XLI CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU PARANÁ - BRAZIL

Prediction of Covid-19 contagion in the State of Maranhão using Neural Networks and the SIR epidemiological model

Luis Fillype da Silva[1], Caio Magno Carvalho[1], Davi Nascimento[1], Marta Barreiros[1], Jonathan Araujo Queiroz[1], Allan Kardec Barros[1]

[1]*Dept of Electrical Engineering, Federal University of Maranhão*
São Luís, Maranhão, Brazil
silvaluis_@outlook.com;magno.caio91@gmail.com;davi.nascimento777@gmail.com;marta-barreiros@hotmail.com; prof.queiroz.jonathan@gmail.com;akduailibe@gmail.com

Abstract. Coronavirus disease 2019 (COVID-19) is a public health emergency of international interest, being declared by the World Health Organization (WHO) as a pandemic, and has instituted essential measures for prevention and confrontation. Because it has high rates of transmission and rapid dissemination around the world, medical authorities worldwide have realized its potential danger and the possible collapse of the Health Systems of the affected countries. Therefore, specific predictive methods are urgently needed to predict the risk of COVID-19 and to assist in the decision-making process in the control of epidemics and / or pandemics. Here, the number of cases for the State of Maranhão in the period from 27/03 to 27/05 of this year was estimated, with observations up to the Lockdown decree using the epidemiological model SIR (Susceptible-Infected-Recovered) and a neural network Multilayer Perceptron (MLP). The samples were obtained through Kaggle, collected from the Ministry of Health. In order to compare the algorithms, the population was divided into three groups: susceptible, infected and recovered. MLP averaged 5507.95 cases, while the actual number was 5787.56. An epidemiological mathematical model estimated the number of cases with a larger lag (about 1.8 million cases). The results obtained by the proposed method can be used to support COVID-19 decision making. The neural network has shown to have more reliability in the results and is closer to the real cases disclosed until May 17, 2020.

Keywords: SIR. COVID-19. Neural networks. Prediction.

1 Introduction

At the end of 2019, a new infectious viral disease called COVID-19 (Coronavirus Disease 2019) was caused by a virus called the SARS-CoV-2 coronavirus (Severe Acute Respiratory Syndrome Coronavirus) [1]. Because it has high rates of transmission and rapid spread of the disease around the world, medical authorities worldwide have realized its potential danger and the imminent collapse of the Health Systems in the affected countries. To assist in prediction and analysis of epidemics such as COVID-19, mathematical models of epidemiological nature are used, as they are quite efficient in assisting the decision-making process, as well as their application provides an analysis of the development of the observed epidemic, which may represent an important tool for decision makers. Since the detection of the outbreak of COVID-19, several scientific studies have been published with the aim of analyzing the behavior of the epidemic in the population, as pointed out in [4]. There are basically two ways of analyzing this type of phenomenon, using mathematical (analytical) models and using computer simulation models. There are several studies already published dealing with the evolution of the epidemic and the comparison between possible scenarios, aimed at a specific country, but there is still a need for simulation models to predict the cases of COVID-19 in the states of Maranhão. The Study pointing at [3] uses the model proposed in this work to make predictions in the states of São Paulo, Ceará, Minas Gerais. Thus, this research predicted the cases of infected by COVID-19 using neural networks and the epidemiological model SIR (Susceptible-Infected-Recovered) in three different approaches: without measures adopted in the state, with social distance and with social distance and quarantine.

2 Estimation models

Mathematical models can predict or simulate the most diverse types of infectious diseases that affect the human population, whether at the community level (local epidemic) or worldwide (pandemic), such as COVID-19, which the world is currently facing [4] . So, in order to analyze the behavior and evolution of a disease among individuals in a population over time. This analysis is intended to assist in controlling the spread of the disease over a period of time, in order to prevent the spread of epidemics. Among several existing epidemiological models, the SIR model (Susceptible - Infected - Removed) proved to be the most suitable for this type of approach. In addition to epidemiological models, machine learning algorithms are recommended to predict time series learning from past data. The algorithms most used in research of this type may be artificial neural networks, such as the Multilayer Perceptron (MLP) network, Time-Delay Neural Networks (TDNN), radially based function networks (Radial Basis Function - RBF) and the Auto-Regressive Integrated Moving network (ARIMA) are frequently used to estimate temporal data [3]. In a recent study, neural networks were used to estimate the number of contagion and deaths in the state of Pará, aiming to analyze the impact on the capacity of care in the beds of the Intensive Care Unit, serving as a support for decision making by health surveillance bodies [3].

