ARTICLE TITLE

Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence

AUTHORS

László Róbert Kolozsvári[1], Tamás Bérczes[2], András Hajdu[2], Rudolf Gesztelyi[3], Attila Tiba[2] , Imre Varga[2], Gergő József Szőllősi[1], Szilvia Harsányi[4], Szabolcs Garbóczy,[5] Judit Zsuga[4]

[1] Department of Family and Occupational Medicine, Faculty of Public Health, University of Debrecen, Debrecen, Hungary

[2] Faculty of Informatics, University of Debrecen, Debrecen, Hungary.

[3] Department of Pharmacology and Pharmacotherapy, Faculty of Medicine, University of Debrecen, Debrecen, Hungary

[4] Department of Health Systems Management and Quality Management in Health Care, Faculty of Public Health, University of Debrecen, Debrecen, Hungary

[5] Department of Psychiatry, Kenézy Hospital, University of Debrecen, Debrecen, Hungary

*** CORRESPONDING AUTHOR**

Dr. László Róbert Kolozsvári (LK)

E-mail: kolozsvari.laszlo@sph.unideb.hu

ORCID ID: https://orcid.org/0000-0001-5426-0858

MANUSCRIPT

Abstract

Objectives

The current form of severe acute respiratory syndrome called coronavirus disease 2019 (COVID-19) caused by a coronavirus (SARS-CoV-2) is a major global health problem. The aim of our study was to use the official epidemiological data and predict the possible outcomes of the COVID-19 pandemic using artificial intelligence (AI)-based RNNs (Recurrent Neural Networks), then compare and validate the predicted and observed data.

Materials and Methods

We used the publicly available datasets of World Health Organization and Johns Hopkins University to create the training dataset, then have used recurrent neural networks (RNNs) with gated recurring units (Long Short-Term Memory – LSTM units) to create 2 Prediction Models. Information collected in the first t time-steps were aggregated with a fully connected (dense) neural network layer and a consequent regression output layer to determine the next predicted value. We used root mean squared logarithmic errors (RMSLE) to compare the predicted and observed data, then recalculated the predictions again.

Results

The result of our study underscores that the COVID-19 pandemic is probably a propagated source epidemic, therefore repeated peaks on the epidemic curve (rise of the daily number of the newly diagnosed infections) are to be anticipated. The errors between the predicted and validated data and trends seems to be low.

Conclusions

The influence of this pandemic is great worldwide, impact our everyday lifes. Especially decision makers must be aware, that even if strict public health measures are executed and sustained, future peaks of infections are possible. The AI-based predictions might be useful tools for predictions and the models can be recalculated according to the new observed data, to get more precise forecast of the pandemic.

Introduction

Coronavirus

High and low pathogenic species may be distinguished within the coronavirus family, with the former including 4 viruses that are responsible for 10-30% of mild upper respiratory diseases (e.g. common cold), and the latter known to cause a more severe form of acute lung injury: SARS (Severe Acute Respiratory Syndrome) and MERS (Middle East Respiratory Syndrome) CoV (coronavirus).[1]

SARS-CoV originated in Guangdong Province, China and started to spread in 2002, causing over 8,000 illnesses in 29 different countries all over the world, with a crude fatality rate of 10%.[2,3,4] The disease spread to Hong Kong in 2003 causing an outbreak of severe acute respiratory syndrome (SARS). A novel coronavirus was isolated and was suggested to be the primary cause of the infections.[5] Few years later, in 2007 Cheng et. al issued a warning that "the presence of a large reservoir of SARS-CoV-like viruses in horseshoe bats, together with the culture of eating exotic mammals in southern China, is a time bomb".[4]

MERS-CoV began spreading in Saudi Arabia in 2012 and to date has led to a total of 2519 laboratory-confirmed cases in several countries around the world.[6,7] Its case-fatality rate reached 37.1% over the course of the past 8 years.[7]

COVID-19

The current form of severe acute respiratory syndrome called COVID-19, is caused by a new variant of formerly known highly pathogenic *Coronaviridae*. The infection allegedly began to spread from a market in Wuhan, the capital of Hubei province, China, at the end of 2019. Early PCR analysis has found that the new virus, called 2019-nCoV by the World Health

Organization (WHO) and SARS-CoV-2 by the International Committee on Taxonomy of Viruses, shows a 79.6% homology with SARS-CoV, and has 96% sequence identity with bat coronavirus suggesting a common origin from SARSr-CoV (severe acute respiratory syndrome related coronavirus). According to analyzes the suspected host is a bat species, Rhinolophus affinis (a horseshoe bat), but the virus probably needs an intermediate host.[8]

Symptoms associated with the disease include fever (83%), cough (82%), shortness of breath (31%), muscle aches (11%), confusion (9%), headache (8%), sore throat (5%), runny nose, chest pain, diarrhea, nausea and vomiting.[9] According to a meta-analysis that complied data from more than 50 000 patients, the incidence of fever (0.891, 95% confidence interval (CI): [0.818; 0.945]) and cough (incidence of 0.722, 95% CI: [0.657; 0.782]) were the highest respectively, followed by muscle soreness and fatigue.[10]

The incubation period of the COVID-19 disease is estimated between 1-14 days (5 days on average).[11] There is no definite data concerning the transmissibility of the virus. Several transmission routes have been identified: direct lung, other mucous membranes, direct bloodstream and possibly fecal-oral transmission.[12] It seems probable that those with the fulminant disease are most infectious, but reports have identified asymptomatic and presymptomatic virus shedding as well. There is also lack of definite data regarding tertiary and quaternary spreading among humans, but it seems probable that the person who has been exposed to the infection has acquired some (at least temporary) immunity. [13]

According to WHO data, there were 1 914 916 confirmed cases and 123 010 fatalities globally as of 15th of April 2020, which corresponds to a case-fatality rate of about 6.42 %.[14]

R0, the basic reproduction number, denoting the transmissibility of a virus indicates the average number of new infections induced by an infectious person in a susceptible, infection naïve population. The transmissibility of the virus was apparently underestimated initially by the

WHO with R0 suggested to range between 1.4 and 2.5. More recent analyzes have indicated higher R0 values around 3 (with the mean and median R0 for published estimates being 3.28 and 2.79, respectively).[11,15]

The daily number of the newly diagnosed infections - epidemic curves

The initial epidemic curves of the COVID-19 outbreak from Hubei, China showed a mixed pattern, indicating that early cases were likely from a continuous common source e.g. from several zoonotic events in Wuhan, followed by secondary and tertiary transmission providing a propagated source for the later cases.[16]

The propagated (or progressive source) epidemic curve visualizes the spread of an infectious agent that may be transmitted from human to human starting from with a single index case, that continues to infect numerous other individuals. This shows up as a series of peaks on the epidemic curve, that starts with the index case, followed by successive waves of the infection set apart with respect to the incubation period of the pathogen. The waves continue to follow each other, until appropriate mitigation measures, prevention or treatment are implemented, or the pool of the susceptible population becomes infected. This is a theoretic curve, that is generally influenced by lots of other factors.[16]

