Variational-LSTM Autoencoder to forecast the spread of coronavirus across the globe

Mohamed R. Ibrahim[1], James Haworth, Aldo Lipani, Nilufer Aslam, Tao Cheng and Nicola Christie

Department of Civil, Environmental and Geomatic Engineering,
University College London (UCL)
[1]Corresponding author: mohamed.ibrahim.17@ucl.ac.uk,

***Abstract*—** Modelling the spread of coronavirus globally while learning trends at global and country levels remains crucial for tackling the pandemic. We introduce a novel variational LSTM-Autoencoder model to predict the spread of coronavirus for each country across the globe. This deep spatio-temporal model does not only rely on historical data of the virus spread but also includes factors related to urban characteristics represented in locational and demographic data (such as population density, urban population, and fertility rate), an index that represent the governmental measures and response amid toward mitigating the outbreak (includes 13 measures such as: 1) school closing, 2) workplace closing, 3) cancelling public events, 4) close public transport, 5) public information campaigns, 6) restrictions on internal movements, 7) international travel controls, 8) fiscal measures, 9) monetary measures, 10) emergency investment in health care, 11) investment in vaccines, 12) virus testing framework, and 13) contact tracing). In addition, the introduced method learns to generate graph to adjust the spatial dependences among different countries while forecasting the spread. We trained two models for short and long-term forecasts. The first one is trained to output one step in future with three previous timestamps of all features across the globe, whereas the second model is trained to output 10 steps in future. Overall, the trained models show high validation for forecasting the spread for each country for short and long-term forecasts, which makes the introduce method a useful tool to assist decision and policymaking for the different corners of the globe.

***Keywords*—** Coronavirus spread, pandemic, forecast, Variational LSTM- Autoencoder, COVID-19, deep learning

1. INTRODUCTION

As a new contagious disease in human inhabitants, COVID-19, has been currently reaching 803,126 confirmed cases with 39.032 death in 201 countries across the World (Wordometer, 2020). Although there are a number of the statistical and epidemic models to analyse COVID-19 outbreak, the models are suffering from many assumptions to evaluate the impact of intervention plans which create a low accuracy as well as unsure prediction (Hu et al., 2020). Therefore, there is a vital need to develop new frameworks/methods to curb/control the spread of Coronavirus immediately (Botha and Dednam, 2020; Hu et al., 2020).

The epidemic outbreak of COVID-19 in literature is investigated using mathematical compartmental model named Susceptible-Infected-Recovered (SIR) (Kermack and McKendrick, 1927). The SIR model represents a population under three categories: 1) Susceptible (the number of people presently not infected), 2) the number of people currently infected, and 3) the number of people either recovered or died. The model describes as differential equations. The model is completely determined by transmission rate, the recovery rate, and the initial condition, which can be estimated using least square error, Kalman filtering or BMC. The model is sometimes renamed based on the new parameters such as Susceptible-Infectious-Quarantined-Recovered (SIQR) or Susceptible-Exposed-Infected-Recovered (SEIR). The main idea in the version of all SIRs models are four-fold; first, identification and better understanding current epidemic (Crokidakis, 2020), second, simulation the behaviour of the system (Castro, 2020), third, forecasting of the future behaviour (Toda, 2020), and last, how we control the current situation (Sameni, 2020). However, the results of the models including accuracy only valid based on their assumptions in a slice of available data/moment and have their scopes to assist healthcare strategies for the decision-making process.

On the other hand, agent-based modelling is utilised to explore and estimate the number of contagions of COVID-19, specifically for certain countries (Chang et al., 2020; Simha et al., 2020). Also, statistical methods (Singer, 2020), simple time series modelling (Deb, 2020), and logistic map (Al-qaness et al., 2020) are utilised for similar objectives, whereas (Botha and Dednam, 2020), focused on modelling the spread of coronavirus based on the parameters of basic SIR in a (3-dimensional) iterative maps to provide a wider picture of the globe. Petropoulos and Makridakis (2020) forecasted the total global spread relying on exponential smoothing model based only on historical data. Put all together, the drawbacks of their models are not flexible to fit for each country or region due to the lack of necessary measures, government responses, and spatial factors related to each specific location.

There are few examples of predictive modelling of the coronavirus spread based on machine learning approaches, whether through shallow or deep models. While it is can be explained due to the limitation of data since the early stage of the outbreak, it remains an essential tool. According to Pham and Luengo-oroz (2020), machine learning approaches certainly could assist in forecasting by with improved quality for prediction. One of the few studies is presented by (Hu et al., 2020). They have applied real-time short-term forecasting using the compiled data from 11th Jan to 27th Feb 2020 collected by the World Health Organization (WHO) for the 31 provinces of China. The data is trained on a deep learning model for real-time forecasting of new cases for the provinces. Their model has the flexibility to be trained at the city, provincial, or national level. Besides, the latent variable of the trained model is used to extract necessary features for each region and fed into a K-

means to cluster similar features of the infected or recovered features of patients. Bearing this in mind, there is still a knowledge gap for machine learning models to predict coronavirus cases at a global as well as regional scales (Pham and Luengo-oroz, 2020).

While SIR models with their different types, in addition to the aforementioned ones, are essential, the challenges remain in forecasting different regions and countries across the globe with a single model without any assumptions or scenario-based rules, but only with the current situations, features related to countries, and measures amid to reduce the impact of the outbreak. Accordingly, in this paper, we introduce a new method of learning and encoding information related to the historical data of coronavirus per country, features of countries, spatial dependencies among the different countries, and last, the time and location-dependent measures taken by each country amid towards reducing the impact of Coronavirus. Relying on deep learning, we introduce a novel variational Long-Short Term Memory (LSTM) autoencoder model to forecast the spread of coronavirus per country across the globe. This single deep model aimed to provide robust assistance to policymakers to understand the future of the pandemic at both a global level and country level, for a short-term forecast and long-term one. The main advantages of the proposed method are: 1) It can structure and learns from different data sources, either that belongs to spatial adjacency, urban and population factors, or various historical related data, 2) the model is flexible to apply to different scales, in which currently, it can provide prediction at global and country scales, however, it can be also applied to city level. And last 3) the model is capable of learning global trends for countries that have either similar measures, spread patterns, or urban and population features.

After the introduction, the article is structured in five sections. Section 2 introduces the method and materials used. In section 3, we show model evaluations and the experimental results at country and global levels. In section 4 we discuss our results, compare our model to any existing base models and highlights limitations. Last, in section 5 we conclude and present our recommendation for future works.

2. METHODS

2.1 Hypothesis and assumptions

The model algorithms are constructed based on four assumptions that we assume the model needs to learn to predict the next day spread: First, the model needs to extract features regarding the historical data of coronavirus spread for a given country bearing in mind the historical values of the virus spread in the other countries simultaneously before it outputs a prediction for a given country. Second, before the model gives a predicted value for each country, it should consider the predicted values of all other countries instantaneously, similar to the first point. Third, the spatial relationship between different countries is multidimensional; it can vary based on geographical location, adjacency, accessibility, or even policies for banning accessibility. The model needs to deal with variations of time and location of the different inputted scenarios while sampling outcomes. Last, apart from the virus features, for each country, there are unique demographic and geographical features that show association to the spread of the virus that may show association with the virus, in which the learning process of the model needs to consider each time before it gives a predicted value.

The structure of the input data is key for any model to learn. Figure 1 shows the concept of the overall structure of the proposed graph of multi-dimensional data sets for forecasting the spread. It illustrates how different types of data can be linked and clustered for the model to learn the spread of a virus. This data can be seen as dynamic features related to both virus and the location with long temporal scales (i.e. the population data) or short ones (t_i). It shows how local and global trend for a virus can be forecasted for a given country (n_z), with urban features that include both spatial and demographic factors (x_m), that share a spatial weight (g_j) with other countries in the graph, whereas government mitigated measures (r_q) are applied. Put all together, the model needs to differentiate between factors that characterise countries or regions, and those which characterise the virus spread to understand the patterns of spread at global and country levels.

2.2 Translation to the machine

To meet these hypotheses and assumptions during the learning process, the architecture of the proposed model is based on the combinations of three main components: 1) LSTM, 2) Self-attention, and 3) Variational autoencoder graph.

