Computers, Materials & Continua CMC, vol.64, no.3, pp.1343-1357, 2020

Data Driven Modelling of Coronavirus Spread in Spain

G. N. Baltas[1, *], F. A. Prieto[1], M. Frantzi[2], C. R. Garcia-Alonso[1] and P. Rodriguez[1, 3]

Abstract: During the late months of last year, a novel coronavirus was detected in Hubei, China. The virus, since then, has spread all across the globe forcing Word Health Organization (WHO) to declare COVID-19 outbreak a pandemic. In Spain, the virus started infecting the country slowly until rapid growth of infected people occurred in Madrid, Barcelona and other major cities. The government in an attempt to stop the rapssid spread of the virus and ensure that health system will not reach its capacity, implement strict measures by putting the entire country in quarantine. The duration of these measures, depends on the evolution of the virus in Spain. In this study, a Deep Neural Network approach using Monte Carlo is proposed for generating a database to train networks for estimating the optimal parameters of a SIR epidemiology model. The number of total infected people as of April 7 in Spain is considered as input to the Deep Neural Network. The adaptability of the model was evaluated using the latest data upon completion of this paper, i.e., April 14. The date range for the peak of infected people (i.e., active cases) based on the new information is estimated to be within 74 to 109 days after the first recorded case of COVID-19 in Spain. In addition, a curve fitting measure based on the squared Euclidean distance indicates that according to the current data the peak might occur before the 86th day. Collectively, Deep Neural Networks have proven accurate and useful tools in handling big epidemiological data and for peak prediction estimates.

Keywords: Coronavirus, deep neural network, machine learning, Monte Carlo simulation, SIR model.

1 Introduction

Just a few days before the beginning of this year, a novel coronavirus (nCoV) with highest medical significance was detected in Wuhan, capital of the province Hubei, China. As later classified, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [Gorbalenya, Baker, Baric et al. (2020)], is a single stranded positive sense RNA virus [(+)ssRNA virus], causing respiratory disease which is called coronavirus disease 2019 or COVID-19. Because of the severity of the symptoms and the high degree of

[1] Loyola Institute of Science and Technology, Universidad Loyola Andalucía, Seville, 41704, Spain.

[2] Department of Biomarker Research, Mosaiques Diagnostics GmbH, Hannover, 30659, Germany.

[3] Renewable Electrical Energy Systems, Technical University of Catalonia, Terrassa, 08222, Spain.

[*] Corresponding Author: Gregory N. Baltas. Email: ngbaltas@uloyola.es.

Received: 28 April 2020; Accepted: 21 May 2020.

contagiousness, the World Health Organization (WHO) declared COVID-19 outbreak, first as a Public Health Emergency of International Concern (on 30/01/2020) and subsequently as a pandemic on March 11th, 2020 [WHO (2020a)].

Since then, COVID-19 has spread all across the globe infecting almost two million people[4] (1,434,167 as of April 7, 2020), with nearly 82,063 patients having succumbed to the disease and more than 302,218 showing a recovery. Considering also that the basic reproduction number ($R0$) of the virus has been estimated at 2.2, strict measures are necessary to prevent health systems reach their capacity, an occasion where difficult decisions will need to be made such as prioritization of patients to be treated. Briefly, R0=2.2 means that if no population immunity and no preventive measures are taken, from each single infection 2.2 new infections are expected [Li, Guan, Wu et al. (2020)]. However, due to the nature of this virus, the infectivity rate and as a result the reproduction rate can be controlled through social distancing as suggested by Kwok et al. [Kwok, Wong, Wei et al. (2020)].

Beside the high social impact that this new pandemic is causing (including high number of deaths, distress and panic), the measures that were put in force to halt the spreading of the virus also affect the global economy indicating a "*domino*" effect that can last even longer than the probable eradication of COVID-19. In particular, according to a study conducted by Verikios et al. [Verikios, Sullivan, Stojanovski et al. (2015)] economies depending on tourism are expected to be affected the most due to travel restrictions, while in addition based on a study by Haimar et al. [Haimar and Santos (2014)] regarding the H1N1 pandemic, government agencies can suffer great losses despite low inoperability.

Estimating the evolution of COVID-19 is imperative for enhancing the efficiency of health systems. In addition, during a pandemic allocating resources like masks and respirators is vital for reducing infectivity among clinicians and paramedics. A study conducted to assess the impact of the H1N1 pandemic amidst 2009 with respect to resource availability revealed that protective gear can become insufficient both in national and international level [Murray, Grant, Bryce et al. (2010)].

In the battle against COVID-19 pandemic, machine learning (ML) and artificial intelligence (AI) analytical approaches are necessary tools in the researchers' "*faretra*", with multiple applications as recently reviewed by Bullock et al. [Bullock, Luccioni, Hoffmann et al. (2020)] including among others, improved diagnostic methods [Jiang, Coffee, Bari et al. (2020); He, Yu, Hong et al. (2017)], applications in drug development (discovery and drug re-purposing) [Réda, Kaufmann and Delahaye-Duriez (2019)] as well as epidemiological predictions [Koo, Liew, Mohamad et al. (2013)]. For instance, the applications of neural networks in modelling and analyzing the dynamics of diseases have been well studied demonstrating their effectiveness over conventional methods [Hamamd, Abdel-Wahab, De Claris et al. (1996)].