Several studies investigated the impact of different interventions with respect to minimizing contact rates in the population to slow the infection spread, minimize COVID-19 mortality rates and health care utilization or to suppress the epidemic per se. Flattening the curve by reducing peak incidence may limit overall case fatality rates. Nevertheless, most of the forecasts and simulations thus far started out from Bell-shaped curves, that fail to account for the progressive nature of the current outbreak given the known secondary, tertiary even quaternary transmissibility of the virus. Taking this into account it is suggested that the number of cases will rise once again, after pandemic control measures are no longer in effect.[17]

133 Prediction

134 There are different mathematical models that may demonstrate and predict the dynamics of the
135 different infectious diseases.[18] These models, used to simulate the dynamics of infectious
136 diseases, may be based on statistical, mathematical, empirical or machine learning methods.[19]

137 The first attempts to use Artificial Intelligence (AI) in medicine were made in the 1970s.
138 Initially AI was used to implement programs to help clinical decision making, but to date its
139 use is gaining more and more widespread acceptance in biomedical sciences.[20]

140 One class of AI, a form of artificial neural networks, the Recurrent Neural Networks (RNNs)
141 with Long short-term memory (LSTM) were previously used to model and forecast the
142 influenza epidemic, with strong competitiveness and reliable results.[21, 22, 23]

143 The aim of the current study was to use the available official data as a training dataset, followed
144 by predicting the possible outcomes of the COVID-19 pandemic using AI-based RNNs, then
145 compare the predictions with the observed data.

146

147 # Materials and Methods

148 Data

149

150 We used the publicly available datasets of WHO and Johns Hopkins University from the
151 following countries to create the training dataset: Austria, Belgium, Hubei (China), Czechia,
152 France, Germany, Hungary, Iran, Italy, Netherlands, Norway, Portugal, Slovenia, Spain,
153 Switzerland, United Kingdom, United States of America.[13,24] Given that most infected people
154 in China were from Hubei province, only data from that province was included. For each
155 country, the date of the first infection was set as day 1 for the disease time scale. (Fig 1)

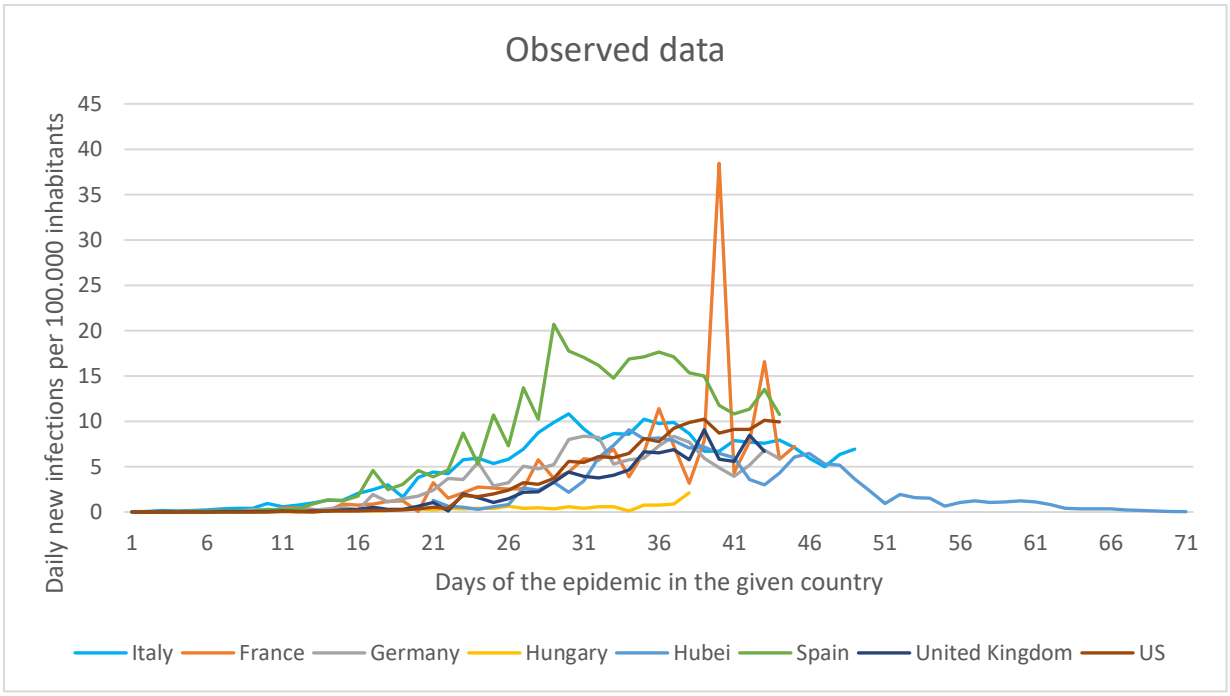

Fig 1. The historical datasets from different countries

When determining the date of the first illness, point source outbreaks were omitted (e.g. those cases where single verified cases were isolated, and no further transmission has occurred). This was important to avoid distortion of the propagated epidemic curves. In Belgium, for example, the first illness occurred on 04/02/2020 and there was no further case reported for up to 26 days. The next illness occurred on 01/03/2020. Inclusion of the early case from February would contribute to a false learning rule for the AI, hence corrupting the results. As for Hubei Province, the first officially available data is of 22/01/2020. This cannot be considered as the first day of the illness, thus the first infection was arbitrarily defined to occur on 01/01/2020. To account for the extreme variability of daily incident cases probably reflecting delays in reporting, a moving average was used (covering 3 days) for Hubei dataset.

Accordingly, an epidemic curve was obtained for each country with a time series where the first day denotes the day of the first confirmed case, and each successive day indicating the number of newly confirmed cases that day. To account for the country-specific differences in the size of population, the number of daily new cases was normalized for 100 000 inhabitants in each

country. The observation period varies for each country, given the difference of time elapsed since the disease initiation in that country. Accordingly, the longest time series covers the observation period of 90 days. e.g. in Hubei, with the first 22 days lacking valid data and the next 68 days having data. The shortest observation period was in Slovenia with only 30 days.

