Article

A Hybrid Model for COVID-19 Monitoring and Prediction

Luis Fernando Castillo Ossa [1], Pablo Chamoso [2], Jeferson Arango-López [3], Francisco Pinto-Santos [2], Gustavo Adolfo Isaza [3], Cristina Santa-Cruz-González [4], Alejandro Ceballos-Marquez [5], Guillermo Hernández [4] and Juan M. Corchado [2,4,6,*]

1 Grupo de Investigación Inteligencia Artificial, Departamento de Sistemas e Informática, Universidad de Caldas, Manizales 170004, Colombia; luis.castillo@ucaldas.edu.co
2 BISITE Research Group, University of Salamanca, 37007 Salamanca, Spain; chamoso@usal.es (P.C.); franpintosantos@usal.es (F.P.-S.)
3 Grupo Investigación GITIR, Departamento de Sistemas e Informática, Universidad de Caldas, Manizales 170004, Colombia; jeferson.arango@ucaldas.edu.co (J.A.-L.); gustavo.isaza@ucaldas.edu.co (G.A.I.)
4 Air Institute, IoT Digital Innovation Hub, 37188 Salamanca, Spain; csantacg@gmail.com (C.S.-C.-G.); guillehg@air-institute.org (G.H.)
5 Grupo de investigación CLEV, Universidad de Caldas, Manizales 170004, Colombia; alejandro.ceballos@ucaldas.edu.co
6 Department of Electronics, Information and Communication, Faculty of Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan
* Correspondence: corchado@usal.es

Abstract: COVID-19 is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has a case-fatality rate of 2–3%, with higher rates among elderly patients and patients with comorbidities. Radiologically, COVID-19 is characterised by multifocal ground-glass opacities, even for patients with mild disease. Clinically, patients with COVID-19 present respiratory symptoms, which are very similar to other respiratory virus infections. Our knowledge regarding the SARS-CoV-2 virus is still very limited. These facts make it vitally important to establish mechanisms that allow to model and predict the evolution of the virus and to analyze the spread of cases under different circumstances. The objective of this article is to present a model developed for the evolution of COVID in the city of Manizales, capital of the Department of Caldas, Colombia, focusing on the methodology used to allow its application to other cases, as well as on the monitoring tools developed for this purpose. This methodology is based on a hybrid model which combines the population dynamics of the SIR model of differential equations with extrapolations based on recurrent neural networks. This combination provides self-explanatory results in terms of a coefficient that fluctuates with the restraint measures, which may be further refined by expert rules that capture the expected changes in such measures.

Keywords: COVID-19; recurrent neural network; LSTM; compartmental models; curve fitting; prediction

1. Introduction

A new virus called SARS-CoV-2 has caused a worldwide health crisis, which has been difficult to tackle as a result of the limited knowledge regarding the virus. The scientific community has only been able to learn about the virus as the pandemic progressed and more information was available. For example, it has been possible to determine morbidity and mortality rates in different population groups, as well as other environmental and social factors. However, there is still a lack of information regarding the SARS-CoV-2 virus. Moreover, it has not been possible to identify the factors that lead to serious illness in some Coronavirus patients.

The first cases of Coronavirus were reported in China, which became the epicenter of COVID-19 in January. Meanwhile, there was only a small amount of cases in other

countries, which had been imported by travelers. By the end of the month, there were 10,000 cases in China and 129 abroad. However, the virus spread quickly and in February there were several outbreaks in South Korea, Italy, Germany and Spain. There have been more than 80,000 deaths and currently, more than 1.4 million cases have been confirmed in most of Central America, South America, North America, Asia, Australia and Europe, and the numbers continue to grow everyday. The spread rate is different in each country; however, the suppression and mitigation policies do not seem to be the only factor behind those variations. Japan, Hong Kong and Singapore have experienced a steady increase in cases since January. In Italy, the virus spread much faster, leading to a state of emergency. Soon, the virus spread to other European countries such as Spain, France or Germany.

The statistics gathered regarding this pandemic are far from representing the real situation and the level of contagion is much greater than reported by official authorities. The reason for this is that there are no adequate and reliable mechanisms that would be capable of diagnosing COVID-19 rapidly. This means that we are measuring two variables simultaneously: the case rate in each country and the capacity of the government to efficiently detect cases.

There have been considerable advances in the field of Precision Medicine (PM), where new techniques give us more opportunities to learn about COVID-19. If healthcare professionals had more knowledge regarding the virus, they would be able to establish an effective treatment for their patients, adapted to their genetic profile and predicted evolution. Until a suitable treatment is found or the mass vaccination campaign is completed, it is crucial to take the adequate measures that will help slow down the spread of the virus. To this end, the short, medium and long term evolution of the pandemic must be predicted, taking into account how it has evolved in a particular area and the changes in trends associated with mobility restrictions, confinement, etc. Therefore, a prediction system has been developed with detailed explanations of the prediction results, which are of help when making decisions. Auto-Explainable: Artificial Intelligence offers the opportunity to comprehend the results of AI systems and the system incorporates a XAI module for this purpose. Unlike in black box models, an intermediate representation is obtained with a direct epidemiological interpretation, which also allows the model to be adjusted by introducing additional hypotheses on the nature of the contention measures.

This paper presents a model developed to forecast the evolution of SARS-CoV-2; it is an artificial intelligence based model that combines several algorithms. The model is capable of assisting healthcare authorities in making decisions, for example, in selecting the most adequate spread suppression mechanisms for a region or in the adequate dimensioning and prioritization of the available medical resources. This model has been developed in the framework of the project Sistema De Inteligencia Epidemiológica Para El Apoyo En La Toma De Decisiones En El Control Del COVID-19 En LatinoAmérica, conv. no. 1015 de 2020, code 1127101576535. Sponsored by the Ministerio de Ciencia, Tecnología e Innovación de Colombia. The model is in use by the local medical authorities.

