, D. Meyerhofer, G.A.G. Crameril, Z.G. Stanga, A. Stettbacher, T.W. Buehrer, J.W. Deuel, Social distancing alters the clinical course of COVID-19 in young adults: A comparative cohort study, Oxford Univ. Press Infect. Dis. Soc. Am. (2020). <https://doi.org/10.1093/gerona/gly169/5057054>.
- [14] Worldometer, COVID-19 Coronavirus Pandemic, (n.d.). <https://www.worldometers.info/coronavirus/>.
- [15] Masks4all, What Countries Require Public Mask Usage To Help Contain COVID-19?, (n.d.). <https://masks4all.co/pt/what-countries-require-masks-in-public/>.
- [16] I.N. da Silva, D.H. Spatti, R.A. Flauzino, L.H.B. Liboni, S.F. dos R. Alves, Artificial Neural Networks: Apractical course, Springer International Publishing, 2017.
- [17] T.S. Serafim, Presença da variante do Reino Unido é estimada em cerca de 60% em Lisboa, Público. (2021).
- [18] R. Challen, E. Brooks-Pollock, J.M. Read, L. Dyson, K. Tsaneva-Atanasova, L. Danon, Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study, BMJ. 372 (2021) 1–10. <https://doi.org/10.1136/bmj.n579>.