2.2.1 LSTM cells

LSTM represents the main component of the proposed model. It has been shown it is the ability to learn long-term dependencies easier than a simple recurrent architecture (Goodfellow et al., 2017; LeCun et al., 2015). Unlike traditional recurrent units, it has an internal recurrence or a self-loop, in which it allows the timestamps to create paths, in which the gradient of the model can flow for a long duration without facing the vanishes issues presented in a normal recurrent unit. Even for an LSTM with a fixed parameter, the integrated time scale can change based on the input sequence, simply because the constants of time are outputted by the model itself. These self-loops are controlled by a forget gate unit ($f_i^{(t)}$) for a given time (t) and a cell (i), in which it fits this weight to a scaled

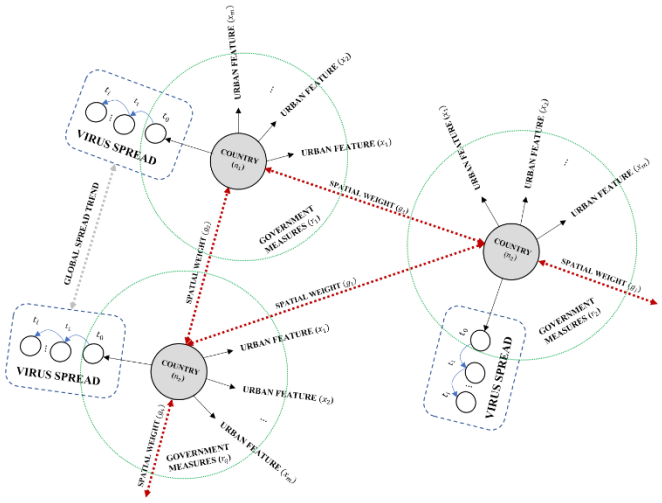

FIGURE 1
THE CONCEPT FOR STRUCTURING THE GRAPH FOR THE PROPOSED VARIATIONAL LSTM AUTOENCODER

value between 0,1 with a sigmoid unit (σ). It can be explained as:

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}\right) \quad (1)$$

Where $x^{(t)}$is a vector for the current input, $h^{(t)}$is a vector for the current hidden layer that contains the outputs of all the LSTM cells, b^f are the biases for the forget gates, U^f is the input weights, $W^f$is the recurrent weights for the forget gates. The internal state of the LSTM is updated with a conditioned self-loop weight ($f_i^{(t)}$) as:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}\right) \quad (2)$$

Where b represents biases, U represents input weights, W represents the current weights into the LSTM cell, and $g_i^{(t)}$ represents the external input gate unit. It is computed similar to the forget gate but with it is own parameters as:

$$g_i^{(t)} = \sigma\left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}\right) \quad (3)$$

Last, the LSTM cell output $h_i^{(t)}$ can also be controlled and shut off with an output gate $q_i^{(t)}$, similar to the aforementioned gate by using a sigmoid unit. The output $h_i^{(t)}$ is computed as:

$$h_i^{(t)} = {}_{\varphi}(s_i^{(t)}) q_i^{(t)} \quad (4)$$

$$q_i^{(t)} = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right) \quad (5)$$

Where b^o represents biases, U^o represents input weights, W^orepresents the current, and ${}_{\varphi}$.represents the activation function such as tanh function.

Put all together, this controls of the time scale and the forgetting behaviour of different units allow the model to learn long- and short-term dependencies for a given vector. Not only the model learns from the previously defined timestamps for each country, but also the model could extract features from the other countries at each given timestamp in which the dimension of the input vector, and cell states, includes the dimensions of the different countries. It is worth mentioning that the input for the LSTM cells is can be seen as a three-dimensional tensor, representing the sample size for both training and testing, the defined timestamps for the model to look back, and the timestamps of the other countries as a global feature extractor.

2.2.2 Self-attention mechanism

While the LSTM cells learn from their input sequence to output the predicted sequences through the long and short dependencies of the time constants and their additional features for each country, the relations between its inputs remains missing. A self-attention mechanism allows the LSTM units to understand the representation of its inputs by relating the positioning of each sequence (Goodfellow et al., 2017; Vaswani et al., 2017). This mechanism in the case of the proposed model is crucial to assist the model to which piece of information to consider and what to forget when making a prediction. The self-attention mechanism can be explained as:

2.2.3 Variational autoencoder graph

We initialise the first graph based on the spatial weight of the geographical locations of all infected countries (more details will follow in sub-section 3.1.4), however, despite the attempts of trying to create a sophisticated adjacency matrix among the infected countries (based on flight routes, spatial network, migration network, etc.), the output may misleading for any learning method over time or for a given location. The spatial weight since the outbreak of the model may look completely different from the initial day to the latest day. These due to different policies and measures that are taken by countries. However, due to its high uncertainty and variation. Inputting the model with a static graph or even a dynamic one based on limited data may exacerbate the learning process. Accordingly. the third vital components in our model represent the variational autoencoder (VAE) component that allows the model to generate information from a given input. It can be defined as a generative directed method that makes use of the learned approximate inference (Goodfellow et al., 2017; Ha and Schmidhuber, 2018). The model is based on the idea of passing latent variables $\boldsymbol{z}$ to the coded distribution $p_{model}(\boldsymbol{z})$ over samples $\boldsymbol{x}$ using a differentiable generator network $g(\boldsymbol{z})$. Subsequently, $\boldsymbol{x}$ is sampled from the distribution of $p_{model}(\boldsymbol{x}; \boldsymbol{g}(\boldsymbol{z}))$ which is equal to the distribution of $p_{model}(\boldsymbol{x}|\boldsymbol{z})$. The model is trained by maximising the lower bound of the variation $\mathcal{L}(q)$ that belongs to $\boldsymbol{x}$ as:

$$\mathcal{L}(q) = \mathbb{E}_{\boldsymbol{z}\sim q(\boldsymbol{z}|\boldsymbol{x})} \log p_{model}(\boldsymbol{z}, \boldsymbol{x}) + \mathcal{H}(q\,(\boldsymbol{z}|\boldsymbol{x})) \quad (6)$$

Equation (6) describes the joint log-likelihood of the visible and hidden variables under the approximate posterior over the latent variables $\log p_{model}(\boldsymbol{z}, \boldsymbol{x})$, and the entropy of the approximate posterior $\mathcal{H}(q(\boldsymbol{z}|\boldsymbol{x})$, in which q is chosen to be a Gaussian distribution with a noise that is added to the predicted mean value. In traditional variational autoencoder, the reconstruction log-likelihood tries to equalise the approximate posterior distribution $q(\boldsymbol{z}|\boldsymbol{x})$ and the model prior $p_{model}(\boldsymbol{x}|\boldsymbol{z})$. However, in the case of our model the encoded $q(\boldsymbol{z}|\boldsymbol{x})$ is conditioned and penalized based on the output of a predicted value of the next forecast of the spread, instead of the log-likelihood of the similarity with $p_{model}(\boldsymbol{x}|\boldsymbol{z})$, which will be explained further in the proposed framework.

2.3 Proposed model framework

We propose a sequence-to-sequence architecture relying on a mixture of VAE and LSTM. The model comprises two branches trained in parallel in an end-to-end fashion. Figure 2 shows the overall proposed framework.

The first branch is a self-attention LSTM model that feeds by Spatio-temporal data of coronavirus spread per day and per country, the government policies per day and per country, and the urban features per country, in which the vector is repeated to cover the duration of training (The urban features used are three features: population density, urban population percentage and fertility rate, which will be covered in detail in the upcoming section). Each input is reshaped as a 3D tensor of shape (samples, timestamps, number of features X number of countries). The three-input data are concatenated at the last axis of the data (the dimension of the feature) and passed to the first branch of the model through two parts: 1) the self-attention LSTM sequence encoder, and 2) the LSTM sequence decoder. The first sequence encodes the input data and extracts features for the second part of the LSTM sequence to output the prediction of the spread for the next day (in case of the short-term forecast) per country.