During the health emergency generated by the novel coronavirus SARS-CoV-2 and COVID-19 pandemic, in Caldas department, Colombia, the need of computer tools developed to measure and predict the socioeconomic impacts that can affect Caldas department from the beginning of the pandemic to an estimated time in the future.

Manizales, the capital city of Caldas department, showed a special requirement, due to a much higher population density in comparison with other municipalities of the department and in consequence with higher probability of new cases and fatalities caused by COVID-19. Health authorities need information in real time that supports the decision making related with the capacity and occupation of the available intensive care units (ICUs). The last, due to the remissions of patients arriving from different municipalities of the entire department in addition to the local demand. Considering the economic reactivation as well as relevant information needed to prioritize the opening of certain sectors that greatly boost the economic growth of the city and the region.

A review of Artificial Intelligence models for predicting the evolution of time series is presented. The mixture of experts developed for Manizales is then described and 83 evaluated. Finally some conclusions are presented.

2. Predicting the Pandemic with Artificial Intelligence Models

Numerous AI models are being used to predict the evolution of the pandemic. This article does not attempt to present an exhaustive review of the existing approaches. Rather, it attempts to show some varied examples of proposals which include intelligent algorithms, and which are representative of the state of the art. A list of some of the approaches used for the prediction of the evolution of COVID-19 infections is presented. Some are proposals based on local data while others are more ambitious, international studies. There are proposals that use deep learning [1–7], others use the Random Forest model [8,9], networks based on domain models [10] and SIR models [11]. The hybrid model proposed for the prediction in the present research includes an SIR model. Auto encoders [12,13] and binary classifiers [14] are also used in the prediction of COVID-19 evolution. Support Vector Regression and stacking-ensemble have also been proposed with some success [15], in addition to XGBoost classifier [16].

Artificial Intelligence has experienced significant growth over the last decades, giving rise to numerous technologies and approaches within the AI paradigm. However, separately, all of these approaches have some weaknesses which can be mitigated by combining the approaches, techniques or concepts from the different subfields of AI. This practice is called hybrid AI [17] and includes neural computing, machine learning, fuzzy logic and evolutionary algorithms [18–20].

For example, connectionist approaches in AI focus on the use of neural networks, while symbolic approaches involve the representation of problems in a way that can be understood by humans. Unfortunately, the results generated by neural networks [21] are opaque while symbolic approaches tend to generalize solutions. Thus, thanks to hybrid AI, both approaches are able to offer their benefits and to overcome their limitations.

Moreover, thanks to the variety of the technologies used in hybrid AI, solutions can be found to countless problems in a number of disciplines, including bioinformatics, control engineering, software verification, visual intelligence and ontological learning.

To illustrate this with a specific example, Deep Neuro Fuzzy algorithms [22] are implemented in smart systems; they jointly employ techniques based on fuzzy logic and deep neural networks to make the results scalable. The optimal performance of this approach has been demonstrated in practice in a series of case studies, offering advanced reasoning for the identification of intricate relations among data and the ability to build predictive algorithms (e.g., vehicle speed prediction [23] or overlapping networks).

Explainable AI (XAI) algorithms combine reasoning and learning. Separately, both have been the focus of countless researches over the years. Now, the interest of the research community lies in uniting these two approaches [24]. These algorithms join the conceptual approach of symbolic reasoning, which offers human readable results, with connectionist models so that links between the symbols may be created [25].Trustworthy AI [26] is the basis that underpins XAI, striving towards AI developments being ethical, constitutional and reliable.

This research makes progress in the field of XAI, which has attracted the attention of the scientific community because it is able to provide solutions that can be explained, for example, by transparently showing the process in machine learning when decisions are made [27].

Symbols have the capacity to adapt their knowledge to different contexts by transferring reward signals. Symbolic artificial intelligence prevailed in AI studies in the post-war era, until the late 1980s [28]. Symbolic developments include rule engines, expert systems or knowledge graphs (e.g., Google's Knowledge Graph [29]). There are differences in the learning of Machine Learning and traditional symbolic reasoning. In Machine Learning algorithms, the input data is correlated with the outputs to created new rules. In turn,

symbolic reasoning, systems are static, they require a human expert to intervene in the process by establishing and codifying the rules that link two phenomena. However, symbolic systems encounter some hindrances: (i) expert systems are monotonous [30], that is, as new rules are added, the system's knowledge grows, however, he system is unable to unlearn a rule if the new one contradicts the prior. (ii) The computer does not understand the meaning of the symbols (contrary to neural networks, which associate symbols to vectorized representations of data).

To provide the system with explainable results, a model is proposed that combines the numerical solution of a system of differential equations, modeling the dynamics of the disease on the basis of coefficients which can be interpreted. This is combined with a deep learning system based on recurrent neural networks to predict the evolution over time of these coefficients. In addition, the system allows for the incorporation of additional hypotheses about the containment measures taken. In conclusion, the system allows us to interpret and adapt its predictions on the basis of these coefficients.

