



Neural network powered COVID-19 spread forecasting model

Michał Wieczorek, Jakub Siłka, Marcin Woźniak*

Faculty of Applied Mathematics, Silesian University of Technology, Kaszubska 23, 44-100 Gliwice, Poland



ARTICLE INFO

Article history:

Received 15 July 2020

Accepted 10 August 2020

Available online 15 August 2020

2010 MSC:

68T05

68T37

60G25

Keywords:

COVID-19

Prediction

Neural network

ABSTRACT

Virus spread prediction is very important to actively plan actions. Viruses are unfortunately not easy to control, since speed and reach of spread depends on many factors from environmental to social ones. In this article we present research results on developing Neural Network model for COVID-19 spread prediction. Our predictor is based on classic approach with deep architecture which learns by using NAdam training model. For the training we have used official data from governmental and open repositories. Results of prediction are done for countries but also regions to provide possibly wide spectrum of values about predicted COVID-19 spread. Results of the proposed model show high accuracy, which in some cases reaches above 99%.

© 2020 Published by Elsevier Ltd.

1. Introduction

Virus spread prediction is very important on the way to plan situation. The problem with modeling such system is that each day of COVID-19 and the number of new potential cases can not be determined in a simple mathematical equation. There are many reasons for such problems. Spread of human threads in general depends on various features both dependent from human behavior but also biological structure of the virus itself. In any case the research must be done to biologically describe the virus for the development of medical treatment but also to model spread which will help to prevent new cases and concentrate on locations with highest potential needs.

1.1. Current knowledge about COVID-19

The COVID-19 virus belongs to the Coronaviridae family and order Nidovirales, thus it contains a few taxonomy hallmarks. Those have lipidic envelope or positivesense, single strand RNA within large ribonucleoprotein (RNP) core. All of such viruses contain three characteristic proteins anchored in the viral membrane: the envelope protein (E), the transmembrane glycoprotein (M) and Spike (S) glycoprotein. As it is worth noticing, the viral envelope exhibits a large thickness, almost twice that of a typical biological membrane [1]. The genome of SARS-CoV-2 consists of six major

open-reading frames (ORFs), frequently identified in other CoVs. It seems, that some of the genes shared less than 80% nucleotide sequence identity to SARS-CoV [2].

COVID-19 is sensitive (like other CoVs) to ultraviolet rays and heat - it is thought that this virus can be inactivated at 27 °C. Moreover, this virus might be inactivated by ether (75%), ethanol, chlorine-containing disinfectant, peroxyacetic acid and chloroform except for chlorexidine [3]. The thought of the vast impact that coronaviruses may have on Public Health is underpinned by the fact that throughout the history, a lot of them were able to cross species barriers and finally affect human health by causing common cold, or more severe diseases, like SARS and MERS [3].

A large-scale study based on 1995 cases have shown that main clinical symptoms are fever (88.5% of cases), cough (68.6% of cases), myalgia or fatigue (35.8% of cases), expectoration (28.2% of cases), and dyspnea (21.9% of cases), while the minor ones include headache or dizziness (12.1% of cases), diarrhea (4.8% of cases), nausea and vomiting (3.9% of cases) [4]. The transmission of novel coronavirus, alike many pathogens is believed to occur through respiratory droplets, thereby the vast majority of spread cases are limited to the close spaces. The incubation time last usually 3–7 days, however the symptoms might occur to up to 2 weeks [3].

1.2. Related COVID-19 prediction methodologies

In recent months several important models were presented. In [5] machine learning was applied to estimate how the outbreak of this thread will go. Nevertheless it not easy to predict situation in

* Corresponding author.

E-mail addresses: michal_wieczorek@hotmail.com (M. Wieczorek), kubasilka@gmail.com (J. Siłka), marcin.wozniak@polsl.pl (M. Woźniak).

case of COVID-19, since there are many factors which determine rapid changes [6]. Therefore many approaches were used to help.