In this report, an AI approach, based on Deep Neural Networks (DNN), is designed to predict the evolution of COVID-19 active cases in Spain. The method consists of a data generation process based on Monte Carlo simulations of SIR epidemiology models, the development, optimization and validation of a DNN prediction model. At first, a brief

[4] Source: https://www.worldometers.info/coronavirus/.

summary of the virus evolution in Spain is provided, along with a description of the methodology that formed the basis of the peak prediction. Subsequently, the results are discussed along with conclusive statements, referring to the Spanish population.

2 Evolution of the COVID-19 in Spain

The first case of SARS-CoV-2 in Spain was detected on January 31st and was an isolated case in Las Islas Canarias, where the virus began to spread, with six cases reported by February, 24th. A second incidence was reported in Baleares on February 9th. By March 24th there has been no official registration of SARS-CoV-2 infection outside the Spanish islands, however on the following day several SARS-CoV-2 cases were detected in Spain's largest cities, i.e., Madrid, Catalonia and Valencia. Subsequently, from then on, the number of SARS-CoV-2 cases within Spain started to grow steadily.

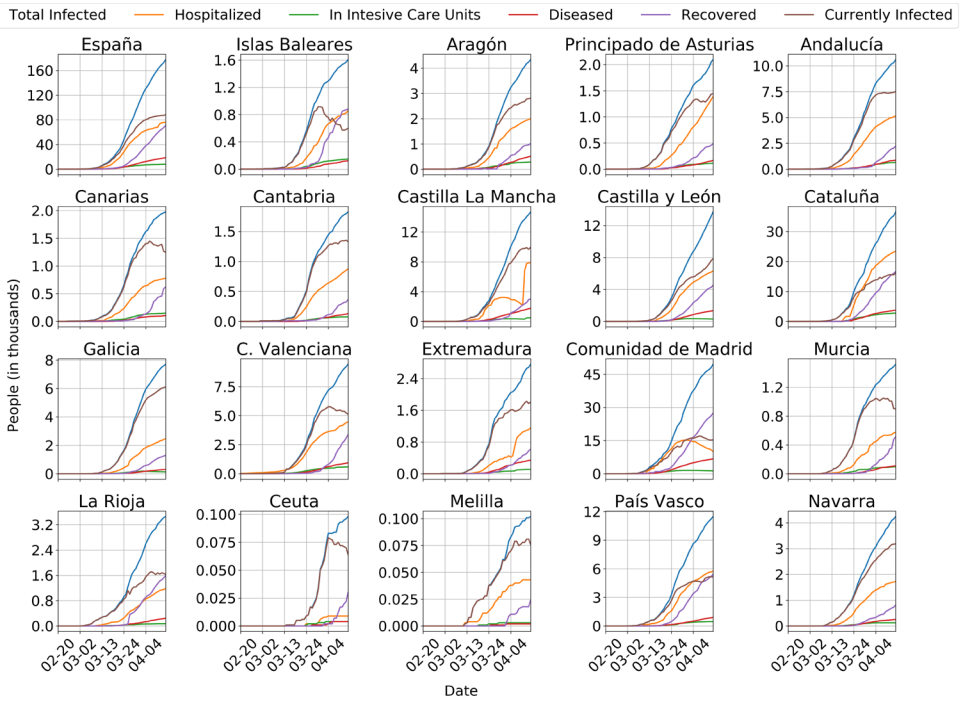

Figure 1: COVID-19 infections in Spain and Spanish provinces

The evolution of COVID-19 is depicted in Fig. 1, for each Spanish region separately. According to the underlying data (Fig. 1), Madrid and Catalonia are the two provinces that COVID-19 has spread more radically possibly due to the high number of inhabitants and population density. Fortunately, the number of recovered patients is constantly increasing in Madrid and Cataluña.

In this work, the COVID-19 peak estimation is considered for the national level, therefore only the data shown in the upper left graph are used. Data are available online by the Spanish Ministry of Health [Gobierno de España (2020)] for all provinces, which are update it regularly with information about the new cases, recoveries and deaths.

3 Methodology

3.1 The SIR model

A widely used mathematical tool for analyzing the spread of a virus is called SIR (Susceptible-Infected-Recovered) model. In its simplest form SIR model is based on a few strong assumptions [Weiss (2013); Weisstein (2020)]. Firstly, the number of population is constant for the duration of the analysis, meaning that we assume that there are no natural births and natural deaths occurring. Secondly, the incubation period (latent period) is zero, meaning that individuals become infectious directly at the time of their infection. Thirdly, SIR assumes a homogeneous mixing (mass-action principle), where populations of susceptible and infected individuals are homogeneously distributed and do not mix mostly in any smaller subgroups [Weisstein (2020)]. Yet, despite these assumptions, this model can provide good estimates on the evolution of an epidemic.