3. Proposed Model and Case Study: Caldas COVID-19 Evolution

To obtain predictions on the evolution of the disease in Caldas, a hybrid model has been proposed in which artificial intelligence and an expert system are incorporated into the SIR epidemiological model. A diagram of this model can be seen in Figure 1. The model is described in more detail below.
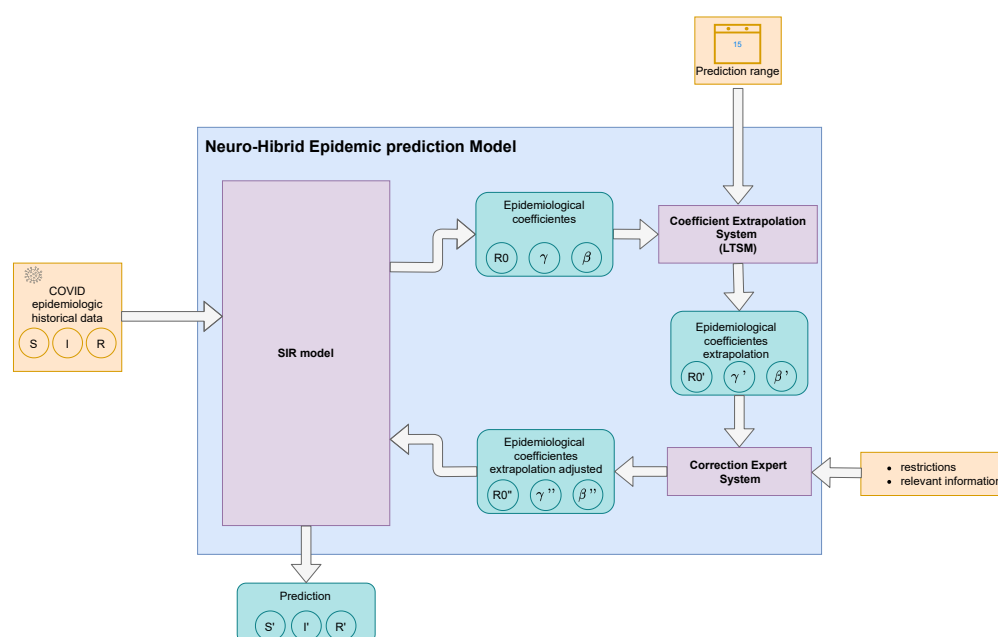

Figure 1. A historical dataset is used to calculate how COVID-19 evolves over a certain period of time and the curves S (the time-dependent susceptible population), I (the time-dependent infected population) and R (the time-dependent removed (recovered, death) population) are extracted over this period of time. These variables are used to fit an SIR model using sliding windows. The differential equations are solved using a Runge–Kutta method and the fit is performed in the sense of mean squares, thus extracting the unknown parameters of the model: β and γ, as well as the basic reproductive number R_0, all of them being functions of time. To extrapolate these parameters to higher time values (β', γ', R_0'), an LSTM neural network is used, the results of which are further refined by using an expert system that takes into account possible future changes in the constraints imposed by the government (β'', γ'', R_0'' in the diagram). Finally, by solving the SIR model again with these extrapolated coefficients, the predictions in the evolution of the S', I' and R' curves are obtained.

3.1. Extraction of Input Variables

Let $d. \in D$ be a tuple with information on patients who have been infected by COVID-19 in Caldas and whose components we denote as d.diagnostic_date. A representation of the information for each patient is required by this model, as shown in Table 1.

Table 1. Input data.

Case ID	Symptoms Onset Date	Diagnostic Date	Recovery Date	Death Date
2592	2020-03-26	2020-04-11	2020-04-23	-
2593	2020-03-25	2020-04-11	2020-05-25	-
2594	2020-03-25	2020-04-11	2020-05-05	-
2595	2020-03-25	2020-04-11	2020-05-16	-
2596	2020-03-28	2020-04-11	-	2020-04-01
2597	2020-03-22	2020-04-11	2020-04-16	-
2699	2020-04-09	2020-04-11	2020-04-21	-

From individual instances, the curves of each compartment of the SIR model are obtained using

$$
\begin{cases}
S(t) = \#\{d \in D | d.\text{infection\_start\_date} < t\}, \\
I(t) = \#\{d \in D | (d.\text{infection\_start\_date} \geq t) \cap (d.\text{recovery\_date} < t \\
\qquad\qquad \cup (d.\text{death\_date} < t)\}, \\
R(t) = \#\{d \in D | d.\text{recovery\_date} \geq t \cup d.\text{death\_date} \geq t\}.
\end{cases}
$$

where the infection start date (infection_start_date) is identified with the date when symptoms began to appear, or with the date of the diagnosis in the case of asymptomatic individuals.

3.2. SIR Model

The SIR model is an epidemiological model that is given by the following system of differential equations:

$$
\begin{cases}
\dfrac{\mathrm{d}S}{\mathrm{d}t} = -\beta \cdot \dfrac{IS}{N} \\
\dfrac{\mathrm{d}I}{\mathrm{d}t} = \beta \cdot \dfrac{IS}{N} - \gamma I \\
\dfrac{\mathrm{d}R}{\mathrm{d}t} = \gamma I
\end{cases}
\tag{1}
$$

where S corresponds to the number of people without immunity to the infectious agent, i.e., susceptible population, I to the infected population and R to the individuals who are immune to the infection (recovered or deceased). All these variables have a time dependence that has been omitted from the notation to make it clearer. In addition, the model states that

$$
N = S + I + R
\tag{2}
$$

where N is a constant that refers to the total population. For the case we are dealing with, Caldas, takes the value of 993870 [31].

Therefore, the parameters of this model are β and γ, variables that evolve over time. These are obtained by solving the system of differential Equation (1) with a sliding window adjustment. The Runge–Kutta 45 method is used to solve it, this means a Runge–Kutta whose error is controlled by assuming the accuracy of the fourth-order method but taking fifth-order steps in the accuracy formula.

3.3. Coefficient Extrapolation

The series $\beta(t)$ is extrapolated using a recurrent neural network called LSTM [32].