In [7] prediction of thread was done by mathematical model, in which undetected infections were estimated for China region. Sometimes even very simple methodologies are used. When decision is necessary immediately we can start prediction based on data preprocessing in which some cases are simply removed for applied model on Euclidean network [8]. In Japan prediction models also evaluated first symptoms [9]. One of the first presented models for Italy was using Gauss error function and Monte Carlo simulation on recorded cases [10]. Also stochastic predictors give potential help in first days when not much data is available for machine learning approaches [11]. Such stochastic models are also presented to work even for very large societies like in India [12]. Therefore when Artificial Intelligence is applied in the first days of prediction periods the results are mostly for single region or country. One of the first approaches for China was presented in [13]. An interesting discussion of principles for using mathematical modeling was presented in [14]. Some methodologies not only predict numbers of new cases but also make some assumptions regarding dynamics of growth [15]. There are many sources of information to predict situation. As reported in [16] social media can bring valuable information not only about confirmed cases but further spread. Relations between new cases and speed or reach of growth can be transformed to prediction in other location, as shown in [17] such knowledge transfer to model another region was done between Italy and Hunan province in China. Diamond Princess ship case was discussed in [18]. There are also models which estimate situation in larger regions or more than one country. In [19] and [20] applied prediction model was defined to work with data from China, Italy and France. Some models treat only the total number of cases in the whole World altogether [21].

The model proposed in our article is a complex solution. Proposed Neural Network architecture was developed to flexibly predict new cases in various countries and regions of the World. The architecture is composed of seven layers and the output predicts the number of new cases. Our predictor is trained by NAdam, which was chosen in tests for the best efficiency and shortest training time. Several other test enabled us to select the best time window in which data is fed to the network. As the results have shown such composed model is able to predict new cases with very high efficiency, which in some regions and countries is above 99%. In the following section we discuss our approach by showing how the data set was selected and preprocessed. Further we discuss network construction and results for tested training models. In final section we present results and compare them to classic statistical approaches and draw conclusions from our research.

2. Dataset

In our work we were using dataset provided by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University on their github page. It contains combined data from numerous different sources such as:

- World Health Organization (WHO),
- European Center for Disease Prevention and Control (ECDC),
- DXY.cn. Pneumonia. 2020,
- COVID Tracking Project,
- National Health Commission of the People's Republic of China (NHC),
- China CDC (CCDC),
- Washington State Department of Health,

and many other smaller, regional health departments. This dataset is free for academic purposes and contains time series data with daily-updated total confirmed cases for countries and some bigger

Table 1
Comparison of Accuracy for different time-steps for error margin of 0.2.

Time Step	Accuracy
6 days	80.45%
8 days	79.58%
10 days	79.39%
12 days	79.90%
14 days	80.62%
16 days	78.75%

regions. Because it's all combined in a single csv file it is very accessible to use for our neural network predictor.

2.1. Data preparation

The dataset contained three categories of data:

- Region name,
- Geolocation in the form of Latitude and Longitude,
- Total cases count per day.

In our model we are focusing on predicting total cases count so we have removed the region name and localization data from the dataset giving each country numerical ID to later identify predicted data and show it in the form of a table, however the ID is not used in the training process. In order to prevent overfeeding we have sorted discerningly our region by the maximum cases count and then split them to separate files containing maximum of 30 regions per file. Using 20 regions/file resulted in decreased performance by about 3% compared to 30 r/f and using 40 and 50 also decreased accuracy by 1% and 1.67%. After division we train a separate network for each file and then combine all results in one, final file.

2.2. Finding the best time-step

Because we are working with time-step data we had to specify on how many previous days we want to predict future values. In order to choose the best option we had to prepare an experiment on which we tested few different time-steps and analyzed the results. Gathered data are shown in [Table 1](#) and [Fig. 1](#).

