Application of an Artificial Neural Network to Classify countries with COVID-19 adjusting to SIR model

J. Juárez[1*], A. Sánchez[1], R. Díaz[1], L. Altamirano[1]
[1] Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México.
*jjlucero@inaoep.mx

Abstract

A backpropagation artificial neural network was employed in MathWorks MATLAB software to analyze the data of 96 countries infected with COVID-19. They were evaluated with the fitVirusCV19v3 (COVID-19 SIR model) function from MathWorks to classify them in two classes. The first-class has an R-square precision greater than 0.995 when evaluating the SIR model and the other with precision less than 0.995 to zero.

The data of the first class has only one maximum and can predict the final of the epidemic with the SIR model. The data of the second class has more than one peak, and it is not possible to find one maximum to evaluate the final of the disease in that country.

To training the artificial neural network, 70% of the data were used, 15% of data were taken for validation, and 15% to testing. In the testing, 5 to 7 countries were well classified and 2 of 7 were classified incorrectly. Then, the model has 75% of precision obtained with the confusion matrix, and the ROC curve shows sufficient success in the classification.

This first approximation to classify an epidemiologic phenomenon through an ANN with the SIR, the model could help to health professionals to know a possible panoramic about the infection course related to sickness, considering the mathematical conditions of the SIR model.

Keywords: COVID-19, Neural network classifier, SIR.

1. Introducción

A worldwide pandemic was started by a new coronavirus SARS-CoV-2 (Chih-Cheng, Tzu-Ping, Wen-Chien, Hung-Jen, & Po-Ren, 2020), (Lin, y otros, 2019). This new virus belongs to a new family of coronavirus that is considered a virus to jump between species form baths to humans (Phan, 2020), (Lupia, Scabini, Mornese-Pinna, Di-Perri, & Corcione, 2020).

The neural networks have been used in biomedicine to many applications (Krenek, y otros, 2014), (Boutaba, y otros, 2018), (Saravanan & Sasithra, 2014) as protein recognition (Wang, Sun, Li, Zhang, & Xu, 2017), medical image analysis (Litjens, y otros, 2017), or a cervical cancer detection (Elayaraja & Suganthi, 2018). The artificial neural network most used is back-propagation. It has a supervised learning model and needs to be trained, validated, and tested to develop a model to predict and classify data (Lau, Sun, & Yang, 2019). An activation function is used to train the model and it can be validated to test its performance (Kose & Arslan, 2017) . A method to classify data was used as an implementation of machine learning conceived as a neural network to identify inputs into a set of target categories, (Rosebrock, 2017), (Lippmann, 1987), (Jahnavi, 2017).

Lin et al. (Lin, y otros, 2019) in 2019 propose a conceptual model to maintain the control of an infectious disease like COVID-19, this model prevents outbreaks and reduce the pandemic time exposition if the people have an extension of holidays, make a travel restriction, and put the population in quarantine. This method can model over susceptible, infected, and recovery people of the disease,

dx.doi.org/10.24254/CNIB.20.59

with this information the SIR mathematic model can predict the duration time and the end of the quarantine taking in account the end of the disease in the population (Cadoni & Giusepe, 2020), (Ball, Briton, Leung, & Sirl, 2019), (Agrawal, Tenguria, & Modi, 2017). The graph of the SIR model is similar to a Gaussian function, with a maximum of infected people with the virus and is decreasing until zero during a time interval.

However, not all country's data distribution follows the Gaussian function and many of them at different times have two or more maximum, these have a precision near to zero in the SIR model. So, this paper aims to apply an artificial neural network for classifying all countries in the Tom Hopkins database considering the daily infected people of COVID-19 in countries that follow a distribution with one peak and can be adjusted to the SIR mathematic model or not.

2. Methodology

The first section describes SIR model characteristics and the second the neural network application to classify countries.