The series $\gamma(t)$ is extrapolated by taking the median of the series $\gamma(t)$ for $t < t_{max}$. This parameter is does not fluctuate much as it is the inverse of the time it would take for a person to recover from the disease, therefore, it is a constant.

LSTM is a type of neural network where each memory cell, c_j, is built around a central linear unit with a fixed self-connection (the CEC). In addition to the net_{c_j}, c_j obtains input from a multiplicative unit out_j (the "output gate"), and from another multiplicative unit in_j (the "input gate"), holding that

$$
\begin{aligned}
y^{\mathrm{out}_j}(t) &= f_{\mathrm{out}_j}(net_{\mathrm{out}_j}(t)) \\
y^{\mathrm{in}_j}(t) &= f_{\mathrm{in}_j}(net_{\mathrm{in}_j}(t))
\end{aligned}
\tag{3}
$$

where

$$
\begin{aligned}
\mathrm{net}_{\mathrm{out}_j}(t) &= \sum_u w_{\mathrm{out}_j u} y^u(t-1) \\
\mathrm{net}_{\mathrm{in}_j}(t) &= \sum_u w_{\mathrm{in}_j u} y^u(t-1)
\end{aligned}
\tag{4}
$$

and

$$
\mathrm{net}_{c_j}(t) = \sum_u w_{c_j u} y^u(t-1).
\tag{5}
$$

The index u of the summations sweeps through the different memory cells present in the neural network. Therefore, at time t the output of the cell c_j, which we denote by $y^{c_j}(t)$, is defined as

$$
y^{c_j}(t) = y^{\mathrm{out}_j}(t) h(s_{c_j}(t)),
\tag{6}
$$

where the internal state $s_{c_j}(t)$ is

$$
\begin{cases}
s_{c_j}(0) = 0 \\
s_{c_j}(t) = s_{c_j}(t-1) + y^{\mathrm{in}_j}(t) g(\mathrm{net}_{c_j}(t)) & \text{for } t > 0.
\end{cases}
\tag{7}
$$

With this structure, LSTM is able to overcome the problems of error backpropagation by either blowing up or vanishing, due to the fact that the time evolution of the backpropagated error depends exponentially on the size of the weights. In fact, the multiplicative input gate unit protects the memory contents stored in j from perturbations by irrelevant inputs just as the multiplicative output gate unit protects the other units from perturbations by currently irrelevant memory contents stored in j.

In this work, a multi-step LSTM with tanh activation for the hidden state and the output hidden state, and a sigmoid activation for the input, forget, and output gates was used to perform predictions up to a 20-day horizon.

3.4. Expert System for Restraint Measure Modeling

The implementation of different measures that the government of Caldas is forced to impose in order to stop the expansion of the disease in critical moments, causes considerable changes in the $\beta(t)$ parameter of the SIR model. For this reason, it is necessary to incorporate an expert system that allows to make predictions on the basis of the possible exceptional measures that may be imposed in the future to counteract the pandemic.

The expert system's key is, depending on the type of legal measure applied by the government (isolation, restriction of mobility,...), to modify the parameter $\beta(t)$ for $t > t_{\max}$, which was obtained with the LSTM neural network. To extract some general rules with which to modify these parameters, it is first necessary to classify the measures according to their degree of restriction or relaxation and observe the effect that each of the measures has had in the past (Figure 2).
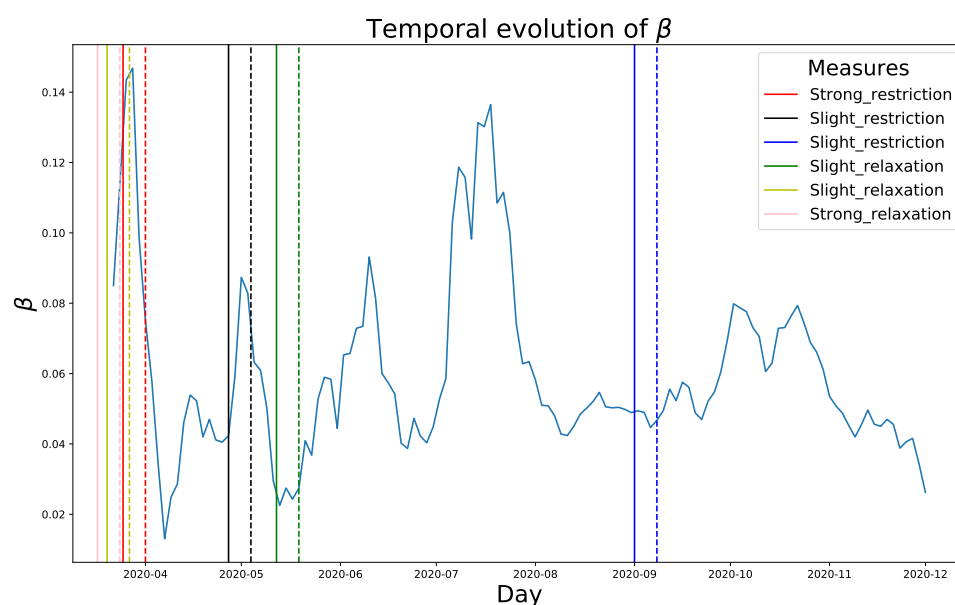
Figure 2. Representation of the variation of the β parameter without applying the expert system (blue curve). The horizontal coordinate represents time, while the vertical coordinate represents this coefficient. The higher the value is, the more contacts have infected individuals. The solid vertical lines represent the dates on which the government imposed a new measure and the dotted vertical lines are the approximate dates on which the consequences of these measures are expected to become visible. This time lag, which has been considered to be 10 days, is due to the existence of an incubation period of the disease, where no symptoms are present even though contagion has occurred, and an adaptation period until the measure is fully established.