Briefly, the model consists of three coupled Ordinary Differential Equations (ODE), namely $S(t)$ representing the susceptible people, $I(t)$ for the infected and $R(t)$ for the recovered, all as function of time t. Using numerical integration techniques, these equations are solved for a pre-specified period of time, as in Eqs. (1)-(3) where β and γ are the infection and recovery rate parameters, respectively. The ratio defined by Eq. (4) is known as the epidemiological threshold ($R0$) and is a key indicator about the evolution of a disease. The initial fixed population is parameter η, which represents the fraction of the population in a country, such as Spain considered in SIR.

$$\dot{S} = -\beta SI \tag{1}$$

$$\dot{I} = \beta SI - \gamma I \tag{2}$$

$$\dot{R} = \gamma I \tag{3}$$

$$R0 = \frac{\beta}{\gamma} \tag{4}$$

3.2 Development of monte Carlo database

Based on the hypothesis that a SIR model can approximate COVID-19, Monte Carlo simulations [Theodoridis (2015)] are implemented for generating a set of probable SIR models by handling β and γ parameters as random variables. The workflow presented in Fig. 2 shows the process that generates the database for training and testing of the estimation models.

For iteration $m = [1, \cdots, M]$, where M is the total number of iterations, a random set of parameters is drawn from a uniform distribution, which is used to develop a unique SIR model. Ultimately, the two sets of data consist by the numerical integration of SIR for the infected agents (i.e., $\boldsymbol{X} \in \mathbb{R}^{M\times 365}$) and their corresponding parameters (i.e., $\boldsymbol{Y} \in \mathbb{R}^{M\times 2}$).

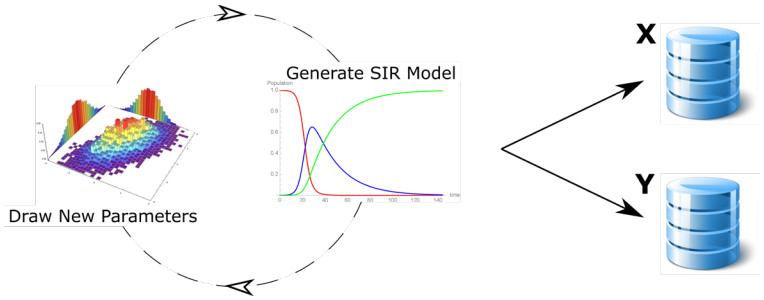

Figure 2: Monte Carlo database generation

To limit the simulations to generate only relevant data, a constraint has been added to maintain β and γ within the estimated range of WHO [WHO (2020b)]. As such, the epidemiological threshold is constrained within the open set of $1 < R0 < 5$. This is visualized in Fig. 3, where for a vast range of possible β and γ combinations the epidemiological threshold is plotted in the z-axis. The blue region represents the search space for the Monte Carlo Simulations according to the restriction described above. In this study, the number of simulated SIR models are $M = 4 \cdot 10^5$ based on random selection of β and γ to generate the different possible outbreak outcomes given a fixed population. In addition, several levels of η are considered for taking into account the impact of initial population in the SIR models.

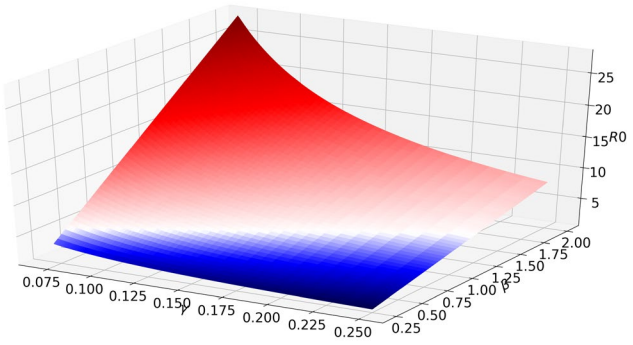

Figure 3: Monte Carlo search space

3.3 Artificial intelligence and deep neural networks

Inspired by the human brain cells, Deep Neural Networks (DNN) are the cornerstone of modern Artificial Intelligence (AI). A typical DNN consists of hundreds (or thousands) of neurons grouped in at least four layers: an input, an output and two hidden layers. At each neuron two operations occur: a summation of the weighted neuron inputs and a transformation of that sum through a mapping function. Several mapping functions exist such as the Sigmoid, Hyperbolic Tangent or the Rectified Linear Unit. Overall, DNN may differ in both size and structure however, all are typically known as universal function approximators due to their ability to solve any possible problem [Schaul,

Horgan, Gregor et al. (2020)]. DNN can be implemented for Supervised, Unsupervised and Reinforced Learning related tasks. Reasonably, the structure will differ depending on the type of learning.