SIR model evaluation

The SIR mathematical model was calculated using the fitVirusCV19v3 software to assess if from the distribution of the infected people the parameters susceptible (S), infected (I), and recovery (R) people can be calculated to obtain the statistic and the R-square precision that the SIR model needs. This model can be solved by ordinary differential equations solutions with the initial conditions $S(0) = S_0 > 0, I(0) = I_0 > 0, R(0) = R_0 > 0$. If with data of one country it is possible to do this, then the duration time of the disease can be modeled with the SIR model.

$$\frac{dS}{dt} = \frac{-\beta}{N} IS \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta}{N} IS - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

S(t) is the susceptible, I(t) the infected, and R(t) the recovered people at the time t. β is the constant that represents the contact rate and γ is the inverse of the average infectious period. N is the total population depending on the other variables

$$N = S(t) + I(t) + R(t) \tag{4}$$

Fig. 1 shows an example of the two countries evaluated with the fitVirusCV19v3 software, in (A) the United Kingdom graph of infected people and (B) the USA one. (A) can be modeled by SIR mathematic model with one peak and the data were decreasing until zero (prediction of the end of the pandemic in that country), but (B) cannot be modeled because the distribution of the data makes not possible to find the maximum in this country so the parameters cannot be calculated and the precision is almost zero.

A) B)

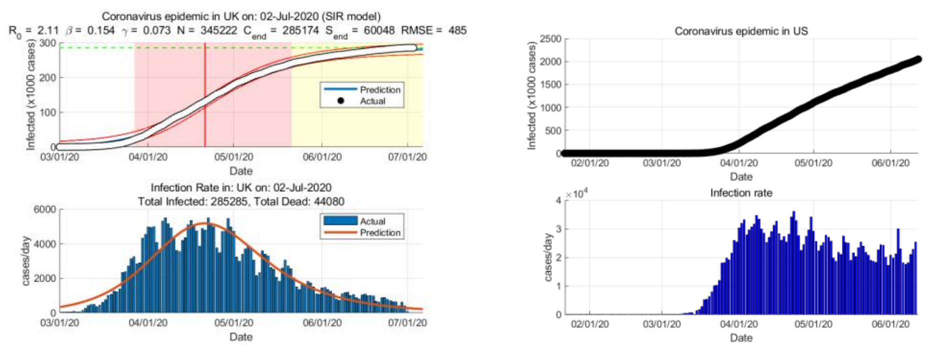

Fig. 1. Distribution of infected people by COVID-19 in A) United Kingdom, B) the United States of America.

The fitVirusCV19v3 (COVID-19 SIR model) function was calculated with the GitHub software of the MathWorks community. The goal was to know which country's data are adjusted with the SIR model. This program was developed by J. McGee solving ordinary differential equations (1-4) follow the methodology proposed by R. Ranjan (McGee, 2020), (Ranjan, 2020). The function also calculates the prediction of the duration of the SARS-CoV-2 pandemic with the Tom Hopkins database of infected people in 96 countries around the world.

Artificial Neural Network Classification

The used back-propagation artificial neural network (ANN) has one hidden layer with 300 neurons as shown in Fig. 2. It was used with an activation sigmoidal function and was performed with MathWorks MATLAB software. We have chosen this architecture with the porpoise of implementing a classifier, taking into account the number of available input data and as an expected result in the output (Bala & Kumar, 2017). The input data were 48 countries with an R-square precision greater than 0.995 and 48 countries whose R-square precision were less than 0.995).

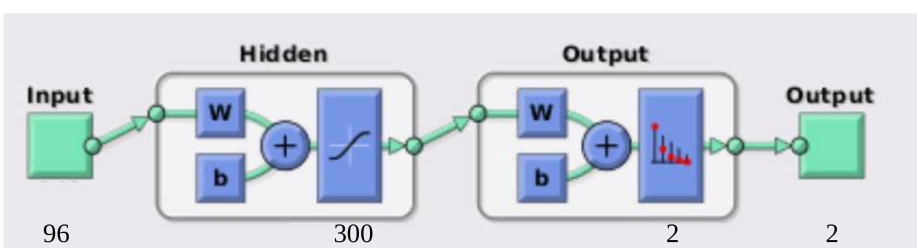

Fig. 2. Artificial Neural Network created in MATLAB software.