Given that we have few measurements to be able to carry out a detailed statistical study, on the basis of the observations made in Figure 2 some general approximations have been extracted from the increase or decrease in spread rate after the implementation of each type of measure. These are summarized in Table 2.

Table 2. Percentage of change according to the type of measure applied, based on the date from the measures taken in Caldas (Figure 2).

Government Measures	Percentage Change	Target Change
Strong restriction	−30%	0.7
Slight restriction	−10%	0.9
Slight relaxation	+10%	1.1
Strong relaxation	+30%	1.3

Taking into consideration that the effect a measure has on the spread rate is not visible for a few days after the implementation, due to the incubation period of the disease (t time lag) and the adaptation period until the restriction becomes fully effective (k time lag), a sigmoidal function (Figure 3) is implemented to modulate these changes in β.

$$f_{t,k}(x) = \begin{cases} 1 & \text{if } x \leq t, \\ \frac{\sigma(x)-\sigma(t)}{\sigma(t+k)-\sigma(t)}(\text{Target\_change} - 1) + 1 & \text{if } t < x \leq t+k, \\ \text{Target\_change} & \text{if } t > t+k, \end{cases} \tag{8}$$

where

$$\sigma(x) = \frac{1}{1+e^{-x}}.$$

As a result of the decomposition carried out by the model and the ability to introduce these additional assumptions, the resulting system makes it possible to understand the causes of the prediction, adapting to the hypotheses to be made about them.
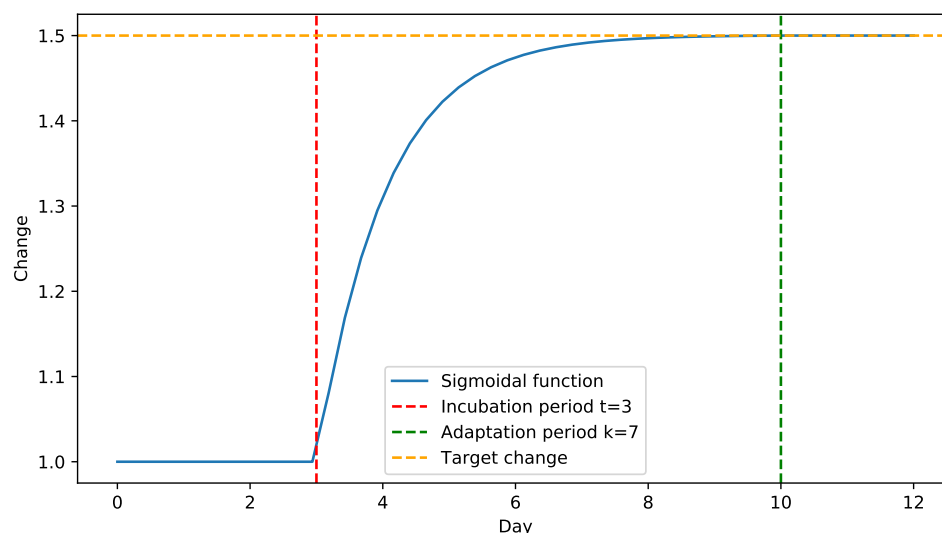

Figure 3. Representation of the sigmoidal function that has been used when applying a strong relaxation measure, assuming an incubation period *t* of 3 days and an adaptation period *k* to the new measure of 7 days. As a result, a smooth transition transition is obtained when a measure is considered in the expert model.

3.5. Platform Achitecture

To tackle this problem, it is proposed to use a microservices-oriented architecture that favours the growth of the system, providing it with flexibility, horizontal and vertical scalability (Figure 4). To achieve this, the following subsystems must be developed:

- Data extraction: periodically (once a day) extracts statistics on the impact of the COVID-19 pandemic from Colombia's open portal (https://www.datos.gov.co/, accessed on 22 February 2021).
- Deep intelligence: This is a platform used to manage the different workflows in projects in the field of Big Data, Artificial Intelligence and Smart Cities. Specifically, it is used in the following ways:
  - Main data warehouse, as it stores both the data resulting from extraction and the data resulting from predictions.
  - Visualization engine: it is able to generate interactive visualizations with the stored data to show the result of the whole process in the most summarized way possible to the user.
- Data analysis: this is a system that periodically takes Deep Intelligence data on the indicators of COVID-19 impact in Colombia, in order to make the relevant predictions with the hybrid model explained above. Finally, it stores the results in Deep Intelligence.

As a result of the infrastructure built, a system for monitoring multiple variables is available, including those from which the variables necessary for the construction of the model are derived.

The origin of all the information recorded in the Epidemiological Intelligence System (Sistema de Inteligencia Epidemiológica (SIE)) comes from the information system that centralizes the data of the department of Caldas called registro COVID [33]. The information in this source is synchronized every hour.

Among the information included within the monitoring system, the following can be found:

- Personal data
    - ID: it is the patient identifier; the identification document is encrypted to preserve the privacy of the information.
    - Municipality: name of the municipality where the patient resides.
    - Department: name of the department where the patient resides.
    - Age Criteria: identifies whether the patient is an adult (A) or a minor (M)
    - Age: it is the numerical value of the patient's age at the time of registration.
    - Gender: gender identifier of the patient, may be female (F) or male (M).
    - Neighborhood/Village: indicates the name of the neighborhood or village of the patient's residence.
    - Address: the exact location of the patient.
    - Patient state: identifies if the patient has recovered or deceased.
    - Management: Describes whether the person was hospitalized (H) or stayed at home (C)
    - Type of contagion: states if the contagion has been within the community, from a relative or imported.
    - State of severity of the patient: a difference between mild, moderate or serious.
- General section:
    - Alert: allows to register the test status of the patient: in study, discarded, confirmed.
    - Record date: corresponds to the date on which the patient information was recorded.
    - Onset of symptoms date: corresponds to the date the patient started having symptoms.
    - Diagnosis date: date the patient's infection was confirmed.
    - Recovered date: corresponds to the date on which the patient was recovered.
    - Date of death: date of the patient's death.