2.1 SIR model

The SIR model describes the spread of a disease in a population [5], dividing it into three classes: (S) Susceptible: the class of healthy individuals, but who can contract the disease; (I) Infected: the class of individuals who are sick; (R) Recovered: the class of individuals who are recovered from the disease. Due to the evolution of the disease, the size of each of these classes changes with time and the total size of the population N is the sum of these classes, according Eq. 1

$$N(t) = S(t) + l(t) + R(t) \quad (1)$$

In this model, there is also the parameter β, which represents the average number of contacts sufficient for the transmission of a person at time t .. In addition, there is the γ, which represents the rate at which infected individuals recover or die, called recovery rate. Based on these definitions, below are listed the rates of change through the differential equations, in Eq. 2, 3 and 4.

$$\frac{ds}{dt} = \frac{-\beta IS}{N} \quad (2)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \quad (3)$$

$$\frac{dR}{dt} = \gamma I \quad (4)$$

2.2 Neural network estimation

Artificial neural networks (ANN) are defined as complex structures interconnected by simple processing elements (neurons), which have the ability to perform operations, such as calculations in parallel, for data processing and knowledge representation [6]. In its learning process, ANN can be supervised or unsupervised. The properties and capabilities that make RNAs potentially useful involve their non-linearity, where an artificial neuron can use linear or non-linear functions; the Input-Output mapping based on input and output examples, the ANN is able to adapt to minimize the mapping error. Among the known structures of these models, MLP is a neural network with multiple connected layers, it has an input layer (without computational function), one or more hidden layers and an output layer [6].

3 Materials and methods

The proposed methodology has an experimental quantitative content. The samples were obtained through Kaggle, collected from the Ministry of Health of the numbers of cases in the State of Maranhão in the period from 27/03 to 27/05 of this year, with observations until the Lockdown decree in the State. The figures released by the Minister of Health were used for the parameters of the SIR epidemiological model according to each scenario: without measures adopted in the state, with social distance and with social and quarantine distance. The MLP was configured to receive the data released per day from the first notified day (03/27) until 05/17/2020 as the network entry. To obtain these results, the algorithm scripts were generated in the Python language.

- **Entry for the SIR model:** Population number (N), transmission rate, duration rate and number of days t.
- **Output to the SIR model:** Graph generated from susceptible, infected and recovered individuals over time

In Fig. 1, there is an illustration for the SIR model.

Figure 1. Methodology for the SIR epidemiological model.

The architecture of the proposed MLP network had two hidden layers and 200 neurons in each layer, with the 'relu' activation function and the Limited-memory Broyden Fletcher Goldfarb Shanno (LBFGS) optimizer due to the small amount of data. The training of the network was established with the days and the number of cases, estimating one day ahead of the data used for training. The algorithm used is available in [8].

- **Input to the neural network**: n days.
- **Output:** Graph of the projection for n + 1 day from the data.

Figure 2 depicts the MLP network architecture used in this study.