The training data set was obtained by averaging the daily incidence rates per 100 000 inhabitants across the 17 countries included, for each day in the time series. When calculating the average, missing data was left blank, i.e. NULL, e.g. countries that did not contain a data for a specific day, were excluded from calculation of average. The resulting training data set is shown in Figure 1. It should be noted that the first part of the data set (up to the initial 30 days since Day 1 of the epidemic) contains data for almost all the countries listed, whereas the end of the data set contains only data from Hubei. (Fig 2)

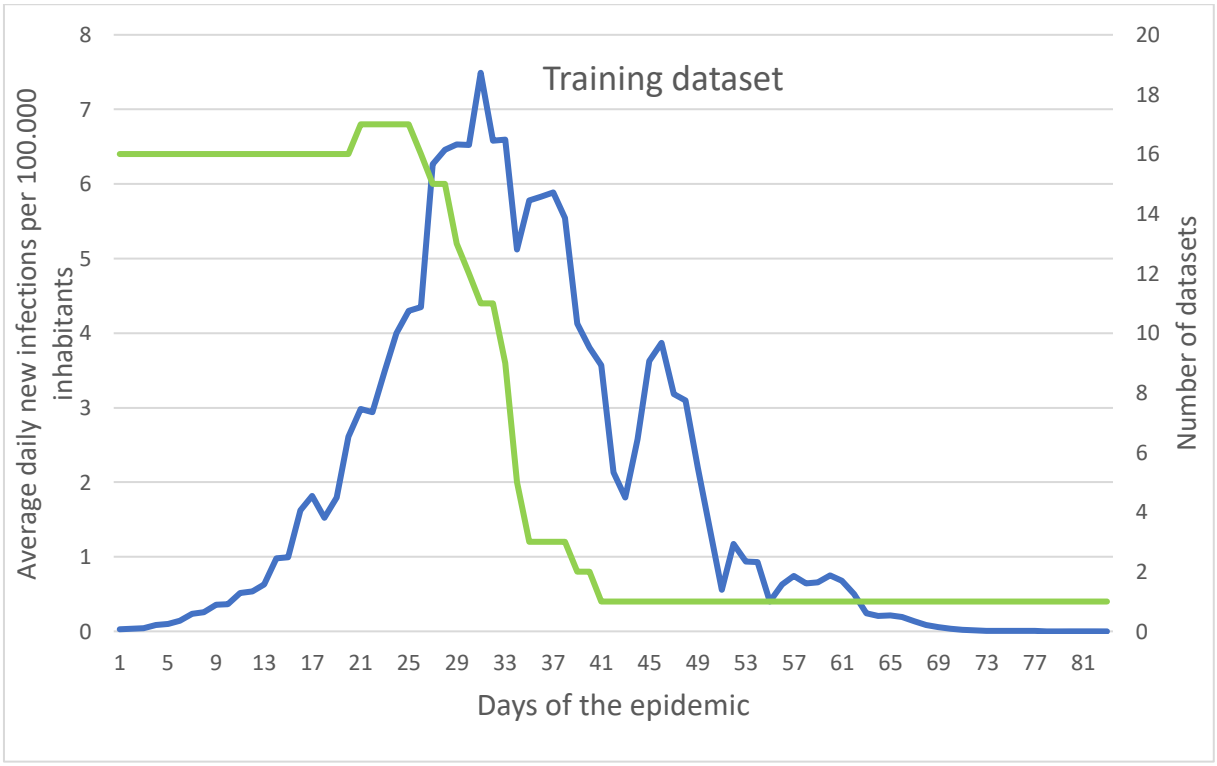

Fig 2. The training dataset. Average daily new infections per 100.000 inhabitants (blue coloured line) and the Number of datasets (green colour line)

RNN-based model for prediction

The state-of-the-art for time series analysis is artificial intelligence-based analytic tools, which have the best prediction performance. Recurrent Neural Networks (RNNs) are specifically designed to cope with sequential input, characteristic of textual or temporal data.[22] This architecture is a neural network-based architecture, that contains hidden layers chained according to the time step, with a possibility to predict the next sequence element(s). A time series has a special temporal form, where the input to the i-th hidden layer is at the i-th time-step that has a corresponding x(i) observation. In its original form a simple RNN tries to predict the next sequence element, however, for the purposes of the current analysis, an encoder-decoder variant is a more natural choice, similarly to machine translation.[25] For our specific scenario this means that the during the encoder phase including time steps 1,…,t the RNN is fed with the already known time series data (the average of the number of new cases normalized to 100 000 inhabitants for day 1…t, respectively), followed by prediction in the decoder phase for the future time steps t+1,…,T. In our analysis T=t+1=90 days is the longest known (Hubei) time interval. Since this covers quite a long data sequence, we have used gated recurring units (namely Long Short-Term Memory – LSTM units) in compliance with the general recommendations.[23] Figure 3 depicts our RNN architecture showing how unknown time series elements are predicted. Figure 3 also shows how the information collected in the first t time-steps are aggregated with a fully connected (dense) neural network layer and a consequent regression output layer to determine a predicted number of new patients as x(t+1). (Fig 3)

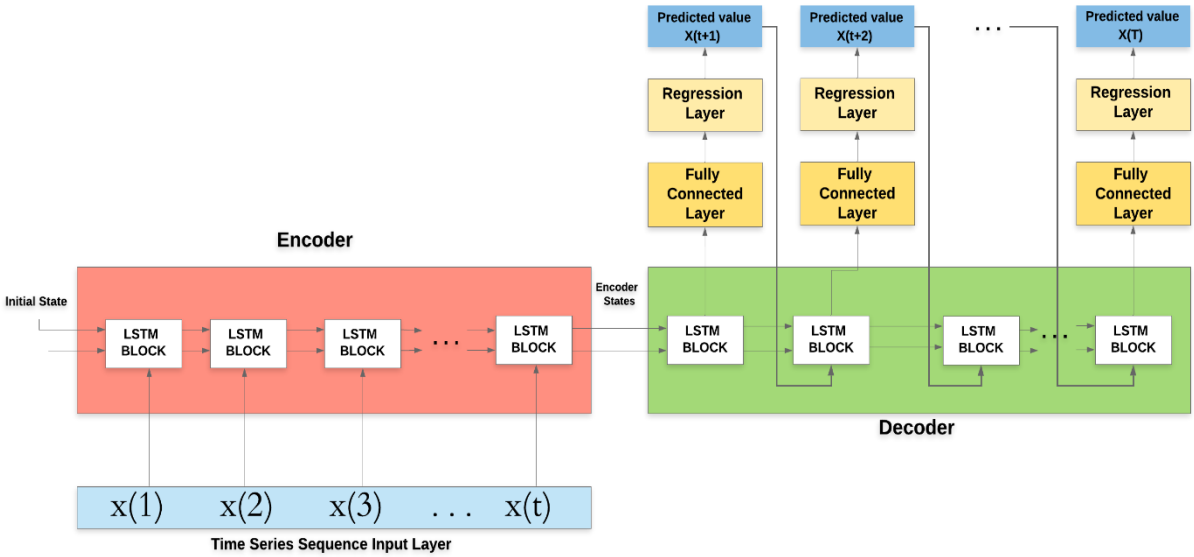

Fig 3. The Recurrent Neural Network architecture used for prediction.