Some of the steps in the construction of this model within the Deep Intelligence platform are shown in Figure 5.
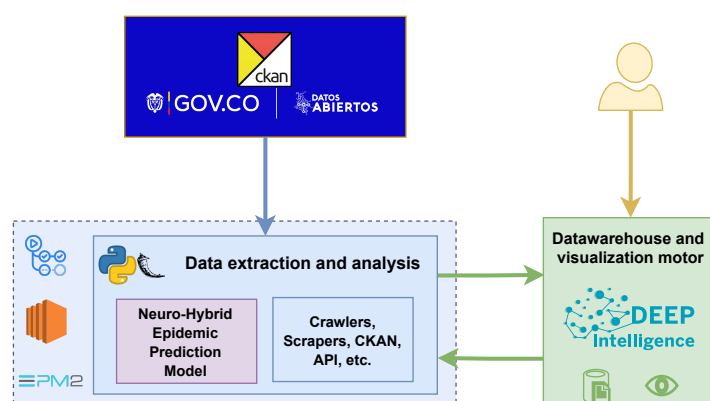

Figure 4. Architecture used to implement the proposed predictive model. The data is obtained from the CKAN server in Colombia's open portal (**upper left**), using a Python-implemented API served with Flask (**lower left**), which also accesses additional information sources. The interaction with the Deep Intelligence platform (**lower right**) is performed by means of its API. Users (**upper right**) are able to access the platform using its graphical interface.

(**a**)

(**b**)

(**c**)

(**d**)

Figure 5. Some of the steps performed to build the system in Deep Intelligence [34]. (**a**) A data source in the platform, showing structured information. (**b**) Monitoring of additional indicators using the visualization tools provided by the platform. (**c**) Global curve visualization from an external source. (**d**) Visualization of the predictions obtained by the model.

4. Results

This section analyzes the results obtained from the conducted case study, as described in detail in the previous section. For this purpose, the chosen input sample is the data collected during 3 months, from mid-August to mid-November, taken from the web [33], where data is accessible under an open license. Using these data it is possible to compare the differences between what happened in reality and the different predictions that the model is able to make. Figure 6 shows the obtained predictions using different assumptions, starting from a randomly chosen date (4 September). That day was a weekday when no change in government measures was imposed, so the obtained results are in line with expectations. The predictions extend up to a 20-day horizon. Further details on the prediction error are given below.
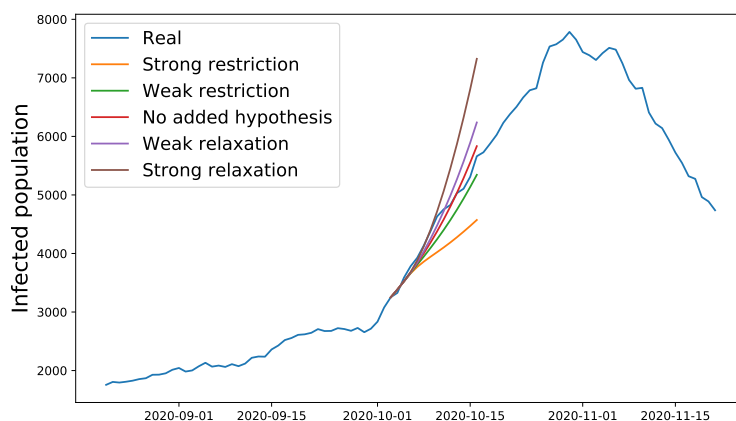

Figure 6. Prediction of the evolution of the number of positive cases from 4 September. The series show the different predictions when various hypothesis on the restriction measures are used in the system. The prediction curves extend up to 20 days, which is the intended prediction range.

On the other hand, Figure 7 shows the obtained results, assuming that no restrictive measures are applied, with an estimated 80% confidence interval.
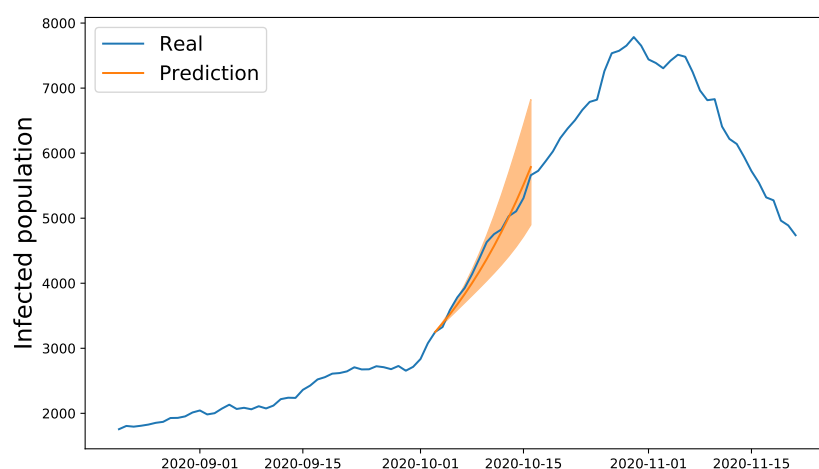

Figure 7. Prediction of the evolution of the number of positive cases from 4 September with a confidence interval of 80 %, assuming that no restrictive measures are applied. The prediction extends up to 20 days, which is the intended prediction range.