As we can see smaller time-step allows the network to adapt more quickly to the changes, so it reduces the delay between first few days of evaluation period in some regions and the adjustment of the network. Because of that there is a growth of accuracy in early days of period. However, because the network is using few days for the prediction, the model is also less stable and has a higher fluctuations if changes are higher than normal, what results in predicted values shaking up and down on the plot reducing overall accuracy. Higher time-step on the other hand, produces very stable data what results in more accurate predictions later during the period, however there is a bigger delay in adaptation in the early stages compared to smaller time-step.

Having all of this pros and cons we had to find the balance between fast adaptation and stability of the predicted data. Because the spread has already started in most regions there is smaller need for fast adjustment but higher for the reliability of the data so we decided to go for some higher time-step. In our tests we found out that the best option for our predictor was 14 days evaluation period which not only produces accurate plot but also has the highest overall accuracy of the model.

2.3. Data normalization

To normalize our data before training we are dividing all values from the dataset by the maximum value found in the pool to get

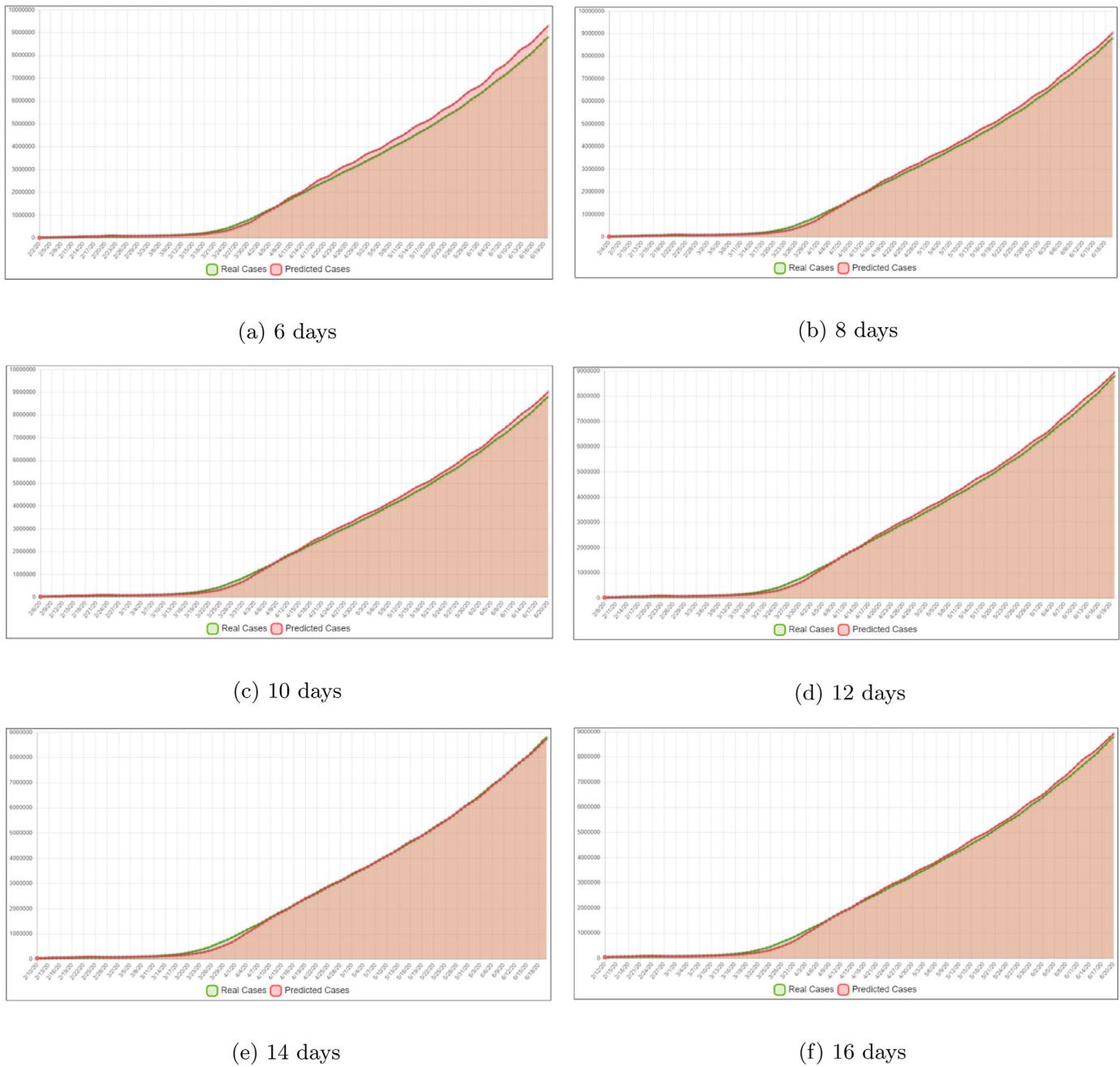


Fig. 1. Predicted total cases curve using different time-steps.

real values in scale 0 to 1. After training, to get the final values we are multiplying results by the same maximum value we were dividing them before and we are rounding it to get the integer value.

2.4. Error margin

Because we are not working with classification but with prediction we need to have a method to compute accuracy of our model. To do this we have used a measure of error margin of 0.2 computed in such way:

$$a = \begin{cases} predicted_value > real_value - (real_value * error_margin) \\ predicted_value < real_value + (real_value * error_margin) \end{cases} \quad (1)$$

where a is match. As we can see the margin value is relative to the total cases count. That gives us possibility to normalize the accuracy calculation and use the same equation for regions with small number of cases as well as for most populated regions.

2.5. Test/train division

After analyzing most of the regions we have noticed that although different regions have some similarities in growth curve there are however some major differences. If we want to accurately predict the virus spread in every region we cannot train the network on some countries and test on other. Such approach would generalize the spread prediction and create universal but semi-accurate model without ability to predict variations. Because of that we have divided our data in a different way - we have trained our model on data from 5 days period and older, and tested on