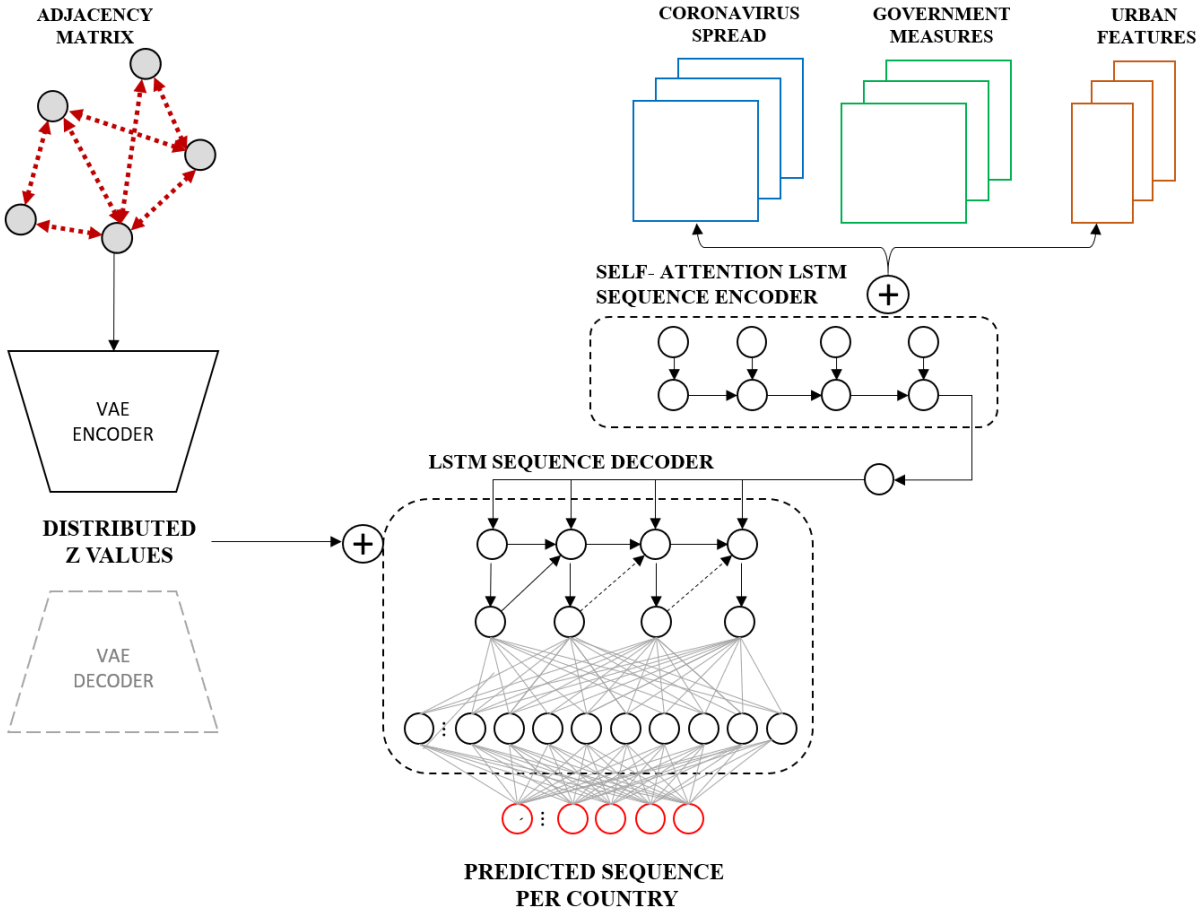

FIGURE 2
THE PROPOSED VARIATIONAL LSTM AUTOENCODER MODEL

In parallel to the self-attention encoder sequence, the second branch of the model is an encoder of VAE. It is feed by a spatial matrix of dimensions (number of countries X number of countries) and repeated for the entire duration of training and timestamps (In the next section, more details will follow on how it is selected and computed). This encoder part is mainly a convolution structure, which consists of three 1D convolution layers of filters 32, 64, and 128 respectively, in which they are all of a kernel size of 1 and activated by a ReLU function and followed by a Dropout layer of size 0.2. After the dropout, two LSTM layers are followed, in which they contain 100, 494 LSTM cells respectively. The first one is activated by a ReLU function, whereas the second one by a linear function. A fully connected layer of neurons equivalent to the number of countries is applied. Last the latent space is defined with a dimension of 10, in which the z-values are generated from sampling over the Gaussian distribution of the previous inputted layer (As explained in section 2.2.3). To visualise the generated graph for representation purposes, It is worth mentioning that the encoder of the second branch of the model can be decoded to output the generated samples for each predicted sequence by passing it into a decoder VAE, where the 1D convolutions layers are transposed to a final output shape equal to the inputted dimension. As for future work, this could be an interesting approach to understanding the variation of the graph for each predicted day for all countries.

Both outputs of the self-attention LSTM encoder and the encoder of the VAE are concatenated over the feature dimension and passed to the LSTM decoder sequence, which contains a single LSTM layer of cell numbers equal to the total number of countries. It is followed by two fully connected layers of shape size (1 X number of countries) for predicting the value of the next day, in case of the short-term forecast, or can be shaped to (numbers of future steps X number of countries) for any number of future steps that model needs to output per each country.

Data sets are split to training and testing on the first dimension of data shape (the total duration of the temporal data), in a way that the model can be tested on the last 6 days. We trained two different models, one as a single-step model for short-term forecast (one day), whereas the other is trained as a multi-step model (10 days forecast). There are two crucial differences between these two models; The output layer, and the dimension of the y-train, and y-test of the first one is shaped as (1 x n), whereas in the other model is output layer is shaped as (10 X n), despite the number of samples. is the structure of the y-train and y-test. The second issue, is the trained and tested sample is not only reduced by the number of timestamps – at the beginning of each sequence- as in the case of the first model, but also reduced by the number of future steps -at the end of the sequence- in the case of the second model. Last, based on trial and error, we structured the data based on 3 timestamps for both models to look back for all the input features for each country, in which we found optimal results.

The weights of the model are initialised by random weights. The model is compiled based on the backpropagation of error of the stochastic gradient descents, relying on ‘adam’ optimiser (Kingma and Ba, 2014), with a learning rate of 0.001 and

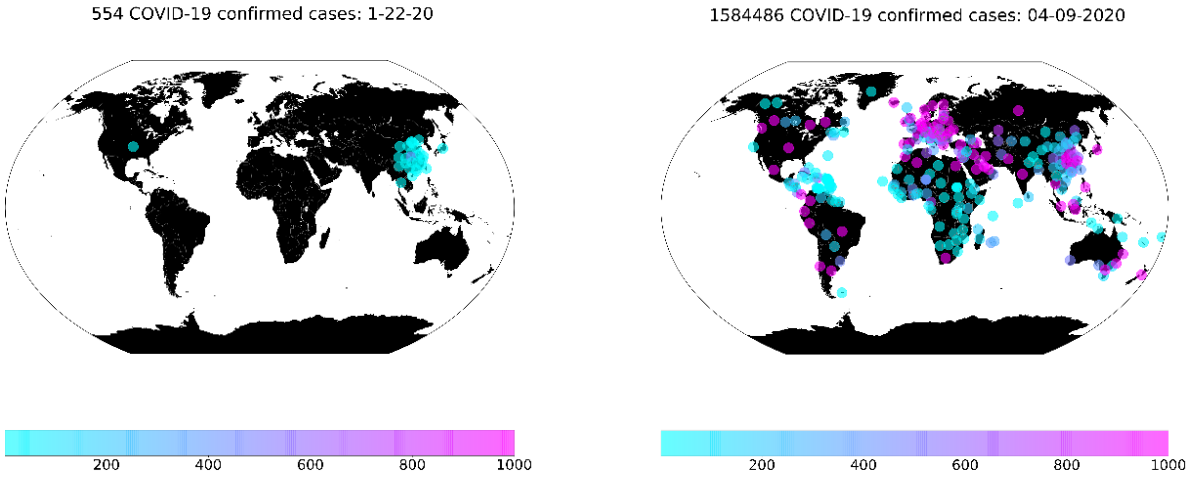

FIGURE 3
CONFIRMED ACCUMULATED CASES GLOBALLY FROM 22/01/2020 TO 09/04/2020

momentum 0.9. The model is trained for 500 training cycles (epochs).

2.4 Evaluation metrics

The performance of the proposed method is evaluated based on three different scales; 1) a global loss-based evaluation, 2) country-based evaluation and last, 3) step-based evaluation. The short-term forecast model (single-step model) relies only on the first two evaluation metrics, whereas the multi-step model includes the three levels of evaluations.

The first loss function evaluates the overall performance of the model at a global level, in which it influenced the adjustment of the model weights during training for both trained models. It is evaluated based on the Mean Squared of Error (MSE) which is calculated as:

$$MSE_{test} = \frac{1}{m}\sum_i (\hat{y}^{(test)} - y^{(test)})_i^2 \quad (7)$$

Where m is the total sample, $\hat{y}^{(test)}$ is the predicted values of the test set, and $y^{(test)}$ is the observed values of the test set.

We also computed Kullback–Leibler divergence (D_{KL}) or so called 'relative entropy 'which measures the difference between the probability distribution of two sequence. It is a common approach for assessing the VAE, nevertheless, it could be a good indicator to evaluate the predicted sequences globally. It is calculated as:

$$D_{KL}(p(x)||q(x)) = \sum_{x\in X} p(x)\ln\frac{p(x)}{q(x)} \quad (8)$$

Where $p(x)$ and $q(x)$ represent the two probability distributions of the two random discrete sequences of x. In the case of the model $p(x)$ and $q(x)$ represents the true distribution of data and the predicted one ($y^{(test)} and\ \hat{y}^{(test)}$). It is worth mentioning $(p(x)||q(x)) \neq (q(x)||p(x))$.