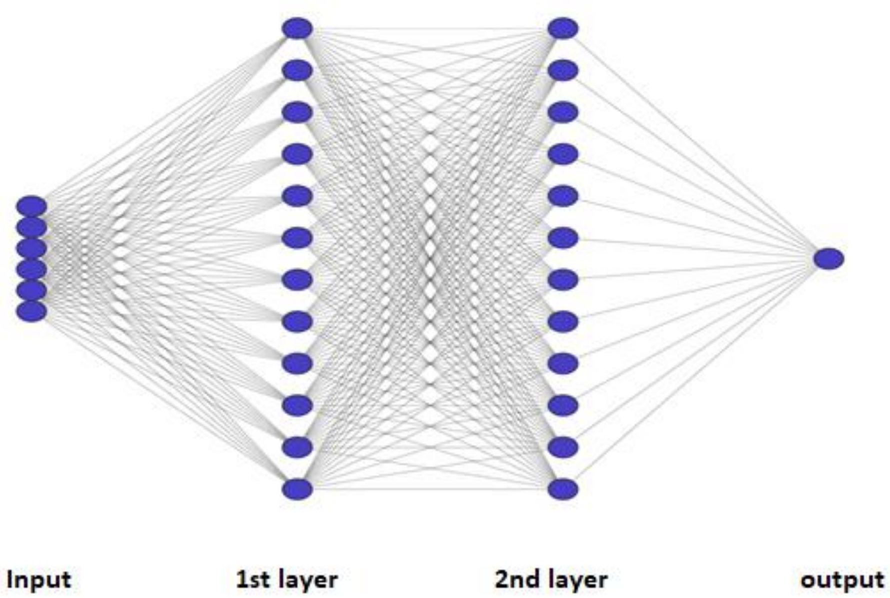

Figure 2. MLP network architecture.

4 Results and discussion

Analyzing the results for 04/30/20, the MLP neural network proved to be a better estimator for the forecasting of cases, in relation to the SIR model. While the MLP network obtained 2910 cases, the SIR epidemiological model reached 20000 cases in the data estimation. Table 1 shows the comparison between the two models used in this study.

Table 1. Comparative between the number of real and estimated cases.

Estimator	Date	Number of cases obtained	Real cases
SIR model	30/04/2020	2910	3190
Neural network	30/04/2020	3190	3190

Table 2 shows all period estimates obtained by the neural network, showing the difference between the predicted value and the real value of the number of cases in the state of Maranhão.

Table 2. Comparative between the number of real and estimated cases with 24 days

Date	Number of cases obtained	Real cases	Delta
25/04/2020	2105	2069	36
26/04/2020	2223	2244	21
27/04/2020	2410	2404	6
28/04/2020	2528	2568	40
29/04/2020	2804	2724	80
27/04/2020	3190	2924	266
01/05/2020	3506	3489	17
02/05/2020	3805	3666	139
03/05/2020	4040	4175	135
04/05/2020	4227	4382	155
05/05/2020	4530	4617	87
06/05/2020	5028	4864	164
07/05/2020	5389	5153	266
08/05/2020	5909	5647	262
09/05/2020	6765	6131	634
10/05/2020	7599	6842	757
11/05/2020	8144	8015	129
12/05/2020	8526	8951	425
13/05/2020	9112	9341	32
14/05/2020	9801	9769	9769
15/05/2020	10739	10399	340
16/05/2020	11592	10747	845
17/05/2020	12492	11993	499
Mean	5507.95	5787.56	656.69

4.1 SIR model

For this scenario, a 100-day forecast is taken from the date of the first case, March 19, in the state of Maranhão. The population parameter was defined as being 6.851 million, with the initial number of infected people equal to 1, and 0 recovered. The infection rates and infection durability per individual, according to the WHO, for this scenario, are 2.2 and 5.2 cases, respectively. The following graphs show a 100-day forecast if isolation measures had not been taken in Maranhão, resulting in a maximum number of cases of 4.8 million, hich would occur on March 30, 2020, 10 days after the first confirmation. As for 04/30/2020, 40 days after notification of the first case, an estimated 19,950 cases are estimated, according Fig. 3.