70% of mixed data were used for training the model, 15% used for validation, and 15% used for testing. The ANN performance was measured with a confusion matrix to calculate classes of data, the prediction accuracy, error state, sensitivity, specificity, and precision. The Receiver Operating Characteristic (ROC) curve was employed to detect the trade-off with true positive and false positive rates to calculate the area under the ROC curve to evaluate the performance of the classifications.
The general diagram is showed in Fig. 3.

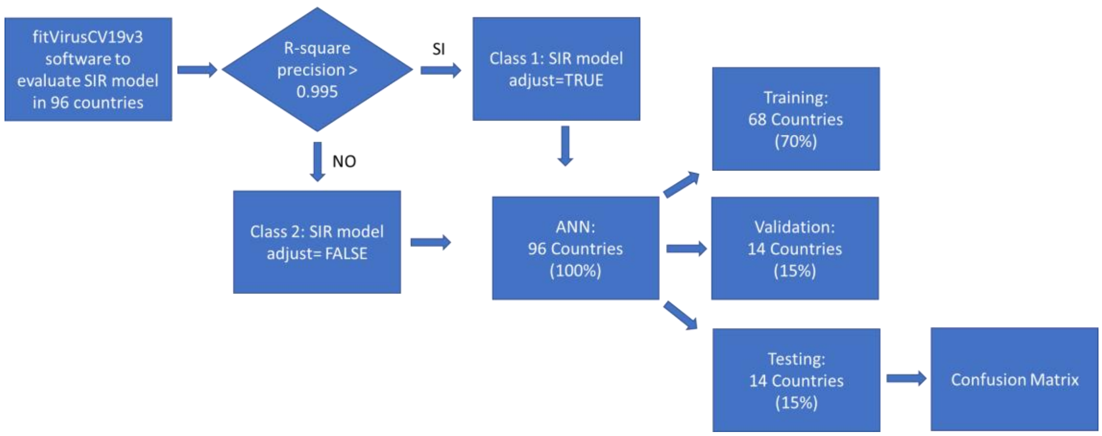

Fig. 3. Flux diagram of the methodology to evaluate the ANN prediction.

3. Results and discussion

The SIR model was evaluated in all data of countries in the Tom Hopkins database considering the number of infected people with COVID-19 disease. The SIR mathematical model was calculated using the fitVirusCV19v3 function to track if the distribution of the infected people (as shown in Fig. 1) is near approximately normal. These 96 countries were divided into two lists, the first 48 belong to a set that satisfies 0.995 or greater precision to the SIR model and the other 48 not, as shown in Fig. 4.

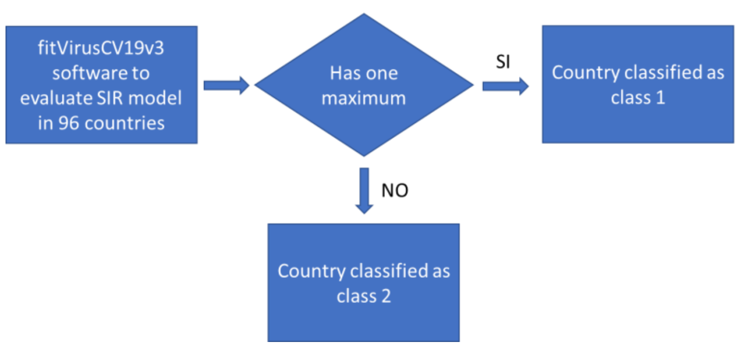

Fig.4. Diagram of countries distribution.

These data were analyzed by an ANN with a hidden layer, 300 neurons, and the activation sigmoidal function with MathWorks MATLAB software (Fig. 2). The error obtained is shown in Fig. 5.

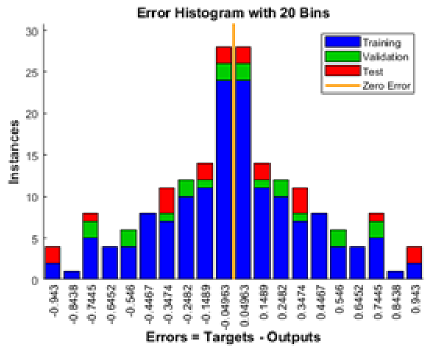

Fig. 5. Error histogram obtained with the proposed ANN.