The second loss evaluates the performance of the model at a local level of each country or region. Strictly, $\hat{y}^{(test)}$ and $y^{(test)}$ ideally fit a statistically significant linear model where the strength of the model with r-squared value can be computed for further interpretation, in addition to the computed MSE or its root, for each county for the entire duration. Similar to the second loss, the performance of the second model (the multi-step model) includes a calculated loss (based on the root of the MSE) for each predicted step.

Last, comparing our results to other models remains a challenge due to the absence of a unified model similar to what we have achieved that forecast each country globally, or also due to the absence of general benchmark data with a common evaluation metrics. However, we try our best to compare and discuss the performance of our method to any existing models such simple or deep time-series model for specific countries or at any specific time.

3. MATERIALS AND FEATURE SELECTIONS

To forecast the spread of the Coronavirus the next day, we synchronised different types of data to allow the model to learn. This wide range of data comprises the historical data of the coronavirus spread by each country, dynamic policies and government responses that amid to mitigate Coronavirus by each timestamp and by each country, static urban features that characterise each country and shows significant correlations with the virus spread, and last, the spatial weight among the different countries. These different data types are integrated and synchronised by countries and -time steps in case of dynamic data – to be feed to the introduced framework.

3.1 COVID-19 confirmed cases data

We used the historical data for Coronavirus spread published by John Hopkins University (Dong et al., 2020; JHU CSSE, 2020). After integrating this data with following data sources, the version we used, contains timestamps from 22/01/2020 till 09/04/2020 (79 days) for 264 regions or countries across the globe as shown in Figure 3 for the confirmed cases for the start and end day of the examined duration.

3.2 Urban features data

We used demographic and locational data that represent the population for each region or country from the aforementioned data set (Worldometer, 2020). There is a wide range of factors, however, we only selected three factors; 1) Population density, 2) fertility rate and 3) Urban population. The two reasons for selecting these features are: First, the selection is based on enhancing the model prediction after several trial and errors with and without several features. Second and most importantly, the selected features show a statistically significant association with the spread of coronavirus over time for all countries across the globe. Figure 5 shows the outputs of the spearman correlation for the three selected factors. In figure 5-a, the population density was significant with decaying positive correlation coefficients (rho) from the starting date until before the last 14 days of the examined duration. This means the higher the population density, the more likely a higher coronavirus spread. In figure 5-b, the fertility rates across the globe show a significant association over the entire tested duration, with negative rho values, which means countries with higher fertility rates are less likely to have a higher spread of coronavirus. This could explain the less spread of the virus in Africa (as shown in figure 3), however, this feature may be a time-dependant or due to reporting inaccuracy or the low percentage of virus testing in Africa. Last, in figure 5-c, the percentage of the urban population started to show a significant association with the spread of the virus with positive rho values from the middle of the tested duration till the last day. This means the higher the countries with a higher percentage of the urban population, are more likely to have higher coronavirus spread.

3.3 Government Response Stringency Index

Different countries took and continuously take different measures and responses amid towards coronavirus outbreak. These time and location dependant measures include 13 indicators, which they are: 1) school closing, 2) workplace closing, 3) cancelling public events, 4) close public transport, 5) public information campaigns, 6) restrictions on internal movements, 7) international travel controls, 8) fiscal measures, 9) monetary measures, 10) emergency investment in health care, 11) investment in vaccines, 12) virus testing framework, and 13) contact tracing. Put all together, Oxford COVID-19 Government Response Tracker (Hale et al., 2020) aimed to measure the variation of the government responses weighted by these indicators in a scaled index, so-called Stringency Index. We used this index to weight the different countries based on the government responses, after integrating and matching the time and location of the previously mentioned data sets.

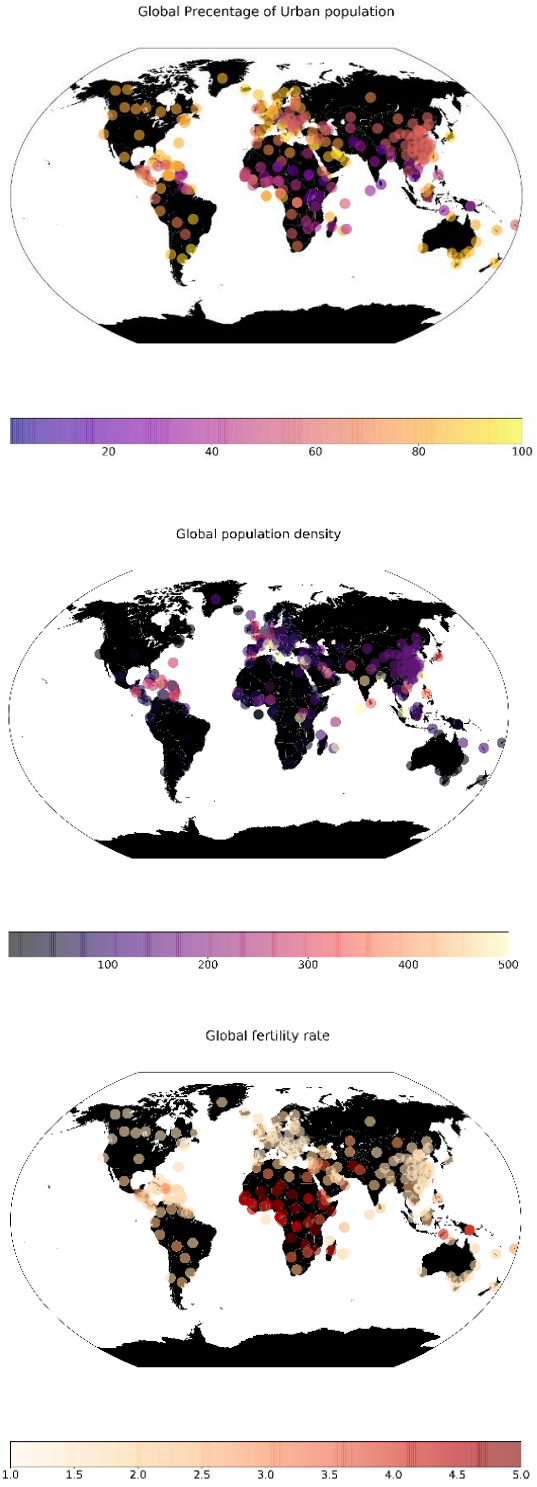

FIGURE 4
THE SELECTED THREE FACTORS FROM TOP TO BOTTOM
(URBAN POPULATION, POPULATION DENSITY, AND FERTILITY RATE)