The training data was described in previous sections. To assess possible specificities regarding the countries two approaches were used for prediction:

- Prediction 1: An algorithm to update training step and subsequent prediction was formulated. This update step is based on the general recommendations of transfer learning that considers the already known time interval for the given country and re-training is done in small increments of the RNN network accordingly.[26] Thus we start predicting the first unknown element x(t+1) from the last 5% of the known data, and the same principle is applied to each subsequent element. Moreover, after each prediction step our RNN architecture is re-trained and the subsequent elements are predicted with this updated RNN.
- Prediction 2: We start predicting the first unknown element x(t+1) from the last known x(t), and all the subsequent elements are predicted only from the preceding ones. Here the rules depicted from the training data set are used, not retraining occurs.

The intuitive interpretations of the difference between Prediction 1 and Prediction 2 are as follows. Prediction 2 makes its predictions utilizes the information derived from the training

data set, reflective of the trends in the average time series. It follows that predictions will comply primarily with the Hubei time series, especially in the far future. Therefore Prediction 2 shows highest fidelity to the country-specific future scenario if the approach to mitigate the epidemic is similar to that in Hubei. Accordingly, this scenario is also reflective of a country-specific future state given the practices of Hubei were followed in said country. On the other hand, Prediction 1 is yielded after the neural network is retrained after any prediction, providing more valid insight into what is expected if the country goes on with the mitigation practices seen during the observation period.

The architecture was trained in 250 epochs with a total number of 100 hidden LSTM layers, to prepare a bit for prediction also after T=90 days. Naturally, the length of the RNN can be freely increased later on.

Validation

To validate the predictions, we first made the above mentioned two predictions based on data available up to 30/03/2020. The resulting daily new morbidity data are labeled “Old Prediction 1” and “Old Prediction 2” on each graph. We then expanded our factual data set with new daily data available until 10/04/2020. These new factual data are labeled “Observed next days” on the graphs. Thus, except for Hungary, we have 11 new daily factual data elements for all countries examined. In the case of Hungary, the data of 10/04/2020 were already available, so in this case 12 new factual data elements are included. Using these data, we validated the two predictions of our model.

The amount of root mean squared logarithmic errors (RMSLE) was used for validation.

In our analysis the possible bias regarding the difference ratios between the observed and predicted values are interpreted using root mean squared logarithmic errors (RMSLE).

Let n be the number of days you for validation. Let p_{1i} and p_{2i} be the number of new cases per day obtained using the two prediction methods in the examined time interval and let a_i be the actual data for the given days. *Err1* and *Err2* be mean squared logarithmic errors (RMSLE) for Prediction 1 and Prediction 2, respectively, where:

$$Err_1 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_{1i}+1) - \log(a_i+1))^2}$$

$$Err_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_{2i}+1) - \log(a_i+1))^2}$$

For each graph, the small graph in the upper right corner contains the daily error values calculated for the predictions. The more accurate the prediction, the smaller the RMSLE error. It should be noted that if the error function is parallel to the x-axis, it means that the trend of the prediction is the same as the real trend, only at a lower or higher scale.

As the next step, using the next 11 new observation data elements after the first prediction and 12 in the case of Hungary, we modified the predictions using both methods. These modified prediction data are labeled New Prediction 1 and New Prediction 2, respectively.

Results

The following section shows the outcomes for Prediction 1 and Prediction 2 for the individual country level data (Figs 4-10).

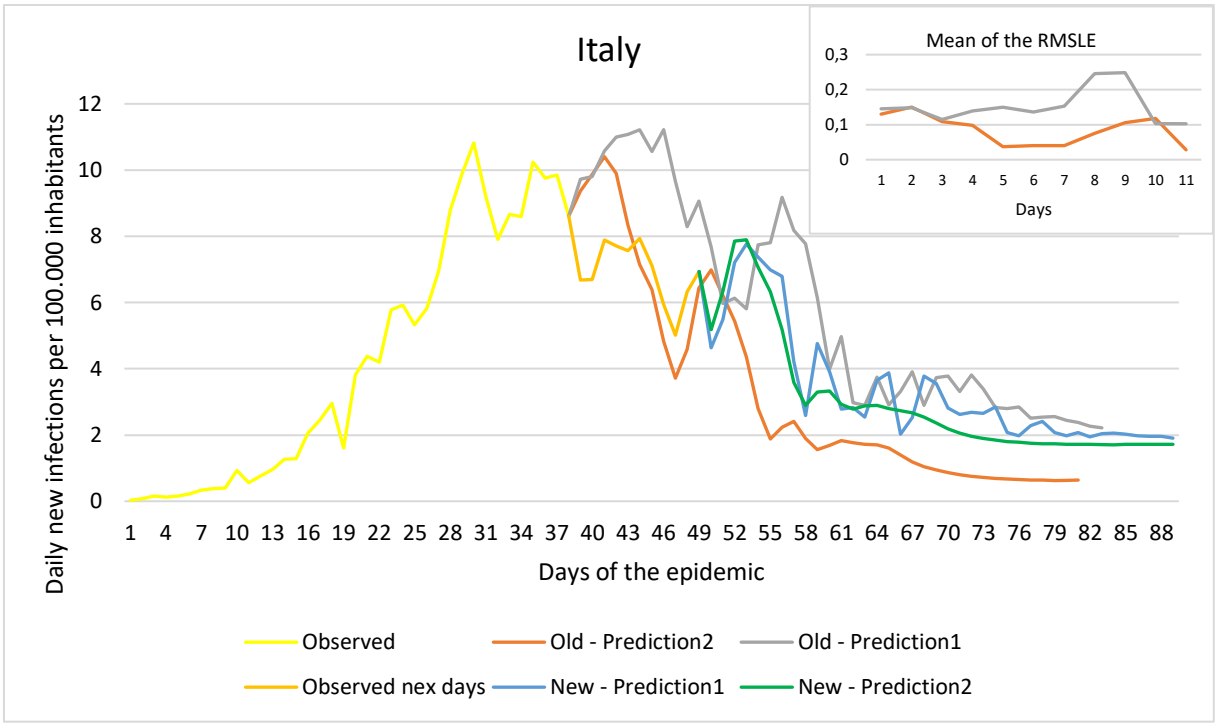

Fig 4. Observation and predictions for Italy. The small graph in the upper right corner shows the daily error values calculated for the predictions.