In this paper, the estimation of the COVID-19 evolution in Spain is formulated as a Supervised Learning task where partially observed SIR curves are labeled by their parameters of their full curves. In other words, the matrix $\boldsymbol{X}_{spatial} \in \mathbb{R}^{M\times 48}$ and matrix $\boldsymbol{Y}_{spatial} \in \mathbb{R}^{M\times 2}$ are the input-output pairs that are used to train the DNN model.

To illustrate this, considering Spanish underlying datasets as of April 14^{th}, Fig. 4 depicts that the number of infected individuals followed a sharp trend upwards for a period of approximately two weeks between March 18 and April 5. Hence, to capture the dynamics of the virus spread, the information within the period 20/02−07/04/2020 is extracted and stored as the sample to be used by the DNN to estimate the optimal SIR model. This is called the observation window and is represented by the gray shaded area of the Fig. 4.

Similarly, for the same period the information about the infected from the generated SIR models is used and "*labeled*" by their β and γ parameters. Concretely, the DNN is trained to detect from this partial information the true parameters β and γ. Once trained the data within the grey area shown in Fig. 4 are fed into the DNN and the optimal SIR model can be created.

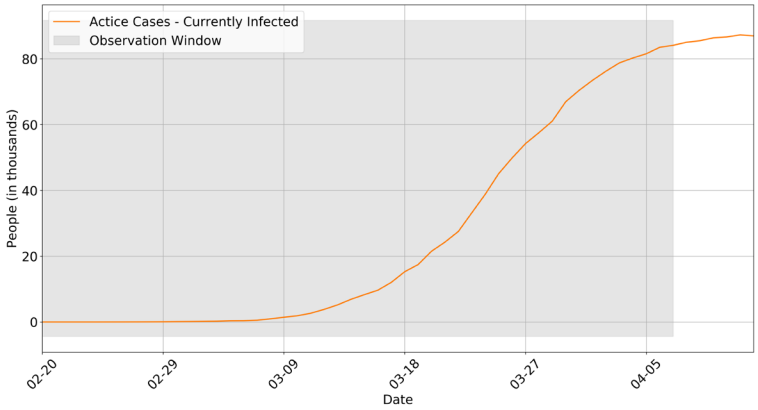

Figure 4: Active cases until 14th of April

4 Results

4.1 Training and optimization of the DNN models

The developed DNN models consist of one input layer with 48 neurons (observation window in Fig. 4) and an output layer with two neurons representing β and γ. Different designs were implemented and evaluated to conclude that the optimal DNN model should consist of two hidden layers each with 50 neurons and a single bias unit, as depicted in Fig. 5. All hidden neurons are using the Rectified Linear Unit (ReLu) activation function. The total number of trainable parameters can be found in Tab. 1. The loss function of the DNN is the Mean Absolute Error (MAE), as in Eq. (5) where y_i and $\hat{y}_i$are the true and predicted values, respectively, of the i-th sample. The loss function is

minimized using the Adam optimizer [Kingma and Ba (2014)] with a learning rate $l = 1e-4$, whereas Mean Absolute Percentage Error (MAPE), as in Eq. (6), is used for evaluating the performance of the optimal model, as it is more convenient to compare performance between different DNN designs.

To avoid over-fitting in DNN, common practice suggests creating subsets of input and output data for training, validating and testing the model [Geron (2019)]. Conversely, two subsets e.g., a training and an independent test set were created using an 80%/20% random split, respectively, from the main dataset, i.e., $\{X_{train}, Y_{train}\}$, $\{X_{test}, Y_{test}\}$.

The independent test set is kept aside for the whole process of training and tuning of the DNN, only to be used at the final step for testing the generalization to unseen cases. For training and validating the DNN a 5-fold cross validation scheme is adopted. Specifically, the training subset is divided in 5 equal sets. Each iteration four out of these five sets are used for training the network while the other for validating. Therefore, the final performance is an average of these models. Note that all sets are scaled between 0 and 1.

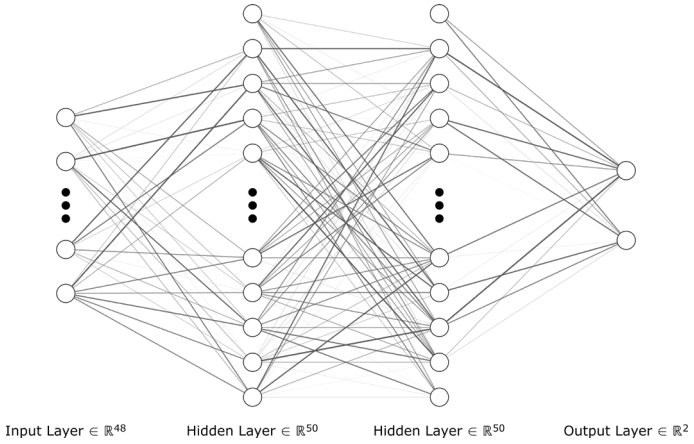

Figure 5: Structure of deep neural network

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{5}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \tag{6}$$

The aforementioned optimal DNN model is trained in Python using Tensorflow and Keras backend for 20 epochs using a model checkpoint callback to prevent over-fitting and store only the best weights according to the validation loss. The training and validation MAE and MAPE are plotted in Fig. 6. As it can be observed the callback will store the model of epoch 18 as its validation MAE is the smallest.