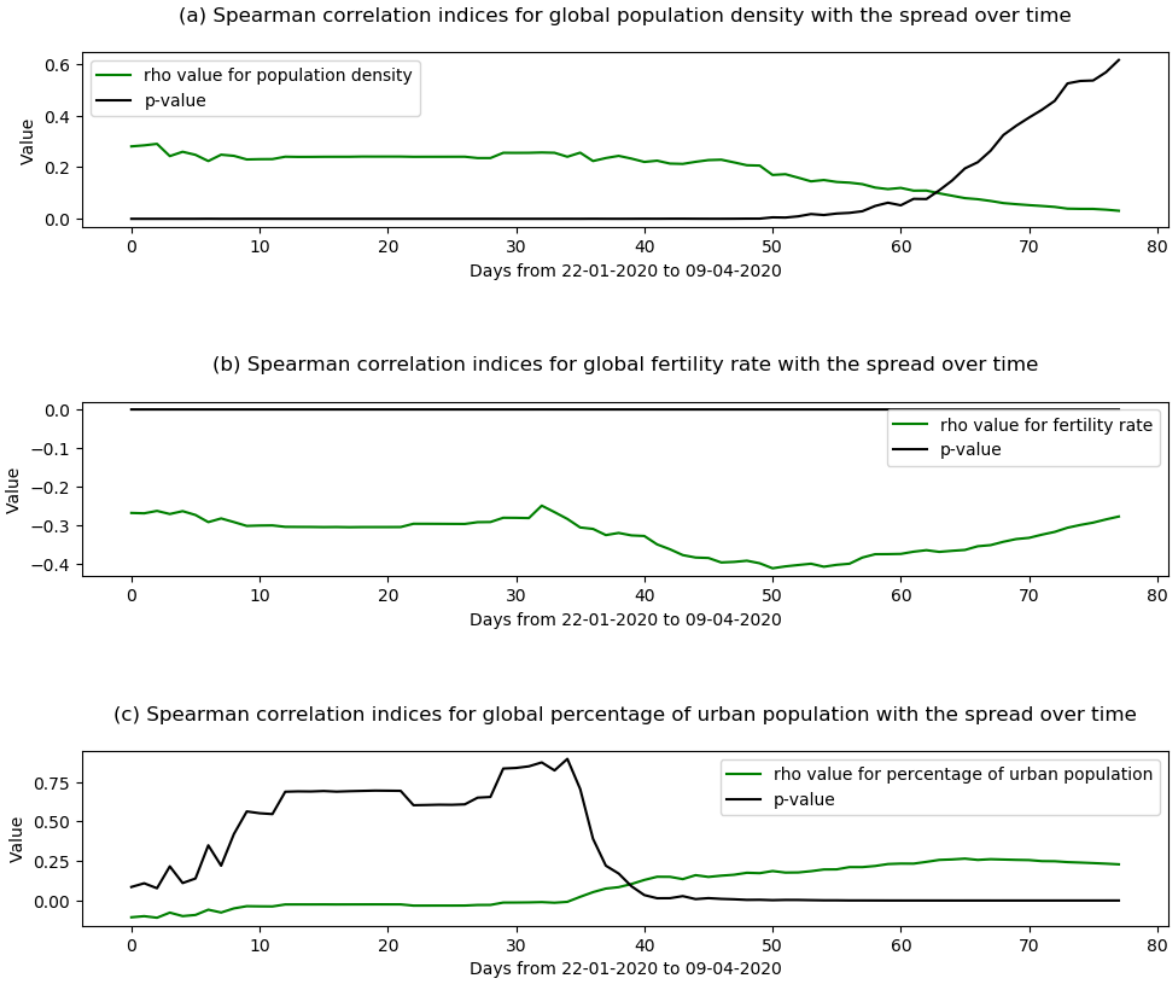

FIGURE 5
SPEARMAN CORRELATION INDICES FOR THE SELECT URBAN FEATURES WITH THE CORONAVIRUS SPREAD
(THE PERIOD BETWEEN 22/01/2020 TO 09/04/2020)

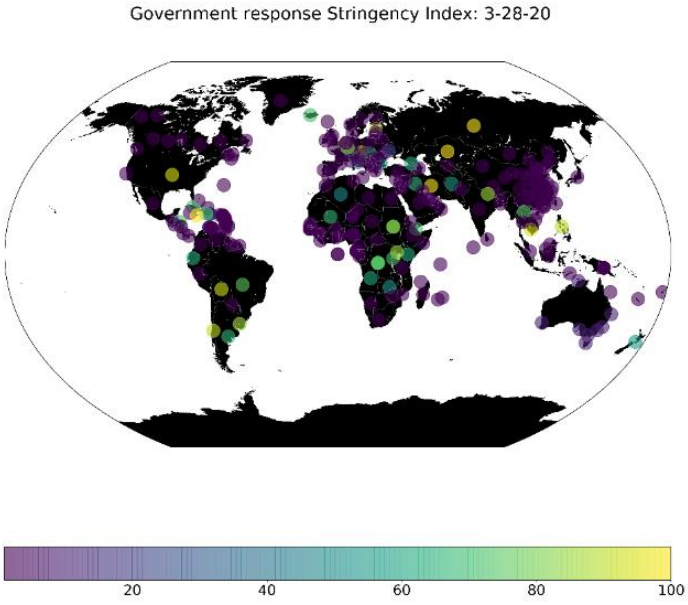

FIGURE 6
AN EXAMPLE FOR THE STRINGENCY INDEX GLOBALLY FOR 28/03/2020

3.4 Spatial weight

We computed a spatially weighted adjacency matrix based on the geolocation of each region or country, relying on the geodesic distance between each region or country. We used haversine formula to compute the distance on the sphere. It calculated as:

$$a = sin^2\left(\frac{\Delta\varphi}{2}\right) + cos\varphi_1 cos\varphi_2 sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (9)$$

$$\mathrm{d} = \mathrm{R}\left(2 \cdot \mathrm{atan2}(\sqrt{\mathrm{a}}, \sqrt{(1-\mathrm{a})})\right) \quad (10)$$

Where φ_1 , φ_2 represent the origin and destination latitudes in radian respectively, $\Delta\lambda$ represents the change between the origin and destination longitudes in radian, and R is the earth's radius.

The adjacency matrix is conditioned based primary on eliminating long-distance connections, which can represent the connection between the US and Europe, the US and China, and direct connection between China and the rest of the world. This hypothetical assumption came from the early international measured took by the US to ban flight to Europe and China for Non-American citizens. Given, this spatial weight may vary or have a higher degree of uncertainty, the model only self- learns from its representation while it generates various samples with the VAE encoder as discussed earlier, instead of using these data as a fixed and constant factor during training and testing. to be in business-as-usual. However, these are only few easily interpretable examples, the challenges for the model is to self-learn the representation of the graph to adjust the different weights and generate graph that could in forecasting the spread globally.

In Algorithm 1, we show how we initialised the adjusted spatially weighted matrix for all countries. It attempts to show three main elements for computing the graph: first, it shows how a complete graph between the origin and destination countries is computed. Second, how the relative distance is computed and conditioned. And last, it shows how the array is scaled and reshaped.

Figure 7 shows examples for the variation that could be more significant and realistic for predicting a given day for a given country. For instance, the first graph in figure 7, can represents countries with strict measures towards international travel, the second one which could be the more likely to be the case during the period of banning travel from US to Europe or China for instance, the last two shows how the world more likely to be in business-as-usual.

ALGORITHM 1: INTIALIZING THE ADJUSTED SPATIALLY-WEIGHTED ADJACENCY MATRIX

INPUT number of countries (n)
INPUT weighted_distance
INPUT latitude
INPUT longitude
Radius = 6372.800
FOR x = 0 to n -1
  i = 0
  $\boldsymbol{\varphi_1}$= radian (latitude[x])
  $\boldsymbol{\lambda_1}$= radian (longitude[x])
  WHILE i < latitude
    $\boldsymbol{\varphi_2}$ = radian (latitude [i]))
    $\boldsymbol{\lambda_2}$ = radian (longitude [i])
    i ← i+1
  END WHILE
  $a = sin^2(\Delta\varphi/2)cos\varphi_2 * sin^2(\Delta\lambda/2)$
  d = Radius (2 * atan2 (√a, √(1−a)))
  IF d < weighted_distance THEN
    OUTPUT d
  END IF
NEXT x
END FOR
computed_distance = d → scaled [0:1]
spatial_weight = computed_distance → reshaped (n , n)
OUTPUT spatial_weight

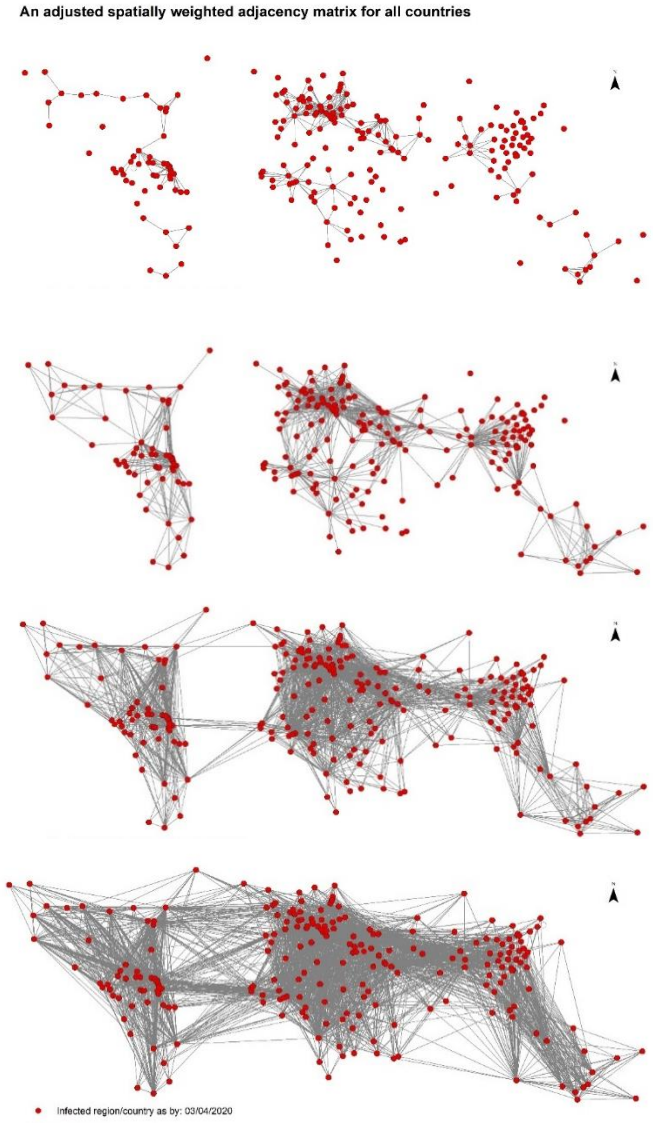

FIGURE 7
FEW EXAMPLES OF DIFFERENT ADJUSTED SPATIALLY WEIGHTED ADJACENCY MATRIX, CONDITIONED BY LIMITING DIRECT CONNECTION THAT WOULD BE GENERATED BY VAE AFTER INITIALISATION

4. RESULTS

4.1 Model evaluation globally

After 500 epochs, the training and testing curves of the model show a steady output with no sign of over fitness, nevertheless, the MSE losses for both curves are at a minimum, with values less than 0.01 whereas the KL loss for the test set is less than 0.37 for both trained model. In figure 8, we show the distribution of the confirmed and predicted cases globally with the single step model. The total predicted cases per day is a close number to the actual data, with a slightly higher confirmed in Africa than what has been confirmed.