The most of error distributions based in ANN predictions were near to zeroth point and the error was decreasing far away to center probing that artificial neural networks were performed successfully.

Table 1 shows the confusion matrix of ANN like a binary classifier studying two classes, 5 of 7 countries that had good precision into the SIR model, and 2 of 7 that do not were classified correctly.

	TRUE	FALSE
TRUE	5	2
FALSE	2	5

Table 1: LTB division results

In the confusion matrix, the true positive rate obtained was 67.92% and the true negative was 72.09% with these values were possible to calculate the error state, sensitivity, specificity, and the precision as shown in Table 2

Metrics	ANN
Accuracy	69.79
Error	30.21
Sentivity	69.92
Specificity	72.10
Precision	75.00

Table 2: Determination of rates in percentage

The sensitivity shows that 67.92% of data were correctly predicted, the specificity reaches 72.10%, so less than 27.9% were predicted incorrectly. This model has a 75% of precision in the accuracy to predict if data test of the infected people in a country has the precision required for the SIR model to find only one peak in the data. Finally, the ROC curve was calculated, and show in Fig. 6 the performance of the artificial neural network studied was successful enough to predict the adjustment by the country to the SIR model.

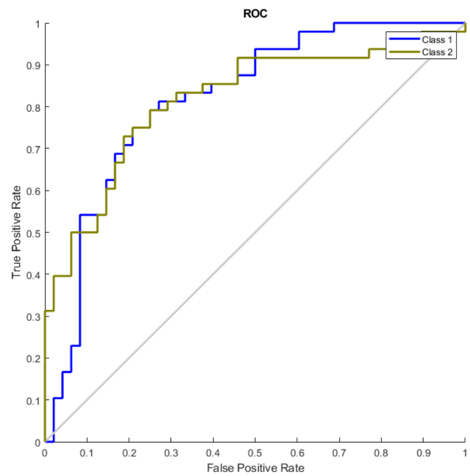

Fig. 6. The ROC curve of the proposed ANN functioning as a classifier.

4. Conclusions

In this paper, an artificial neural network was used to evaluate if the data of infected people by COVID-19 disease of 96 different countries can be used to identified COVID propagation near to SIR model. The back-propagation artificial neural network used has a 75% of precision to validate the performance of the model calculated with the confusion matrix. The ROC curve shows sufficient success to predict if the countries were well classified.

Results show that not all countries can be modeled because the daily disease distribution has two or more peaks and one maximum cannot be found to evaluate the SIR model to the end of the pandemic season. So, in those countries, where the virus is continuously spread with many daily peaks, it is difficult to find the maximum of contagions per day. In the cases of classified countries that adjust to the SIR model, it is possible to use the model to predict the duration time of the infection. For the other countries, it could not be possible to do a prediction because they have many peaks of infected people and other models like SEIR or SEIRD must be evaluated with their data.

This first approximation to classify an epidemiologic phenomenon through an ANN with the SIR model could be of help to health professionals to assess a possible general view about the infection course, considering the mathematical conditions of the SIR model. As future work, we plan to analyze cases with other mathematical models using ANN´s.

Referencias

Agrawal, A., Tenguria, A., & Modi, G. (2017). MATLAB Programming for Simulation of an SIR Deterministic Epidemic Model. *IJMTT, 50*(1), 71-73. doi:10.14445/22315373/IJMTT-V50P509

Bala, R., & Kumar, D. (2017). Classification Using ANN: A Review. *IJCI, 13*(7), 1811-1820.

Ball, F., Briton, K., Leung, D., & Sirl, D. (2019). A stochastic SIR network epidemic model with preventive dropping of edges. *J Math Biol, 78*, 1875-1951.

Boutaba, R., Salahuddin, M., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. (2018). A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *J. Internet Serv. Appl.*, 1-99.