Table 1: Deep neural network parameter list

Layer	Parameters
Input	-
Hidden 1	2,450
Hidden 2	2,550
Output	102
Total	5,102

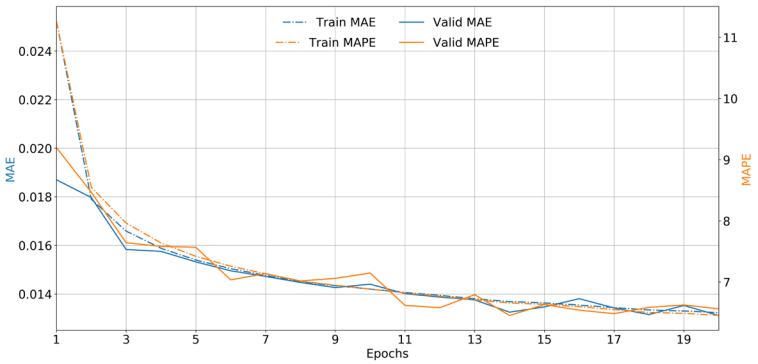

Figure 6: Performance of deep neural network based on 5-fold cross validation

4.2 Validation on unseen cases

As mentioned, to avoid over-fitting, which may hinder the accuracy of the model to generalize in actual cases, an independent test set was kept aside from the development/ optimization process of the DNN. Concretely, a random sample is chosen to be used for predicted and plotting the true and estimated SIR model, as an independent test set.

As shown in Fig. 7 the DNN accurately predicts the parameters of the SIR model according to the partial curve of infected people for a random sample from the independent set. Overall, the cross validated performance of the DNN in the independent set is listed in Tab. 2. From these values, the average MAE and MAPE is about 12.87e-3 and 6.31% respectively, with a corresponding standard deviation of 0.237e-3 and 0.06%.

Table 2: Deep neural network metrics on independent test set

K-Fold Model	MAE (e-3)	MAPE (%)
1	12.96	6.37
2	12.94	6.30
3	12.43	6.20
4	13.14	6.34
5	12.89	6.34

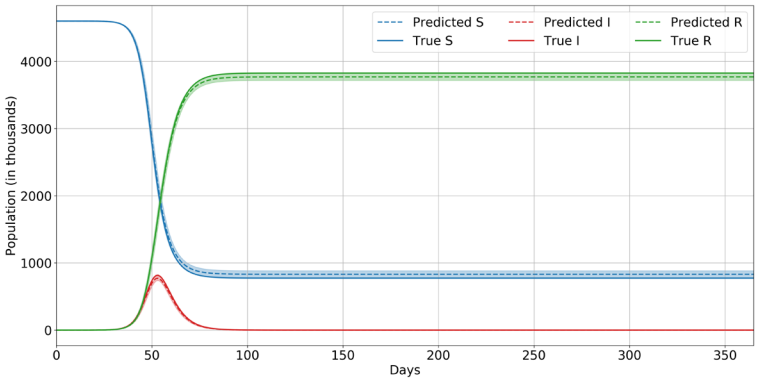

Figure 7: Deep neural network SIR model prediction for random sample from independent test set

4.3 Estimated SIR model for Spain

The purpose of the DNN model is to estimate the optimal parameters of a SIR model for the recorded total of infected individuals over a time span of 48 days. Datasets, over the period from the 20/02/2020 up to 07/04/2020 were considered, as the time point that the DNN was developed and because that is the period where the rapid growth of infected people is observed in Spain.

The results presented in Fig. 8, show the predicted evolution of the COVID-19 active cases in Spain considering data input, up until 07/04/2020. The cross validated SIR parameters predicted by the DNN generate a $\overline{R0} = 2.24$, which is in agreement with estimated by WHO [WHO (2020b)]. The predicted active cases evolution will reach a peak on April 10 and 91,554 infected people. From then on, the active cases are decreasing towards zero.

The SIR model was developed using as initial conditions the number of infected and recovered people on 7 April. As mentioned, the SIR model requires the specification of the population that can be infected by the virus. The predictions shown in Fig. 8 are based on an initial population $\eta = 0.73\%$ from the total population of Spain i.e., $\sim$335,800. This specific value has been found to minimize the squared Euclidean distance between the generated curve and the recorded active cases within the period of 08/04-14/04/2020, as depicted in Fig. 9. In this study, the squared Euclidean distance, defined in Eq. (7) where $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}^n$ are vectors, is used for calculating how well the predicted curve fits the true data.