In Figure 9, we show the sum of the accumulated predicted cases – predicted at a country level - across the globe for each day regarding the actual data. the results are highly accurate at a global level, with a fraction difference between the actual and predicted ones on the last examined day 09/08/2019.

The prediction of the model is nonlinear, however, its output at a given sample when it compared to its ground truth is linear. Therefore, fitting a linear regression model between the predicted result and the observed one and provide a r-squared

FIGURE 8
ACCUMULATED CONFIRMED CASES (LEFT-HAND SIDE) AND PREDICTED ONES GLOBALLY (RIGHT-HAND SIDE) FOR THE LAST THREE DAYS OF THE EXAMINED DATA (THE PERIOD BETWEEN 07/04/2020 TO 09/04/2020).

value could be a good indicator for understanding the model strength. Therefore here, we also show the r-squared value and the root of the MSE metrics (RMSE) for a linear regression fitted model on the predicted and actual values of our single-step model. The computed metrics shows a high linear association among them.

What makes this method a more reliable one than any simple time-series model is that the predicted global curve to the actual one is outputted without the model learns the actual one or without any explicit rules extracted at the global level to mimic that global spread curve of the virus. The model learns the patterns at country levels, whereas error is minimised at both local and global levels. What makes this a very crucial point to discuss is that changes across the globe more likely to happen at a country level, whereas the global level is rather an impact of the difference countries.

4.2 Evaluation of selected countries

Not only the model shows higher performance globally but also at a country level. Figure 10 (In Annexes section) shows the performance of the a single-step model at different countries. Overall, the model shows higher performance in countries with higher spread whereas the performance of the model decreases with countries with fewer cases over short period. However, the model shows overall reliable results at a country level.

4.3 Evaluation of single step and multi-step models

In table 1, we extend on the evaluation of the single step model. We show further variation of prediction in selected countries in different continents. While the model performance varies from a country to country, overall, it shows a reliable result for at a country level.

In table two (In Annexes section), we show the performance of the 10-step model for a group of selected countries. This model is evaluated per country and per step. While the model performance reduces with the increase of the number of steps, compared to the single step model, the result to a higher degree remains consistent at a country level when we reach the 10-step.

TABLE 1: PREDICTION EVALUATION OF SELECTED COUNTRIES WITH ONE-STEP MODEL

COUNTRY/REGION	RMSE	R_SQUARED
UNITED STATES	12722.61605	0.999624301
SPAIN	323.2040961	0.995090427
ITALY	4459.537367	0.997685728
GERMANY	645.0758347	0.997001086
IRELAND	211.2616749	0.992944828
FRANCE	7624.255423	0.984273624
UNITED KINGDOM	2712.825513	0.997394929
IRAN	2097.193217	0.98846854
RUSSIA	604.2631428	0.982704301
ROMANIA	406.3193994	0.996453797
INDIA	271.3863388	0.971975512
EGYPT	176.0606701	0.851700347
SAUDI ARABIA	454.664592	0.997453749
JAPAN	430.3462751	0.86985139

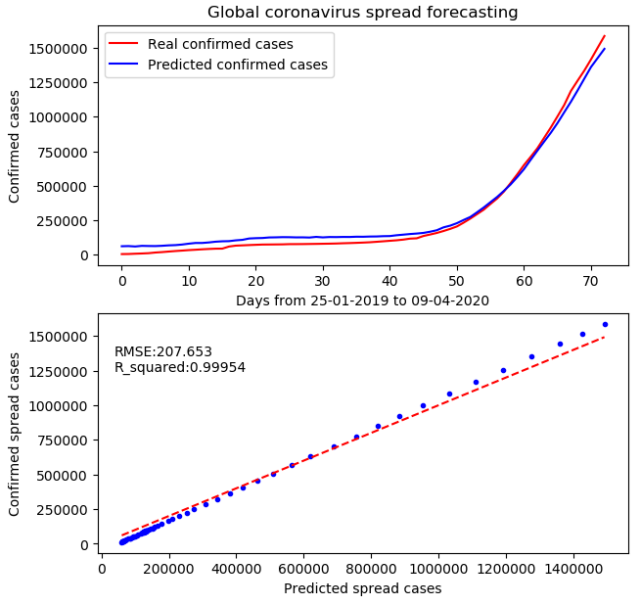

FIGURE 9
THE TOTAL CONFIRMED CASES GLOBALLY AND SUM OF THE PREDICTED CASES AT A COUNTRY LEVEL

5. DISCUSSION

In this article, we introduce a state-of-the-art method for predicting the spread of coronavirus for each country across the globe for both short and long-term forecast. It has three main advantages, first, the model learns not only from the historical data but also the applied governmental measures for each country, urban factors, and the spatial graph that represent the dependencies among the different countries. The second advantage of the model is its ability to be applied at various scales. Currently, it can forecast the spread at a global and country, and region level (i.e. the case of China, UK), however, it can also be applied at the city level. Last, the model can forecast short and long term forecast which could be a reliable tool for decision-making.

5.1 Base model evaluations

There are different attempts for relying on a simple time-series model whether it is relying on machine learning or a simple mathematical rule for a single country or the total cases globally. However, the drawback in such methods is: First, by fitting an exponential smoothing function to a model with no controlled point would mean the virus will continue to spread, regardless of the number of a population, the action is taken. Second, if a simple rule for a given country works for the last days, till when this logic will continue works? What happens when values remain constant, decrease, or even increase at a different rate? There are different possible scenarios that such an approach could not answer. Third, despite the first two arguments, how many rules are needed to fit each country globally at a longer period? Subjectively, a simple time-series model without considering the factors that characterise countries or policies taken to find “general rules and features” would mean finding simple rules for each country at a given time. In most simple ways, when the curve is only increasing at the initial spread time.

Last, even if these previous issues are solved, the world is connected, the spatial weights may vary from country to country or day to day based on the restrictions and measures are taken. If there are simple rules that ultimately can fit the entire countries, the challenges would remain in how to weight the changes around the world. Most importantly, one single case in one region could influence the spread elsewhere.

5.2 Limitation and future work

The generative graph of the model along with other factors of the model empirically has enhanced the predicted values for each country globally (based on trial and errors). However, it remains a challenge that countries with spread over a longer period are more likely to be predicted more accurate than countries with no prior cases. Re-training the model with more data in the future would yield better results at both; global and country levels. However, despite data improvement, there are three main domains that the model algorithms can be advanced in future works. First, finding more significant spatial or demographic factors that show significant associations with the spread may also enhance the forecast of the model Second, applying the same concept and goals of the model to other subjects of coronavirus, could lead to a better understanding of its future. For instance, estimating deaths or recovery, bearing in mind the health system capability and capacity, in addition to the governmental responses could be another assisting tool. Put all together, more data, more factors, different forecasting model could also lead to better long-term forecast (1-3 months) for each country based on the lesson learned from the global and country-level trends of spread.

6. REMARKS AND LESSONS LEARNED

In this article, we introduced a new deep learning model relying on variational-LSTM autoencoder to predict the spread of coronavirus for 264 regions/countries across the globe. The introduced learning process and the structure of the data are keys. The model learned from various types of dynamic and static data, including the historical spread data for each country, urban and demographic features such as urban population, population density, and fertility rate, and government responses for each country amid towards mitigating coronavirus outbreak. Also, the model learned to sample different conditions and adjustments of a spatially weighted adjacency matrix among the different infected countries. Overall, the model shows high validation for forecasting the spread at global and country levels, which makes it a useful tool to assist decision and policymaking for the different corners of the globe.