Cadoni, M., & Giusepe, G. (2020). Size and timescale of epidemics in the SIR framework. *Physica D. Nonlinear Phenomena, 411*, 1-14. Obtenido de https://doi.org/10.1016/j.physd.2020.132626

Chih-Cheng, L., Tzu-Ping, S., Wen-Chien, K., Hung-Jen, T., & Po-Ren, H. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob, 55*(3), 1-9. Obtenido de https://doi.org/10.1016/j.ijantimicag.2020.105924. 2020

Elayaraja, P., & Suganthi, M. (2018). Automatic approach for cervical cancer detection and segmentation using neural network classifier. *APJCP, 19*, 3571-3580.

Jahnavi, M. (2017). *Introduction to Neural Networks, Advantages and Applications.* Obtenido de Towards Data Science: https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207

Kose, U., & Arslan, A. (2017). Optimization of self-learning in computer science engineering course: an intelligent software system supported by artificial neural network and vortex optimization algorithm. *Comput Appl Eng Educ, 25*(1), 142-156. doi:10.1002/cae.21787

Krenek, J., Kuca, K., Bartuskova, A., Krejcar, O., Maresova, P., & Sobeslav, V. (2014). Artificial neural networks in biomedicine applications. En ELSEVIER (Ed.), *In Conference: CENet 2014., 16*, págs. 1-8. Shangai. Obtenido de https://www.researchgate.net/publication/266146057_Artificial_Neural_Networks_in_Biomedicine_Applications?enrichId=rgreq-21031c306f0084281766807e4376ee97-XXX&enrichSource=Y292ZXJQYWdlOzI2NjE0NjA1NztBUzoyODExMzI3ODEzMjYzMzlAMTQ0NDAzODY4NjQyOA%3D%3D&el=1_x_

Lau, E., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences, 1*, 982. Obtenido de https://doi.org/10.1007/s42452-019-0884-7

Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S., Musa, S., & et al. (2019). A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *IJID, 93*, 211-216. Obtenido de https://www.sciencedirect.com/science/article/pii/S120197122030117X

Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4-22. Obtenido de https://www2.cs.sfu.ca/CourseCentral/414/li/material/refs/Lippmann-ASSP-87.pdf

Litjens, G., Kooi, T., Bejnordi, B., Setio, A., Ciompi, F., Ghafoorian, M., & CI., S. (2017). A survey on deep learning in medical image analysis. *Med Image Anal, 42*, 60-88. Obtenido de https://doi.org/10.1016/j.media.2017.07.005

Lupia, T., Scabini, S., Mornese-Pinna, S., Di-Perri, G. G.-D.-R., & Corcione, S. (2020). 2019 novel coronavirus (2019-nCoV) outbreak: a new challenge. *J. Glob. Antimicrob, 21*, 22-27. doi:https://doi.org/10.1016/j.jgar.2020.02.021

McGee, J. (2020). *fitVirusCV19v3 (COVID-19 SIR Model).* Obtenido de MathWorks: https://www.mathworks.com/matlabcentral/fileexchange/74676-fitviruscv19v3-covid-19-sir-model/

Phan, T. (2020). Novel coronavirus: from discovery to clinical diagnostics. *Infect. Genet. Evol.*, *79*, 1-2.

Ranjan, R. (2020). *MathWorks*. Obtenido de Predictions for COVID-19 outbreak in India using Epidemiological models.: https://www.researchgate.net/publication/340314461_Predictions_for_COVID-19_outbreak_in_India_using_Epidemiological_models

Rosebrock, A. (2017). *Deep learning for computer vision with python.* PUBLISHED BY PYIMAGESEARCH.

Saravanan, K., & Sasithra, S. (2014). Review on classification based on artificial neural networks. *IJASA*, *4*(2), 11-18.

Wang, S., Sun, S., Li, Z., Zhang, R., & Xu, J. (2017). Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLos Comput Biol*, *5*(13), 1-34. doi:https://doi.org/10.1371/journal.pcbi.1005324