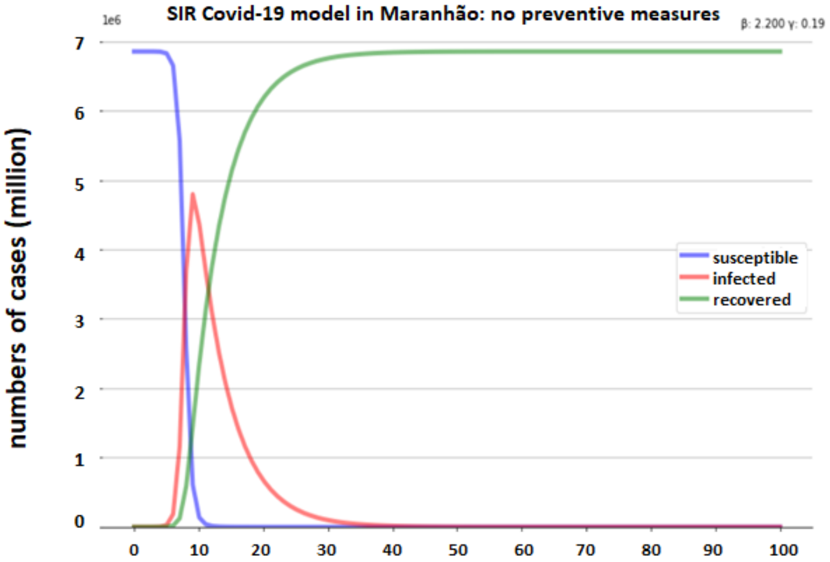

Figure 3. Estimates for the SIR epidemiological model.

4.2 Neural network

It can be seen that in Figure 4, the neural network was able to satisfactorily predict, in relation to the SIR model, the number of cases for 04/30/2020. As a result, the method obtained 2,900 cases, while the reported number of actual cases on that day was 3,190. For the error, the mean absolute error was used, obtaining 290 cases, according Fig.4.

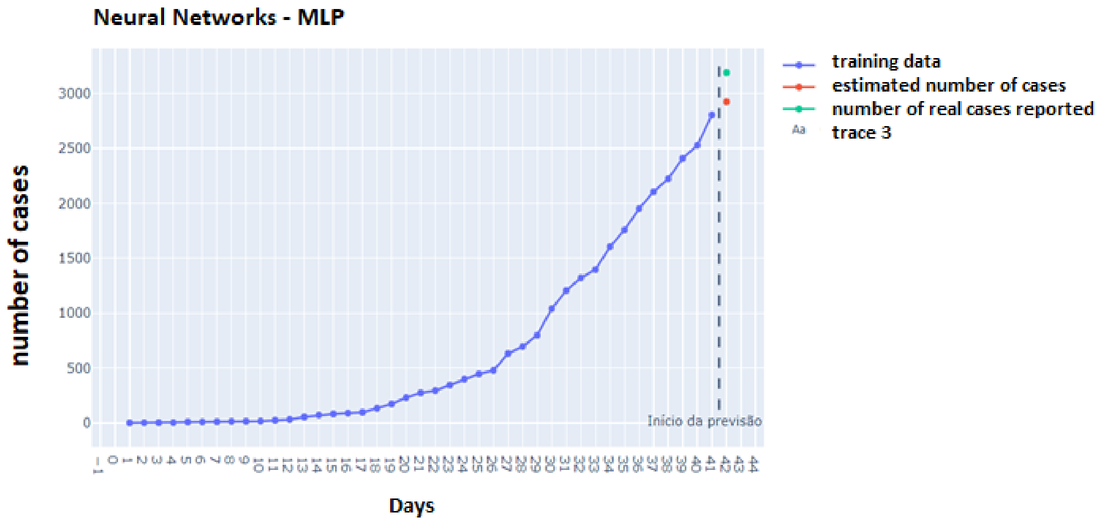

Figure 4. Estimates for one day with the neural network.

We analyzed 24 days, a period comprised of 10 days before the Lockdown decreed by the Maranhão State Government, and 10 days after the Lockdown, to have a closer estimate of the real, 12 thousand cases, as shown in Fig. 5.