To provide more insight on the error distribution, an example is shown in Figure 8a for the case of a one week prediction horizon, aggregated metrics can be provided using its absolute error to prevent sign compensation. The mean value of the absolute value of the relative error is thus depicted in Figure 8b, along with the sample standard deviation of its distribution.
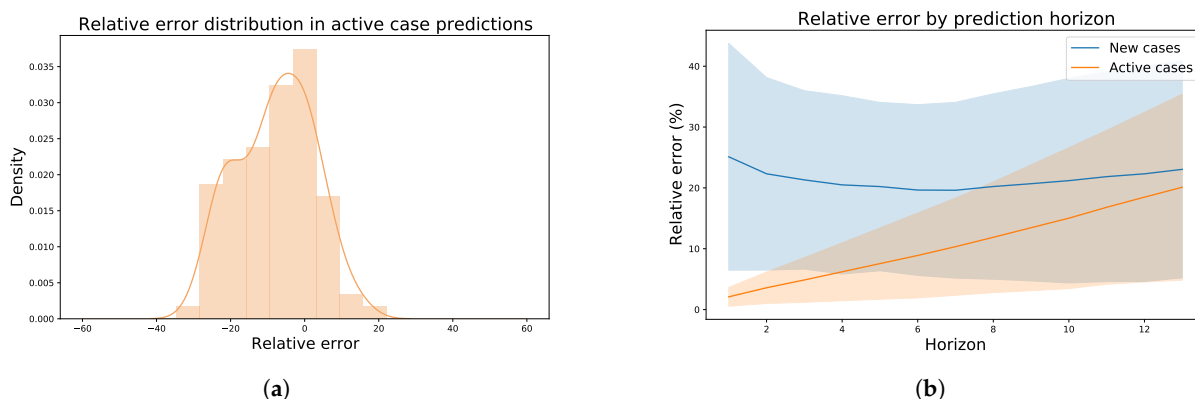

(**a**)

(**b**)

Figure 8. Analysis of the error distribution in the prediction of infected individuals. (**a**) Distribution of the relative error for a 7-day prediction horizon. (**b**) Error as a function of the prediction horizon. Solid lines depict the mean of the absolute value of the relative error, while the shaded regions depict its sample standard deviation.

Overall the relative error in the number of new cases is typically around 25%. The error in the number of active cases is usually below 10% for near-future predictions (below one week), and increase linearly once this limit is extended.

5. Conclusions and Future Work

Predicting the spread of COVID-19 is a complicated task, due to the impact of contingency measures, which continually change the population dynamics. In this work, a prediction system has been proposed that ensures that the evolution occurs within the framework of the dynamics of the differential equation systems of the disease. This makes it possible to interpret the results of the system from the point of view of the measures taken. In addition, a mixture of experts system for the correction of predictions has been designed

to incorporate additional hypotheses on the contingency decisions taken. The system is built with an SIR model and an LSTM, and is capable of representing the state of the pandemic at a given time and projecting it into the future, so that any variations that appear model the system and it is fed back. The recurrent network adapts the coefficients of the SIR model and these explain the evolution of the system and possible changes. The model corrects the prediction once the SIR model identifies structural changes in the way the pandemic evolves and adapts its coefficients and results. In the same way, changes can be introduced into the model once changes in mobility constraints are known, for example. The decomposition made by the model facilitates the understanding of the results, as well as the incorporation of additional assumptions about the restrictions.

The mean squared error in terms of the number of positive cases has been estimated to be around 18% and 22% for the active and new cases respectively in the 2-week predictions. The proposed model has been extended to predict in the medium term 4-8 months and assess the evolution of the pandemic until remission by taking into account the different vaccination scenarios, possible restrictions, mutation effects and regulatory changes. The aim is to create a real decision-support system capable of providing support in making an effective response to the pandemic.

Author Contributions: Conceptualization, L.F.C.O.; data curation, F.P.-S.; formal analysis, G.H.; funding acquisition, A.C.-M.; investigation, P.C., J.A.-L., G.A.I., A.C.-M. and J.M.C.; methodology, J.M.C.; project administration, P.C. and J.M.C.; software, C.S.-C.-G. and G.H.; supervision, A.C.-M. and J.M.C.; writing—original draft, J.M.C.; writing—review and editing, C.S.-C.-G. and G.H. All authors have read and agreed to the published version of the manuscript.

References

1. Abbas, A.; Abdelsamea, M.M.; Gaber, M.M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *arXiv* **2020**, arXiv:2003.13815.
2. Wang, L.; Lin, Z.Q.; Wong, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci. Rep.* **2020**, *10*, 1–12. [CrossRef] [PubMed]
3. Ye, Y.; Hou, S.; Fan, Y.; Qian, Y.; Zhang, Y.; Sun, S.; Peng, Q.; Laparo, K. α-Satellite: An AI-driven System and Benchmark Datasets for Hierarchical Community-level Risk Assessment to Help Combat COVID-19. *arXiv* **2020**, arXiv:2003.12232.
4. Hu, F.; Jiang, J.; Yin, P. Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model. *arXiv* **2020**, arXiv:2003.00728.
5. Ge, Y.; Tian, T.; Huang, S.; Wan, F.; Li, J.; Li, S.; Yang, H.; Hong, L.; Wu, N.; Yuan, E.; et al. A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *bioRxiv* **2020**. [CrossRef]
6. Ardakani, A.A.; Kanafi, A.R.; Acharya, U.R.; Khadem, N.; Mohammadi, A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Comput. Biol. Med.* **2020**, *121*, 103795. [CrossRef]
7. Chimmula, V.K.R.; Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* **2020**, *135*, 109864. [CrossRef]
8. Tang, Z.; Zhao, W.; Xie, X.; Zhong, Z.; Shi, F.; Liu, J.; Shen, D. Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. *arXiv* **2020**, arXiv:2003.11988.
9. Wu, J.; Zhang, P.; Zhang, L.; Meng, W.; Li, J.; Tong, C.; Li, Y.; Cai, J.; Yang, Z.; Zhu, J.; et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. *medRxiv* **2020**. [CrossRef]
10. Imran, A.; Posokhova, I.; Qureshi, H.N.; Masood, U.; Riaz, S.; Ali, K.; John, C.N.; Nabeel, M. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *arXiv* **2020**, arXiv:2004.01275.
11. Chen, Y.C.; Lu, P.E.; Chang, C.S. A Time-dependent SIR model for COVID-19. *arXiv* **2020**, arXiv:2003.00122.