There are several lessons learned while conducting this research. First, concerning urban features, we found several associations of several factors with the spread of coronavirus globally. Most significantly, countries with a higher density of population in one km^2 and larger portion of the population living in urban areas are associated with higher coronavirus spread with different coefficients, and levels of statistical significance during the examined duration, whereas countries with higher fertility rates are associated with fewer spread cases at the given studied period (22/01/2020-08/04/2020). However, we also found an association with other factors that not used in this research such as migration nets. We found that countries with higher migration flows are associated with higher spread which could also be explained with their likelihood of having a higher influx of job opportunities. Second, concerning the computed adjacency matrix graph, we found that at very short distances among the different infected countries with coronavirus spread, Western European countries (such as Germany, Italy, Spain) are fully or partially connected relative to other countries globally that are same distance they are completely isolated. This can be reflected on the relatively shorter distance – as a physical attribute-as among these countries when it compares to other countries, or the non-physical accessibility of the European market which could lead to a higher influx of migration and accordingly higher spread cases.

7. REFERENCES

[1] Al-qaness, M.A.A., Ewees, A.A., Fan, H., Abd, M., Aziz, E., 2020. Optimization Method for Forecasting Confirmed Cases of COVID-19 in China 19. https://doi.org/10.3390/jcm9030674

[2] Botha, E., Dednam, W., 2020. A simple iterative map forecast of the COVID-19 pandemic 1–6.

[3] Castro, C.A. De, 2020. SIR MODEL FOR COVID-19 CALIBRATED WITH EXISTING DATA AND PROJECTED FOR COLOMBIA 1–9.

[4] Chang, S.L., Harding, N., Zachreson, C., Cliff, O.M., Prokopenko, M., 2020. Modelling transmission and control of the COVID-19 pandemic in Australia.

[5] Crokidakis, N., 2020. Data analysis and modeling of the evolution of COVID-19 in Brazil.

[6] Deb, S., 2020. A time series method to analyze incidence pattern and estimate reproduction number of COVID-19 1–14.

[7] Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(20)30120-1

[8] Goodfellow, I., Bengio, Y., Courville, A., 2017. Deep Learning, Adaptive computation and machine learning series. The MIT Press, Cambridge, Massachusetts.

[9] Ha, D., Schmidhuber, J., 2018. World Models. ArXiv180310122 Cs Stat. https://doi.org/10.5281/zenodo.1207631

[10] Hale, T., Webster, S., Petherick, A., Phillips, T., Kira, B., 2020. Oxford COVID-19 Government Response Tracker [WWW Document]. Oxf. Univ. URL https://www.bsg.ox.ac.uk/research/research-projects/oxford-covid-19-government-response-tracker?fbclid=IwAR3siMms7OT2N_cLZmPz2g1tGKrTUYfTXUAyQAgt_2a2kER_uKcZAfVV6Ts

[11] Hu, Z., Ph, D., Ge, Q., Li, S., Boerwincle, E., Ph, D., Jin, L., Ph, D., 2020. Forecasting and evaluating intervention of Covid-19 in the World.

[12] JHU CSSE, 2020. 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE [WWW Document]. GithubCOVID-19. URL https://github.com/CSSEGISandData/COVID-19

[13] Kermack, W., McKendrick, A., 1927. A Contribution to the Mathematical Theory o f Epidemics. Proc. R. Soc. A 115, 700–721. https://doi.org/doi:10.1098/rspa.1927.0118

[14] Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. ArXiv14126980 Cs.

[15] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

[16] Petropoulos, F., Makridakis, S., 2020. Forecasting the novel coronavirus COVID-19. PLOS ONE 15, e0231236. https://doi.org/10.1371/journal.pone.0231236

[17] Pham, K.H., Luengo-oroz, M., 2020. M APPING THE L ANDSCAPE OF A RTIFICIAL 1–14.

[18] Sameni, R., 2020. Mathematical Modeling of Epidemic Diseases ; A Case Study of the COVID-19 Coronavirus 1–11.

[19] Simha, A., Prasad, R.V., Narayana, S., 2020. A simple Stochastic SIR model for COVID-19 Infection Dynamics for Karnataka – Learning from Europe.

[20] Singer, H.M., 2020. Short-term predictions of country-specific Covid-19 infection rates based on power law scaling exponents 1–6.

[21] Toda, A.A., 2020. Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact 1–15.

[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. ArXiv170603762 Cs.

[23] Wordometer, 2020. COVID-19 Coronovirus/ Death Toll [WWW Document]. URL https://www.worldometers.info/coronavirus/coronavirus-death-toll/ (accessed 3.31.20).

[24] Worldometer, 2020. Countries in the world by population (2020) [WWW Document]. Worldometer. URL https://www.worldometers.info/world-population/population-by-country/

8. ANNEXES

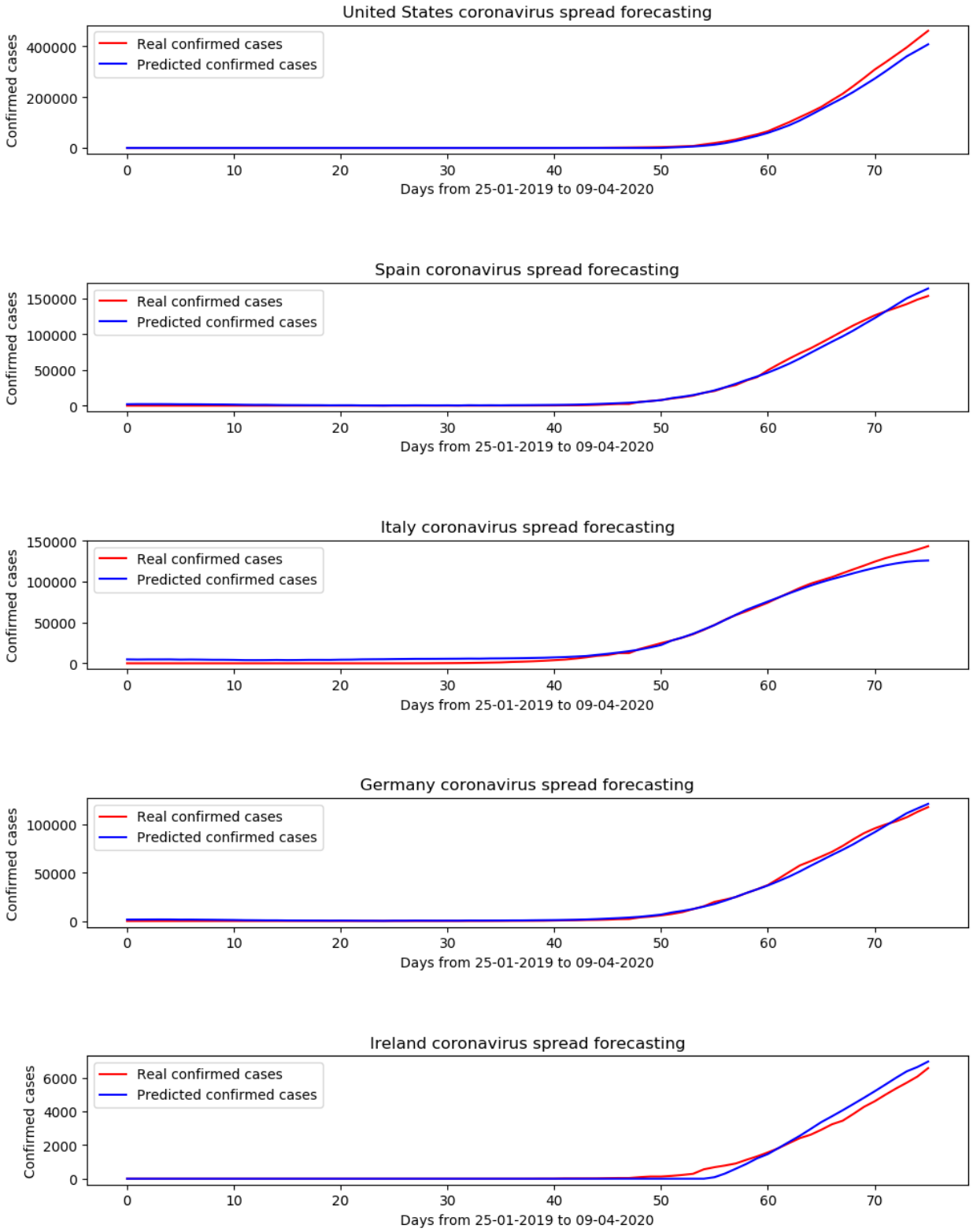

FIGURE 10
MODEL PREDICTION AND GROUND TRUTH FOR SELECTED COUNTRIES WITH THREE TIMESTAMPS AND ONE PREDICTED STEP (THE PERIOD BETWEEN 25/01/2020 TO 09/04/2020).