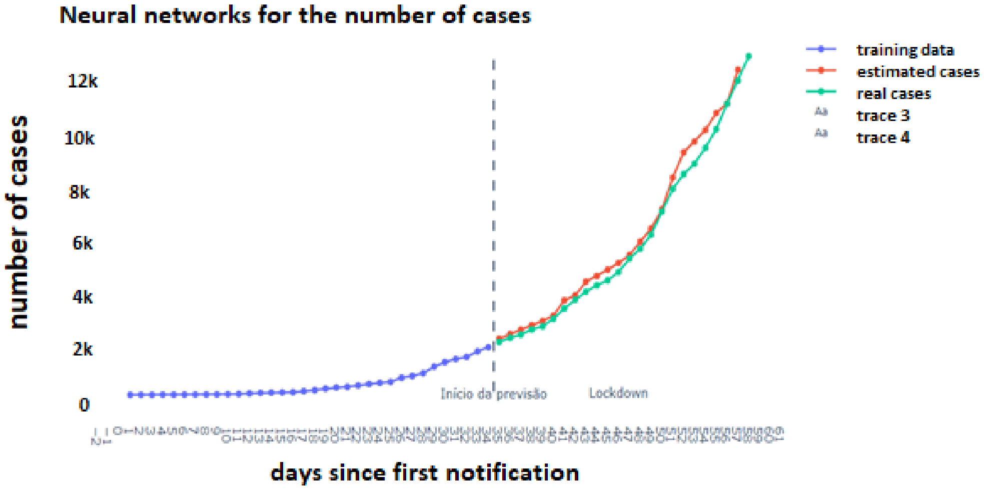

Figure 5. Estimates for many days with the neural network.

Artificial intelligence techniques, such as Artificial Neural Networks, can be promising to help combat the COVID-19 pandemic [5]. Other studies have already used RNA to predict new cases and the number of successes. For example, a study in the state of Pará used a short-term forecasting model based on Principal Component Analysis and RNA, capable of estimating the number of cases and deaths caused by SARS-CoV-2, showing a general error in the proposed model 6.30157% for confirmed cases and 1.06001% for deaths [6].

5 Conclusions

The algorithms used are sufficient to predict the cases of COVID-19 in the state of Maranhão. The MLP neural network showed a better estimate in relation to the SIR epidemiological model. Thus, the epidemiological model used in this research was able to segment the population of different individuals, as well as to simulate the curve of each group in the SIR model. The neural network used was able to satisfactorily estimate, in relation to the SIR model, the number of cases released on the last day of the data in this study. For future work, different parameters and techniques could be used to improve the result, as well as apply methods that allow estimating more days ahead of the released data.

Acknowledgements. We are greatufull to the CNPQ, and the Biológicas Information Processing Laboratory of UFMA.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Rocha JE, et al, Redes neurais artificiais na previsão de contágio e óbitos por covid-19 um estudo no estado do Pará,International Journal of Development Research (2020). Disponível em <http://repositorio.ufra.edu.br/jspui/bitstream/123456789/985/1/REDES%20NEURAIS%20ARTIFICIAIS%20NA%20PREVIS%C3%83O%20DE%20CONT%C3%81GIO%20E%20%C3%93BITOS%20POR%20COVID-19.....pdf>

[2] HAYKIN, S. Redes neurais: princípios e prática. 2.ed. Porto Alegre, Bookman, 2001.

[3] CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL. PREVISÕES A PARTIR DO MODELO EPIDEMIOLÓGICO SIR PARA OS CASOS DE INFECÇÃO PELO COVID-19 UMA APLICAÇÃO PARA OS ESTADOS DE SP, MG, RJ, DF E CE. Minas Gerais. 2020. Disponível em http://www.enf.ufmg.br/images/28.03-Previsoes-de-infecccao-pelo-COVID-19.pdf.pdf. Acesso em 10 de Junho.

[4] DIAS, G.; ARAÚJO, L. Modelagem Matemática para Epidemia de COVID-19 e Estimativa de Casos no Brasil no Curto Prazo. 2020. Disponível em http://covid19cientifico.ime.eb.br/artigo1.pdf. Acesso em 10 de junho

[5] Bullock, J., et.al. Mapping the Landscape of Artificial Intelligence Applications against COVID-19. 2020. Eprint ArXiv:2003.11336.

[6] Rocha, J. E. et al., Redes neurais artificiais na previsão de contágio e óbitos por covid-19: um estudo noestado do pará, Brasil, 2020. International Journal of Development Research, 10, (04), 35416-35421.