12. Hu, Z.; Ge, Q.; Li, S.; Boerwincle, E.; Jin, L.; Xiong, M. Forecasting and evaluating intervention of Covid-19 in the World. *arXiv* **2020**, arXiv:2003.09800.
13. Chenthamarakshan, V.; Das, P.; Padhi, I.; Strobelt, H.; Lim, K.W.; Hoover, B.; Hoffman, S.C.; Mojsilovic, A. Target-specific and selective drug design for covid-19 using deep generative models. *arXiv* **2020**, arXiv:2004.01215.
14. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Acharya, U.R. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* **2020**, *121*, 103792. [CrossRef] [PubMed]
15. Ribeiro, M.H.D.M.; da Silva, R.G.; Mariani, V.C.; dos Santos Coelho, L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos Solitons Fractals* **2020**, *135*, 109853. [CrossRef]
16. Yan, L.; Zhang, H.T.; Goncalves, J.; Xiao, Y.; Wang, M.; Guo, Y.; Sun, C.; Tang, X.; Jing, L.; Zhang, M.; et al. An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.* **2020**, *2*, 283–288. [CrossRef]
17. Bui, D.T.; Bui, Q.T.; Nguyen, Q.P.; Pradhan, B.; Nampak, H.; Trinh, P.T. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agric. For. Meteorol.* **2017**, *233*, 32–44.
18. Hoang, N.D.; Pham, A.D. Hybrid artificial intelligence approach based on metaheuristic and machine learning for slope stability assessment: A multinational data analysis. *Expert Syst. Appl.* **2016**, *46*, 60–68. [CrossRef]
19. Fdez-Riverola, F.; Corchado, J.M. Forecasting red tides using an hybrid neuro-symbolic system. *AI Commun.* **2003**, *16*, 221–233.
20. González Bedia, M.; Corchado Rodríguez, J.M. *A Planning Strategy Based on Variational Calculus for Deliberative Agents*; University of Paisley: Scotland, UK, 2002.
21. Garcez, A.; Besold, T.R.; Raedt, L.D.; Fosddldiak, P.; Hitzler, P.; Icard, T.; Kühnberger, K.U.; Lamb, L.C.; Miikkulainen, R.; Silver, D.L. Neural-Symbolic Learning and Reasoning: Contributions and Challenges. In Proceedings of the 2015 AAAI Spring Symposium Series, Palo Alto, CA, USA, 23–25 March 2015.
22. Kaur, R.; Sangal, A.L.; Kumar, K. Modeling and simulation of adaptive neuro-fuzzy based intelligent system for predictive stabilization in structured overlay networks. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 310–320. [CrossRef]
23. Cheng, Z.; Chow, M.Y.; Jung, D.; Jeon, J. A big data based deep learning approach for vehicle speed prediction. In Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 19–21 June 2017; pp. 389–394.
24. Garcez, A.d.; Gori, M.; Lamb, L.C.; Serafini, L.; Spranger, M.; Tran, S.N. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *arXiv* **2019**, arXiv:1905.06088.
25. Garcez, A.D.; Dutra, A.R.R.; Alonso, E. Towards symbolic reinforcement learning with common sense. *arXiv* **2018**, arXiv:1804.08597.
26. Di Maio, P. Neurosymbolic Knowledge Representation for Explainable and Trustworthy AI. *Preprints* **2020**, 2020010163. [CrossRef]
27. Zhang, Q.; Sornette, D. Learning like humans with Deep Symbolic Networks. *arXiv* **2017**, arXiv:1707.03377.
28. Haugeland, J. *Artificial Intelligence: The Very Idea*; MIT Press: Cambridge, MA, USA, 1989.
29. Singhal, A. Introducing the knowledge graph: things, not strings. *Off. Google Blog* **2012**, *5*. Available online: https://blog.google/products/search/introducing-knowledge-graph-things-not/ (accessed on 22 February 2021).
30. McDermott, D. A critique of pure reason. *Comput. Intell.* **1987**, *3*, 151–160. [CrossRef]
31. Colombia Population—Worldometer. 2021. Available online: https://www.worldometers.info/world-population/colombia-population/ (accessed on 22 January 2021).
32. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 22 February 2021).
33. Casos Positivos de COVID-19 en Colombia | Datos Abiertos Colombia. Available online: https://www.datos.gov.co/en/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data (accessed on 22 January 2021).
34. Corchado, J.M.; Chamoso, P.; Hernández, G.; Gutierrez, A.S.R.; Camacho, A.R.; González-Briones, A.; Pinto-Santos, F.; Goyenechea, E.; Garcia-Retuerta, D.; Alonso-Miguel, M.; et al. Deepint. net: A Rapid Deployment Platform for Smart Territories. *Sensors* **2021**, *21*, 236. [CrossRef]