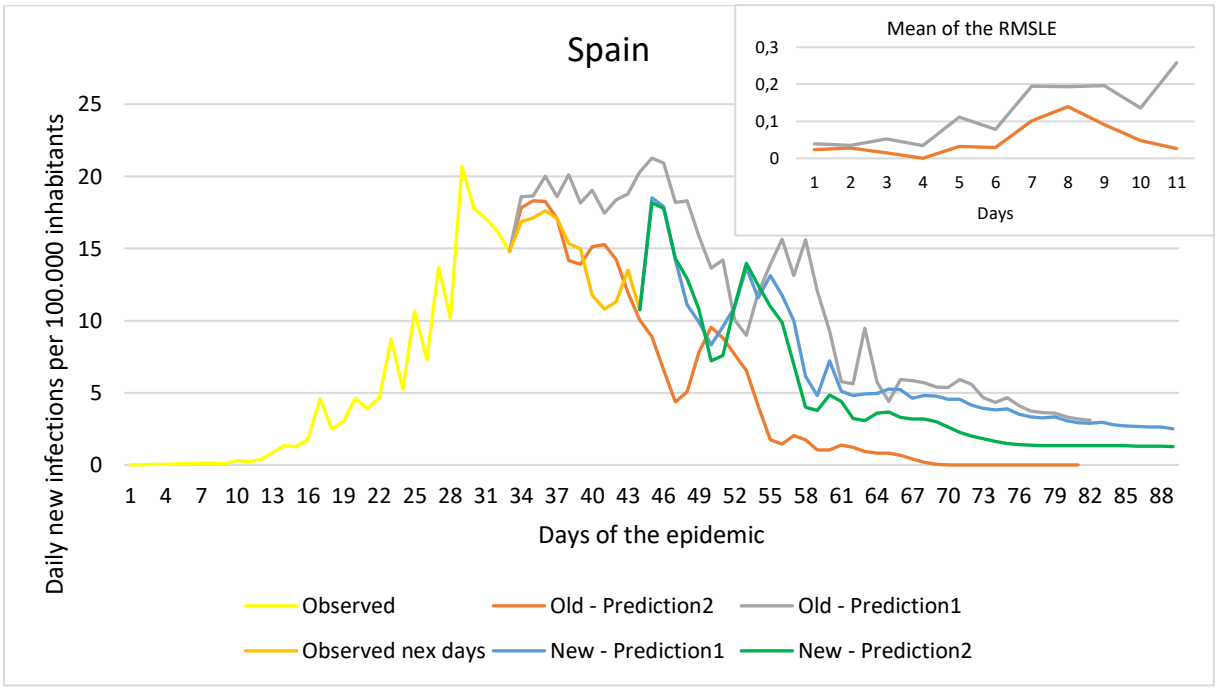

Fig 5. Observation and predictions for Spain. The small graph in the upper right corner shows the daily error values calculated for the predictions.

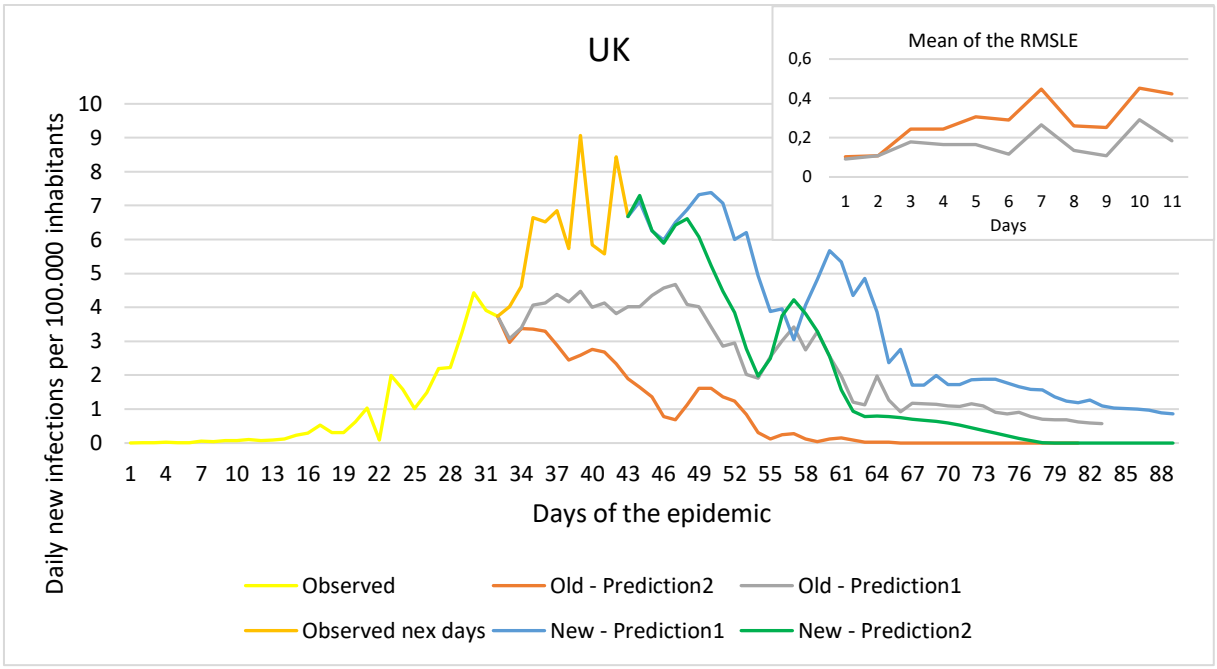

Fig 6. Observation and predictions for the United Kingdom (UK). The small graph in the upper right corner shows the daily error values calculated for the predictions.

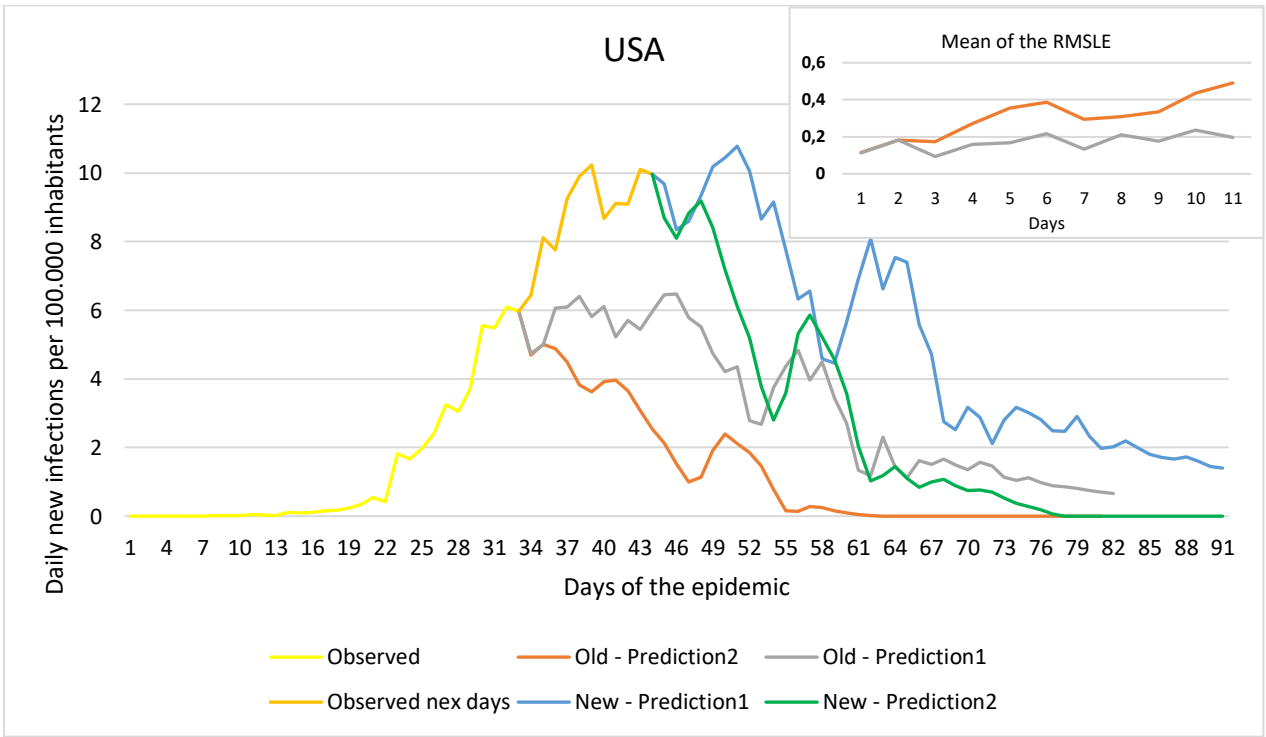

Fig 7. Observation and predictions for the United States of America (USA). The small graph in the upper right corner shows the daily error values calculated for the predictions