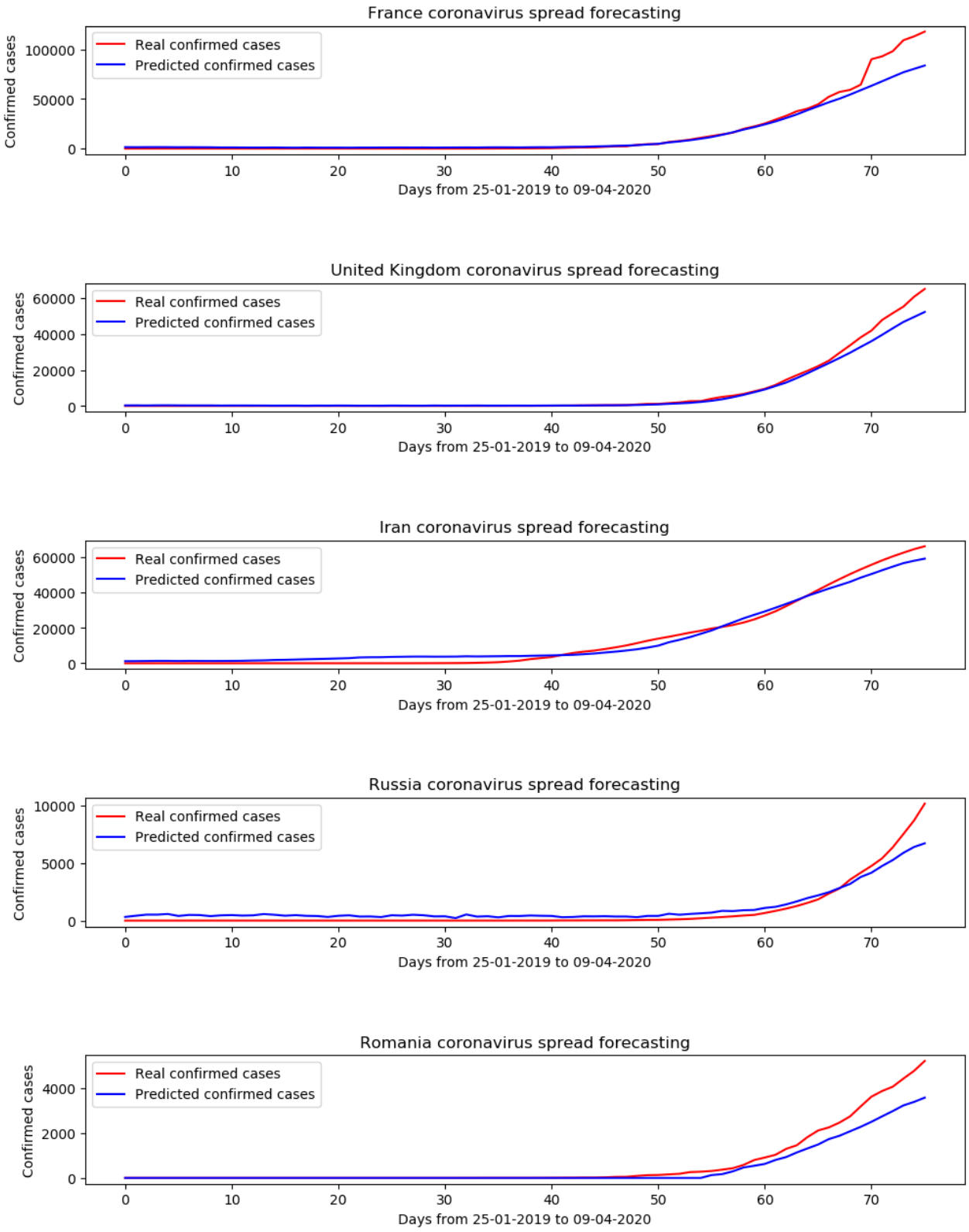

FIGURE 10 CONTINUED
PREDICTSLUMS-ANN ARCHITECTURE MODEL PREDICTION AND GROUND TRUTH FOR SELECTED COUNTRIES WITH THREE TIMESTAMPS AND ONE PREDICTED STEP (THE PERIOD BETWEEN 25/01/2020 TO 09/04/2020)

TABLE 2: PREDICTION EVALUATION OF SELECTED COUNTRIES WITH MULTI-STEP MODEL (10 STEPS)

COUNTRY/ REGION	RMSE_1 STEPS	R2_1 STEPS	RMSE_2 STEPS	R2_2 STEPS	RMSE_3 STEPS	R2_3 STEPS	RMSE_4 STEPS	R2_4 STEPS	RMSE_5 STEPS	R2_5 STEPS
UNITED STATES	14960.74	0.943205	15840.74	0.955945	16103.84	0.966965	15831.43	0.976249	15092.04	0.982463
SPAIN	4056.447	0.990265	3290.202	0.994152	1945.486	0.996188	444.5595	0.996864	1237.293	0.995529
ITALY	5086.556	0.98304	6593.107	0.975113	7932.679	0.967061	9763.755	0.956861	11319.16	0.947332
GERMANY	2111.287	0.9935	1573.153	0.995874	599.7499	0.99657	551.984	0.996137	1866.666	0.994792
IRELAND	127.0115	0.9892	155.7189	0.990045	146.4884	0.991586	174.7025	0.998624	33.05549	0.997493
FRANCE	986.1778	0.997043	1016.473	0.997723	755.85	0.998641	226.8082	0.998696	323.5484	0.997878
UNITED KINGDOM	1599.954	0.966074	1498.96	0.981528	1742.5	0.977681	1671.771	0.982697	1615.876	0.988742
IRAN	1892.061	0.962249	1741.914	0.962841	1886.464	0.962006	1945.66	0.962097	2120.902	0.960991
RUSSIA	254.8093	0.909475	31.27125	0.996975	89.20823	0.992833	144.1582	0.979667	378.2701	0.8988
ROMANIA	182.4116	0.94217	195.6457	0.963924	15.23147	0.99247	60.02222	0.987905	39.6719	0.995176
INDIA	97.88447	0.947919	62.93186	0.969662	100.5075	0.976883	137.2235	0.923134	135.3247	0.967846
EGYPT	39.99932	0.89026	60.26159	0.961026	47.51823	0.759069	51.37424	0.982259	41.00011	0.911897
SAUDI ARABIA	97.04426	0.959215	133.9108	0.98182	15.16398	0.995198	98.18284	0.98754	84.8475	0.989111
JAPAN	109.7728	0.920722	62.4994	0.955439	145.9558	0.961878	166.6307	0.939447	139.498	0.971077

TABLE 2 CONTINUED: PREDICTION EVALUATION OF SELECTED COUNTRIES WITH MULTI-STEP MODEL (10 STEPS)

COUNTRY /REGION	RMSE_6 STEPS	R2_6 STEPS	RMSE_7 STEPS	R2_7 STEPS	RMSE_8 STEPS	R2_8 STEPS	RMSE_9 STEPS	R2_9 STEPS	RMSE_10 STEPS	R2_10 STEPS
UNITED STATES	14227.58	0.988191	12565.34	0.992781	9989.597	0.99581	6497.848	0.997288	1883.936	0.997534
SPAIN	2855.312	0.992616	4937.843	0.987319	7309.107	0.979868	9745.77	0.971048	12388.68	0.96107
ITALY	12907.58	0.936813	14273.12	0.926274	15424.64	0.914829	16449.16	0.903413	17605.19	0.891036
GERMANY	3198.177	0.990744	4518.471	0.985362	5931.181	0.977806	7557.759	0.968247	9405.256	0.959184
IRELAND	66.21991	0.997409	40.00556	0.998592	99.29428	0.994143	158.2693	0.995797	254.4374	0.988625
FRANCE	402.8073	0.991105	634.4987	0.992755	455.7297	0.99392	406.9991	0.99447	346.6128	0.993881
UNITED KINGDOM	1547.32	0.990948	1414.75	0.994414	1323.453	0.996284	1020.778	0.99773	510.1477	0.998256
IRAN	2617.983	0.95746	3199.196	0.952845	3737.035	0.948024	4557.6	0.940859	5036.862	0.936583
RUSSIA	361.8903	0.949766	414.895	0.963938	564.8863	0.965795	491.668	0.970967	514.6991	0.968007
ROMANIA	96.86678	0.995751	59.62687	0.996438	61.96891	0.997107	45.04452	0.990279	53.68652	0.995802
INDIA	93.11648	0.980484	185.4696	0.971165	120.576	0.982155	356.9565	0.950267	395.6706	0.958173
EGYPT	35.27577	0.958364	39.18868	0.963363	73.50435	0.889037	35.52756	0.965417	68.11146	0.976297
SAUDI ARABIA	245.0604	0.957466	74.71041	0.991426	40.16494	0.990361	101.4406	0.986736	184.5986	0.985095
JAPAN	146.1493	0.962084	45.61221	0.969902	97.73392	0.971485	91.91085	0.972004	7.737023	0.978584