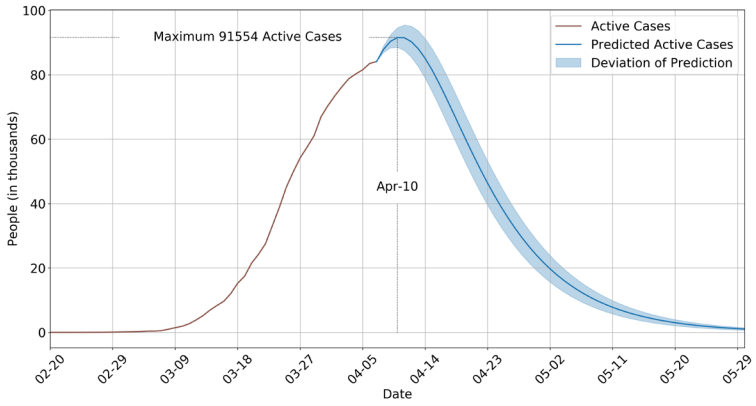

Figure 8: Predicted evolution of active cases in Spain using data until 7th of April

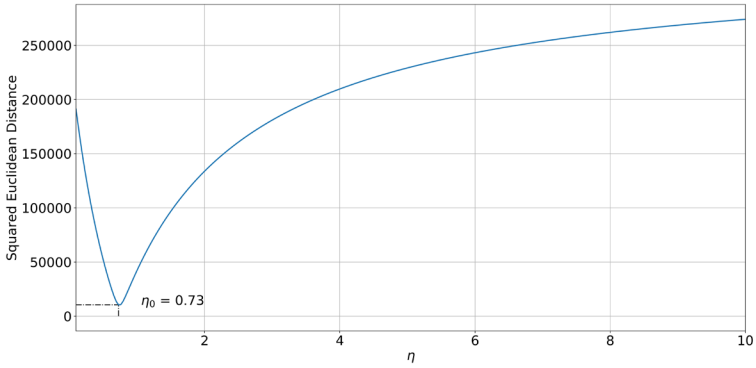

Figure 9: Squared Euclidean distance between predicted and true active cases for various $\boldsymbol{\eta}$

$$sqdist^2 = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{7}$$

Upon finishing this document, new data became available. Therefore, using the same parameters the SIR model was recreated, incorporating the latest information about infected and recovered cases. The generated SIR model is plotted in Fig. 10 revealing that the new peak would be on April 16. Moreover, in contrast to the prediction based on the data until April 7, the maximum active case is approximately 1,000 less. Most importantly, the deviation of predictions in Fig. 10 is also lower indicating a higher confidence result.

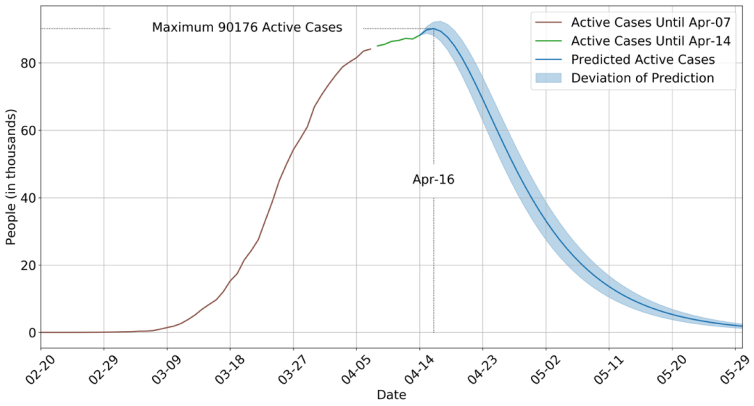

Figure 10: Predicted evolution of active cases in Spain using latest data

The peak value and peak occurrence are accounting for a defined population, which is considered a stable variable throughout the analysis of parameterization of the SIR model. Therefore, using the predicted parameters β and γ accepting them as true, SIR models are sensitive to differences in population size. Nevertheless, the maximum number of infected people appears to be linearly dependent of the initial population considered in the SIR model. That is, according to the Spanish population considered in the SIR model the maximum number of active cases can be between 88 thousand up to 8.7 million people. This means that the infected people may be as low as 0.1% or as high as 19% of the total population in Spain.

Adjusting for differences at the initial population, reveals a time window for the COVID-19 peak in Spain, within the range of 74 to 109 days after February 1st, i.e., 15/04/2020-20/05/2020 as illustrated in Fig. 11. However, considering the square Euclidean distance, shown in Fig. 9, being considerably higher for values above $\eta = 2\%$ we can reasonably assume that peak would be reached before the 86th day i.e., 27/05/2020 unless changes in the current condition of safety measures occur.

5 Discussion

Data-driven approaches in modelling and forecasting the dynamics of an infectious disease such as H1N1, SARS and Ebola have been well studied. To illustrate, Lega and Brown proposed the approximation of the parameters of a SIR model by fitting an inverted parabola to the new versus cumulative cases [Lega and Brown (2016)]. Similarly, an agent-based approach is adopted by Venkatramanan et al. [Venkatramanan, Lewis, Chen et al. (2017)] to model the Ebola outbreak by including qualitative data.