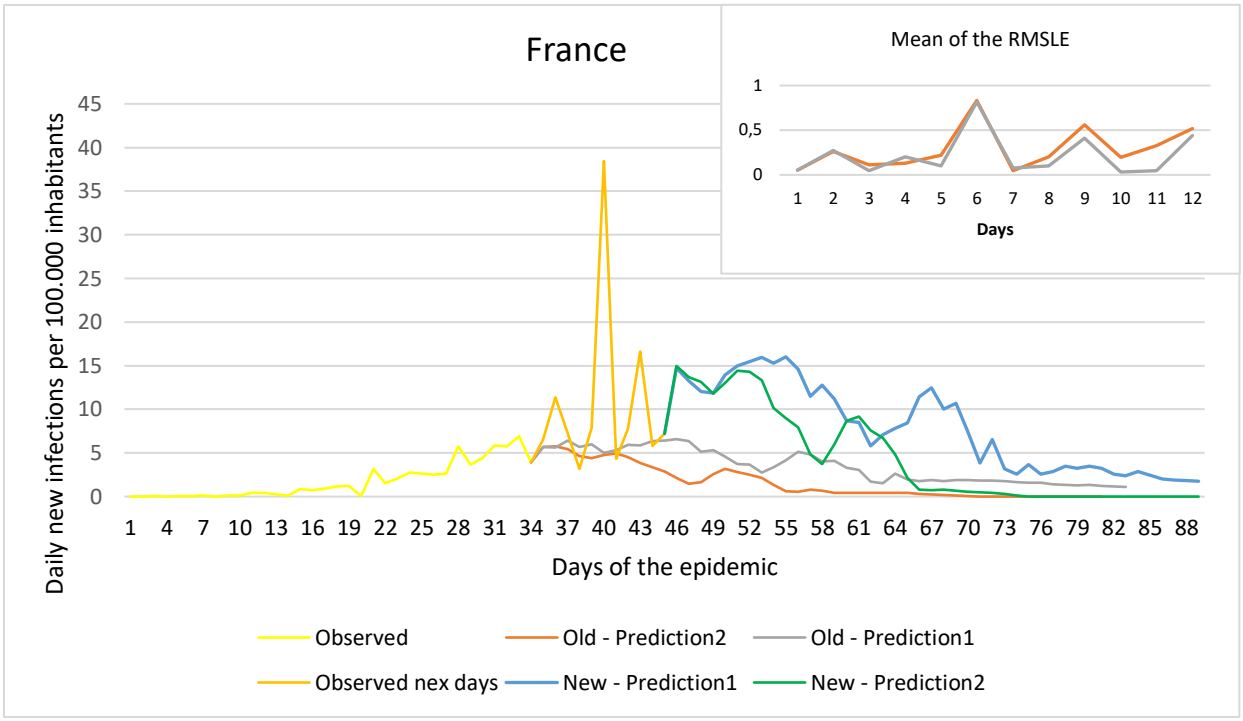

Fig 8. Observation and predictions for France. The small graph in the upper right corner shows the daily error values calculated for the predictions.

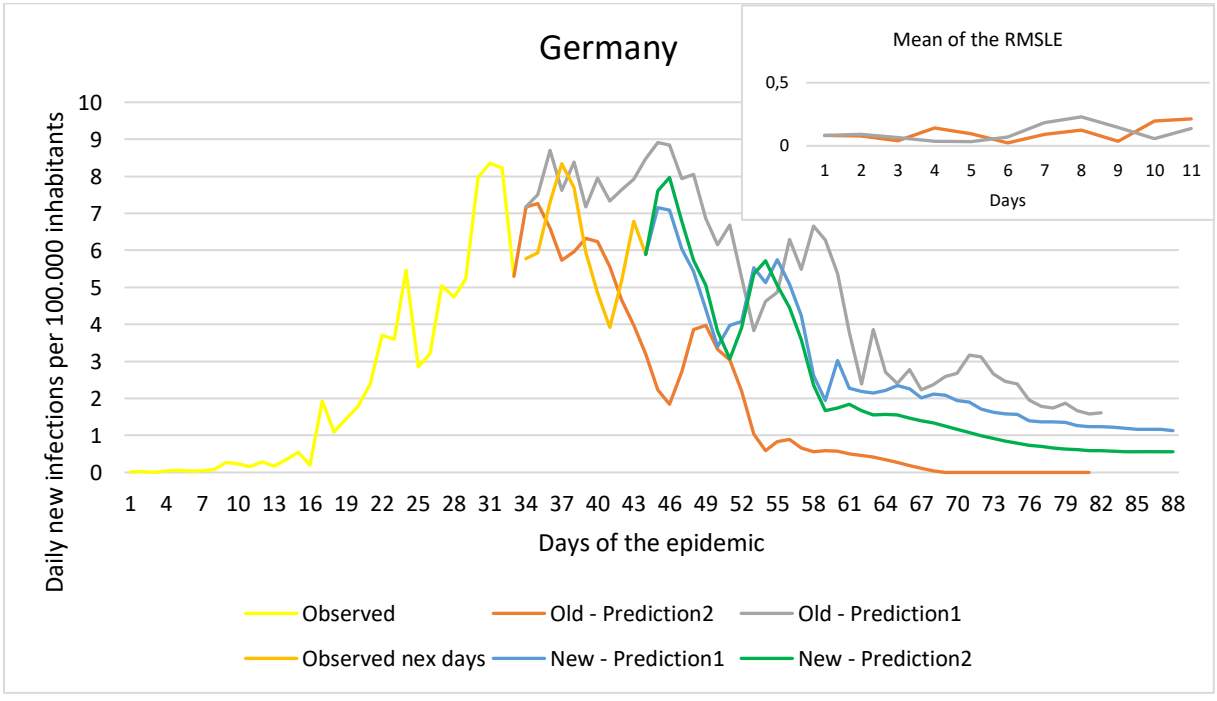

Fig 9. Observation and predictions for Germany. The small graph in the upper right corner shows the daily error values calculated for the predictions.

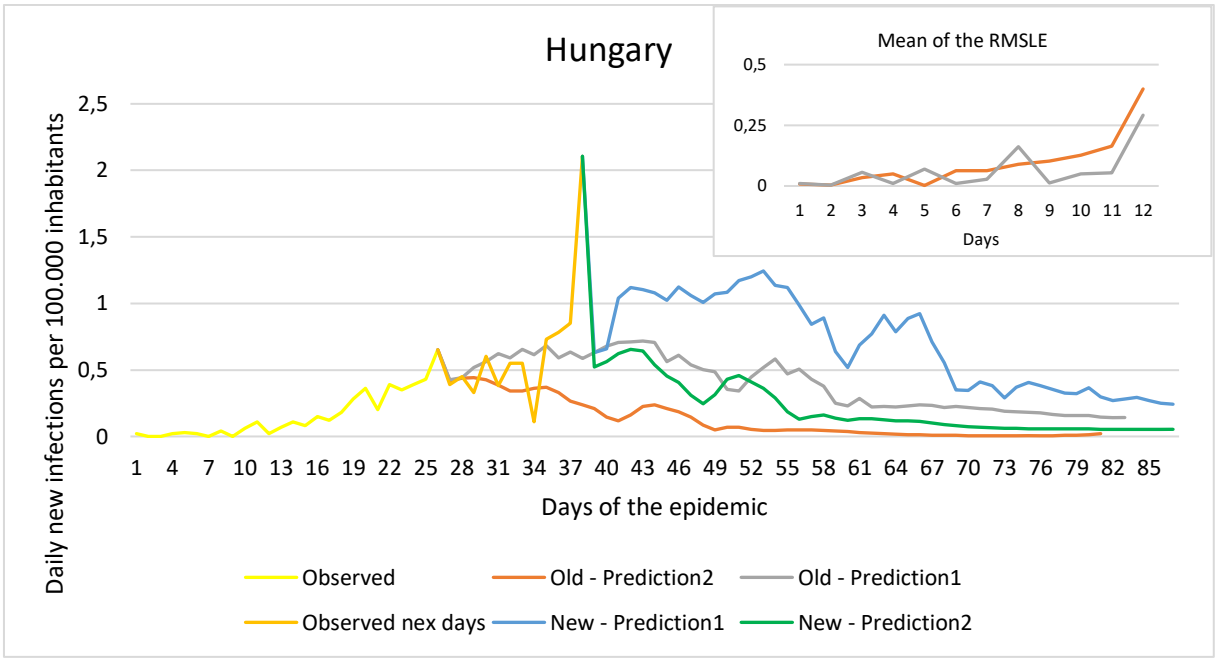

Fig 10. Observation and predictions for Hungary. The small graph in the upper right corner shows the daily error values calculated for the predictions.

The total errors for the entire investigated period, the summarized mean of the predictions (RMSLE) by country shown in Table 1.

Table 1. Total RMSLE error for the entire investigated period

Countries	Summarized mean of old prediction 1 (RMSLE)	Summarized mean of old prediction 2 (RMSLE)
Hungary	0.261	0.316
United Kingdom	0.405	0.533
Italy	0.392	0.291
Spain	0.348	0.221
Germany	0.321	0.319
France	0.443	0.517
USA	0.414	0.552

Discussion

The result of our study underscores that the COVID-19 pandemic is probably a propagated source outbreak, therefore repeated peaks on the epidemic curve (rise of the daily number of newly diagnosed infections) are to be anticipated. Predictions made using AI-based recurrent neural networks, further implicate that albeit majority of investigated countries are near or over the peak of the curve, they should prepare for a series of successively high peaks in the near future, until all susceptible people will be infected by the coronavirus, or effective preventive (eg. vaccination) or treatment options will become available. These scenarios are similar to other known propagated source epidemics, e.g. SARS-CoV and measles.[27] The validation of our first predictions shows a strong correlation with the progression of the newly diagnosed daily cases, the trends of our predictions are similar to the observed data, with relatively small calculated root mean squared logarithmic errors. Our recalculated predictions might be more precise, but the trends are very similar to the previous predictions and observed data.

Albeit suppression and mitigation measures can reduce the incidence of infection, COVID-19 disease, given its relatively high transmissibility reflected by average R0 values of 3.28, will continue to spread, most likely.[14] Accordingly, public health measures must be implemented as the incubation period of the virus may be long (1-14 days, but there are some opinions, that this can be 21 days), during which time asymptomatic or presymptomatic spreading may ensue. Moreover currently it is uncertain, whether those, who were diagnosed with COVID-19 infection, will acquire immunity or not.[11] Finally, data from countries with warm climate suggest that summer is unlikely to stop the pandemic, as the virus already spreading in Australia and South Africa as well.[13, 17] This is why the recurrence of another peaks is very likely, and the end of the pandemic cannot be accurately predicted at this time.

Nevertheless, recent publications showed, that the earlier mitigation attempts are in place (eg. border closure, closing schools, lockdown of the country, curfew), the more effective is the reduction of spread of the epidemic.[17] In fact analyzing the effects of a suppression strategy with respect to COVID-19, it was shown that early implementation of suppression at 0.2 deaths per 100 000 population per week could save 30.7 million lives compared to late implementation of these measures at 1.6 deaths per 100 000 population per week.[28] This seems to be the case in the countries, which had prior knowledge regarding coronavirus infections (eg. China, Singapore, Hong Kong), as they were more prepared to implement public health measures, had more equipment and health care personnel in place to mitigate the spread of the infections. Those countries, that failed to implement efficient and strict mitigation policies in a timely manner, are facing difficulty with controlling the disease, as is the case in Italy, the United Kingdom and the United States.[14]

To the best of our knowledge this is the first study to model the predicted evolution of the newly diagnosed infections using data from official databases with the help of the artificial intelligence-based recurrent neural networks trained on the currently available data, which were validated by root mean squared logarithmic errors calculation. Most studies to date expect a single peak of the epidemic curve, but some fear the emergence of future peaks when mitigation-suppression measures will be discontinued. According to our model, this can even happen, if the strict measures are sustained.

Limitations of our study: As the nature of COVID-19 virus is relatively unknown, and it is prone to mutations, the prediction of the spread of the pandemic is not easy. Factors influencing known new cases per day, for example efficiency of reporting, the different quality and timing of public health measures, the country-specific age-pyramid, chronic disease burden of the population were not included in the training data set, due to lack of reliable data. We did not investigate the number of the deaths and recoveries, as we found no reliable data. Similarly, the

data regarding diagnostic tests performed per country, or death rates were omitted, given they are highly influenced by the countries' economic wellbeing, health care systems, facilities and capacities and other factors. [29, 30] There are lots of unforeseen uncertainties and coincidences, which could not be implemented in our model, for example there were days, when a large number of people were diagnosed with COVID-19 one day (for example in care homes in France or in Hungary) that caused a large increase in the number of the daily new cases.[14]

Summarizing, the COVID-19 disease is a global health challenge, which caused the WHO to declare a "public health emergency of international concern on 30/01/2020".[16] The influence of this global epidemic has dug deep into the day-to-day conduct of everyone, with unforeseen challenges still pending for governments and policymakers. Starting from this, everyone, especially decision makers must be aware, that the current situation might be just the beginning, and even if strict public health measures are executed and sustained, future peaks of infections are possible.