In addition, application of AI (and particularly DNN) has been proven very useful to the fight against COVID-19 pandemic [Wang, Kang, Ma et al. (2020); Ozturk, Muhammed, Yildirim et al. (2020)]. In this study, the application of a DNN has been investigated for the identification and optimization of SIR model parameters based on actual healthcare data, for Spanish population within the time period from 20/02-07/04/2020. Based on this, a prediction of the peak date within the time period of 15/04/2020-20/05/2020 is reported.

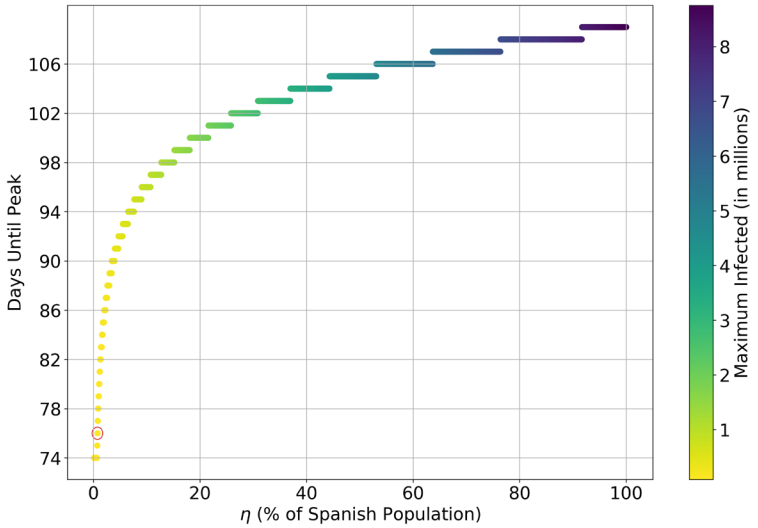

Figure 11: Days until peak[5] of infected cases as a function of $\boldsymbol{\eta}$ (red circle corresponds to the estimations shown in Figs. 8 and 10)

To provide most accurate predictions, the optimized parameters as established in a training set were further validated in an independent test set within the same cohort. Apparently, as estimates are strongly related to the given population at each specific time point, the optimized parameters are subjected to changes occurring over the incubation expansion period of the viral infection within each specific region.

Moreover, in this estimate, population immunity as a form of indirect protection from the infection, is not accounted as a variable considering that for population immunity to occur, a large percentage of a population has to become immune (basically disease and recover from the infection). Additionally, this prediction is based on the certain detection methods that are currently in clinical practice and consider that the detection policy (guidelines on testing of symptomatic population) will remain the same within the next weeks.

Furthermore, the DNN developed in this paper is built upon a database of SIR models generated by random selection of its parameters. Essentially, this means that the behavior of the virus should follow the response of a typical SIR model. In reality, however, this might not be the case. Nevertheless, by approximating the evolution of the virus it is possible to obtain insights on the overall behavior of the virus.

For instance, a regression model is developed by Magdon-Ismail for estimating the evolution of total infected people in USA [Magdon-Ismail (2020)]. The SIR model used in that study incorporates social distancing and lockdown through a lag variable. However, the model requires frequent re-tuning as the safety measures put in force change the evolution of the virus.

Similarly, Dandekar et al. develop neural networks for estimating the evolution of active cases in Wuhan, Italy, South Korea and USA [Dandekar and Barbastathis (2020)]. In this

[5] Since the first case recorded in Spain i.e., 1st of February.

study, the neural networks are used as an ODE integrator meaning that both parameters β and γ as well as the weights of the neural network are optimized at the same time. The models are based on the SIR and the SEIR epidemiology models. The initial population consider in each model is the full population of each region i.e., $\eta = 100$. Nevertheless, this approach overestimates the number of active cases, yet the initial population might also contribute to the over-estimations.

In contrast, the proposed DNN in this paper is able to generalize in the recorded data even though they do not follow a typical SIR response. However, the SIR model approximates the active cases using a bell shaped curve thus exact fitting of the true data is not possible. For this, the squared Euclidean Distance is employed to identify the similarity with the recorded cases. Finally, the DNN does not need retuning and it is not case specific meaning that it can predict the β and γ parameters given data for the active cases for a 48-day period is available.

6 Conclusions

DNN models provide several advantages in handling big epidemiological data and for accurate peak predictions. Although the prediction estimates within this study can inform and guide measures, yet further studies on COVID-19 evolution curves are required to obtain the SIR model parameters, thus generating a population-dependent model.

Nevertheless, the DNN is an advanced technique that has made possible to know the parameters of the SIR model that better adapts to the data of Spain (being this the studied case). Collectively, the simplicity of the proposed approach with the DNN allows identifying the SIR parameters for different COVID-19 evolution curves what it could help the scientific community to identify curves from different population sizes in contact with the virus.