Conclusions

The results of our study underscore that the COVID-19 pandemic is probably a propagated source epidemic, therefore repeated peaks of the rise of the daily number of newly diagnosed infections are to be anticipated.

To the best of our knowledge this is the first study to model the predicted evolution of the pandemic using data from official databases with the help of the AI-based RNNs trained on the currently available data regarding the spread of the disease and validated with comparison of the predicted and observed data. Most studies to date expect a single peak on the epidemic curve, but some fear the emergence of future peaks when mitigation-suppression measures will be discontinued. According to our models, this can even happen, if the strict measures are

sustained. The AI-based predictions might be useful tools and can be recalculated according to the new observed data to get more precise forecast of the pandemic.

Acknowledgement

Authors and contributions

All authors worked on the text writing and editing of the manuscript. The study was designed by LK, TB, ZJ, AH. AH and TB, AT, IV led the data management and analysis, TB and GS made the data extraction, collection and analysis, the computer AI-based analysis was made by AT, BT, IV, AH. The figures were made by AT, BT, IV, GS, AH, LK. The literature search was done by LK, AH, SH, GS, SG, JZ, RG. The interpretation of the literature, methods, results was made in close collaboration by all co-authors.

Funding

This study was supported by the European Union, co-financed by the European Social Fund and European Regional Development Fund [grant No. EFOP-3.6.1-16-2016-00022 "Debrecen Venture Catapult Program" (providing support for SH, LK), by the European Union [EFOP-3.6.2-16-2017-00009 "Establishing Thematic Scientific and Cooperation Network for Clinical Research" (providing support for RG), by the János Bolyai Research Fellowship of the Hungarian Academy of Sciences (providing support for JZ), by the Hungarian Brain Research Program 2.0 under grant number 2017-1.2.1-NKP-2017-00002 and the ED_18-1-2019-0028 providing support for (LK, TB, AT). Research was also supported by the ÚNKP-19-3 – I. New National Excellence Program of the Ministry for Innovation and Technology (AT). Research was supported in part by the project EFOP-3.6.2-16-2017-00015 supported by the European Union, co-financed by the European Social Fund.

The role of the funding source

The funding sources had no role in the writing of the manuscript of the decision to submit it for publications, no involvement in data collection, analysis, or interpretation; trial design; patient recruitment; or any aspect pertinent to the study.

We have not been paid to write this article by a pharmaceutical company of other agency.

Dr. László R. Kolozsvári, the corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of competing interest

All authors declare no conflicts of interest.

Data sharing

All data sources are publicly available and described in the methods section.

Patient and other consents

Non applicable. We involved no individual patient in our study.

References

1 Han Q, Lin Q, Jin S, You L. Coronavirus 2019-nCoV: A brief perspective from the front line. Journal of Infection. 2020;

2 Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med. 2003;

3 Chan-Yeung M, Xu RH. SARS: Epidemiology. Respirology. 2003;

4 Cheng VCC, Lau SKP, Woo PCY, et al. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. Clinical Microbiology Reviews. 2007;

5 Peiris JSM, Lai ST, Poon L, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. The Lancet. 2003; 361(9366): 1319-1325.

6 de Groot RJ, Baker SC, Baric RS, et al. Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. J Virol. 2013;

7 Eastern Mediterranean Region. World Health Organization. March 29, 2020. http://www.emro.who.int/pandemic-epidemic-diseases/mers-cov/mers-situation-update-january-2020.html (accessed March 29, 2020).

8 Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;

9 Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;

10 Sun P, Qie S, Liu Z, et al. Clinical characteristics of hospitalized patients with SARS-CoV-2 infection: A single arm meta-analysis. J Med Virol. 2020; 1–6. published online Feb 28. https://doi.org/10.1002/jmv.25735

11 WHO. World Health Organization. Q&A on coronaviruses (COVID-19). April 08, 2020. https://www.who.int/news-room/q-a-detail/q-a-coronaviruses (accessed April 10, 2020).

12 Wu J, Zha P, Med C. Association of COVID-19 Disease Severity with Transmission Routes and Suggested Changes to Community Guidelines. medrxiv.org. 2020;

13 Yuen KS, Ye ZW, Fung SY, et al. SARS-CoV-2 and COVID-19: The most important research questions. Cell Biosci. 2020;

14 WHO. World Health Organization. Novel Coronavirus (2019-nCoV) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (accessed April 15, 2020).

15 Liu Y, Gayle AA, Wilder-Smith A, et al. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med. 2020;

16 Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. Jama. 2020;

17 Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. March 16, 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf (accessed March 16, 2020).

18 Hethcote HW. Modeling heterogeneous mixing in infectious disease dynamics. Cambridge: Cambridge University Press. 1996; (pp. 215-238).

19 Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. Virulence. 2013; 4(4): 295-306.

448 20 Szolovits P. Artificial intelligence in medicine. Routledge. 2019;

449 21 Zhu X, Fu B, Yang Y, et al. Attention-based recurrent neural network for influenza
450 epidemic prediction. BMC bioinformatics. 2019; 20(18): 1-10.

451 22 Dupond S. A thorough review on the current advance of
452 neural network structures. Annual Reviews in Control. 2019; 14: 200–230.

453 23 Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997; 9
454 (8): 1735–1780. published online Nov 15. https://doi.org/10.1162/neco.1997.9.8.1735.

455 24 Johns Hopkins University. Coronavirus Resource Center. March 30, 2020.
456 https://coronavirus.jhu.edu/data/new-cases (accessed April 12, 2020).

457 25 Yoon S, Yun H, Kim Y, et al. Efficient transfer learning schemes for personalized
458 language modeling using recurrent neural network. In Workshops at the Thirty-First AAAI
459 Conference on Artificial Intelligence. 2017;

460 26 Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks.
461 2014; published online Sep 10. https://arxiv.org/abs/1409.3215v3.

462 27 CDC. Centers for Disease Control and Prevention. Using an Epi Curve to Determine Mode
463 of Spread. April 04, 2020. https://www.cdc.gov/training/QuickLearns/epimode/6.html
464 (accessed April 12, 2020).

465 28 Walker PG, Whittaker C, WatsonO, et al. The Global Impact of COVID-19 and Strategies
466 for Mitigation and Suppression. On behalf of the imperial college covid-19 response team.
467 Imperial College of London. 2020;

468 29 Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS ONE
469 15(3) : e0231236. 2020.

30 Anastassopoulou C, Russo L, Tsakris A, Siettos, C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS ONE 15(3), e0230405. 2020.