Funding Statement: This work was supported by the European Commission under project FLEXITRANSTORE-H2020-LCE-2016-2017-SGS-774407 and by the Spanish Ministry of Science under project ENE2017-88889-C2-1-R. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the host institutions or funders.

Conflicts of Interest: MF is employed by Mosaiques Diagnostics GmbH.

References

Bullock, J.; Luccioni, A.; Hoffmann Pham, K.; Sin Nga Lam, C.; Luengo-Oroz, M. (2020): Mapping the landscape of artificial intelligence applications against COVID-19. arXiv: 2003.11336.

Dandekar, R.; Barbastathis, G. (2020): Quantifying the effect of quarantine control in COVID-19 infectious spread using machine learning. medRxiv: 2020.04.03.20052084.

Flaxman, S.; Mishra, S.; Gandy, A. (2020): Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. arXiv: 2004.11342.

Geron, A. (2019): Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, Inc.

Gobierno de España (2020): Situación de COVID-19 en España. *Ministerio de Sanidad.* https://covid19.isciii.es/.

Gorbalenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, R. J.; Drosten, C. et al. (2020): The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, vol. 5, no. 4, pp. 536-544.

Haimar, A. E.; Santos, J. (2014): Modelling uncertainties in workforce disruptions from influenza pandemics using dynamic input-output analysis. *Risk Analysis*, vol. 34, no. 3.

Hammad, T. A.; Abdel-Wahab, M. F.; DeClaris, N.; El-Sahly, A.; El-Kady, N. et al. (1996): Comparative evaluation of the use of artificial neural networks for modelling the epidemiology of Schistosomiasis mansoni. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 90, pp. 372-376.

He, Q.; Yu, B.; Hong, X.; Lv, B.; Liu, T. et al. (2017): An improved lung sound de-noising method by wavelet packet transform with PSO-based threshold selection. *Intelligent Automation & Soft Computing.*

Jiang, X.; Coffee, M.; Bari, A.; Wang, J.; Jiang, X. et al. (2020): Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua*, vol. 63, no. 1, pp. 537-551.

Kingma, D. P.; Ba, J. (2014): Adam: a method for stochastic optimization. arXiv: 1412.6980.

Koo, C. L.; Liew, M. J.; Mohamad, M. S.; Mohamed Salleh, A. H. (2013): A review for detecting gene-gene interactions using machine learning methods in genetic epidemiology. *BioMed Research International*, vol. 2013, pp. 13.

Kwok, K. O.; Wong, V. W. Y.; Wei, W. I.; Wong, S. Y. S.; Tang, J. W. (2020): Epidemiological characteristics of the first 53 laboratory-confirmed cases of COVID-19 epidemic in Hong Kong. *Euro Surveill*, vol. 25, no. 16.

Lega, J.; Brown, H. E. (2016): Data-driven outbreak forecasting with a simple nonlinear growth rate. *Epidemics,* vol. 17, pp. 19-26.

Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L. et al. (2020): Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199-1207.

Magdon-Ismail, M. (2020): Machine learning the phenomenology of COVID-19 from early infection dynamics. arXiv: 2003.07602.

Murray, M.; Grant, J.; Bryce, E.; Chilton, P.; Forrester, L. (2010): Facial protective equipment, personnel and pandemics: impact of the pandemic (H1N1) 2009 virus on personnel and use of facial protective equipment. *Infection Control and Hospital Epidemiology*, vol. 31, no. 10.

Ozturk, T.; Muhammed, T.; Yildirim, E. A.; Baloglu, U. B.; Yildirim, O. et al. (2020): Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine.*

Réda, C.; Kaufmann, E.; Delahaye-Duriez, A. (2019): Machine learning applications in drug development. *Computational and Structural Biotechnology Journal*, vol. 18, pp. 241-252.

Schaul, T.; Horgan, D.; Gregor, K.; Silver, D. (2015): Universal value function approximators. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, pp. 1312-1320.

Theodoridis, S. (2015): Monte Carlo methods. *Machine Learning: A Bayesian and Optimization Perspective*. Academic Press.

Venkatramanan, S.; Lewis, B.; Chen, J.; Higdon, D.; Vullikanti, A. et al. (2017): Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics*, vol. 22, pp. 43-49.

Verikios, G.; Sullivan, M.; Stojanovski, P.; Giesecke, J.; Woo, G. (2015): Assessing regional risks from pandemic influenza: a scenario analysis. *The World Economy*, pp. 31.

Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M. et al. (2020): A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). medRxiv: 2020.02.14.20023028.

Weiss, H. (2013): The SIR model and the foundations of public health. *MATerials MATemàtics*, vol. 2013, no. 3, pp. 17.

Weisstein, E. W. (2020): Kermack-McKendrick model. From MathWorld-a Wolfram web resource. https://mathworld.wolfram.com/Kermack-McKendrickModel.html.

World Health Organization (2020a): Director-General's opening remarks at the media briefing on COVID-19-11 March 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

World Health Organization (2020b): Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus. https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) 2019 (n-CoV) on 23 January 2020.