the newest data. That allowed our network to create generalized global predictor but also to accurately adapt to every region separately. This method however is not 100% perfect. Because the network is trained on older data, in some cases there may be a little slip in the beginning of COVID-19 spread and the network would need few days to adapt to the new situation and accurately predict new values.

2.6. Finding the best training iterations count

Our model works in such way that it does not have any need to specify the perfect number of iterations to get the best possible accuracy. It saves the network's best performing synaptic weights during the training and after the training it restores them and later makes predictions. Because of that, we can safely give the network more iterations than needed. It is however not recommended because of limited time and resources.

To find the minimum number of iterations that can safely give us the best result we have analyzed accuracy and loss. Based on them we specified the best number of iterations to 3300. In most cases it was more than needed but in some regions the network needed even more iterations to perform the best. During our research we have found out that going over 3300 iterations did not give our model more positive results, so we defined it as the best option.

3. Prediction model

It was proven that neural networks predictors can result in accuracy and better adjustment than other stochastic methods. On the other way there are many different architectures of neural networks. In this article we present predictor which results in highest efficiency. In our research for time-step data we have also tried to use Recurrent Neural Network (RNN) for total cases prediction, however in our case it performed about 1–2% worse than our ANN model. Additionally, even with GPU acceleration network training with the same parameters was about 3 times longer. Because of that, finally we decided to use classic ANN in our predictor. Sample comparisons between RNN and ANN in accuracy and loss are shown in Fig. 3 (for ANN model) and Fig. 4 (for RNN model). As we can see ANN not only scores higher than RNN but also reaches the maximum accuracy in much less iterations making it the most suitable architecture for our problem.

3.1. Applied architecture

In our model we have used classic, fully connected layers in Neural Network presented in Fig. 2. To predict the value of cases we have used two activation functions:

- Hyperbolic Tangent - for all hidden layers,

$$T(x) = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

$$T'(x) = \frac{1}{\cosh^2 x} = 1 - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} \quad (3)$$

- ReLu

$$R(x) = \max(0, x) \quad (4)$$

$$R'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x > 0 \end{cases}$$

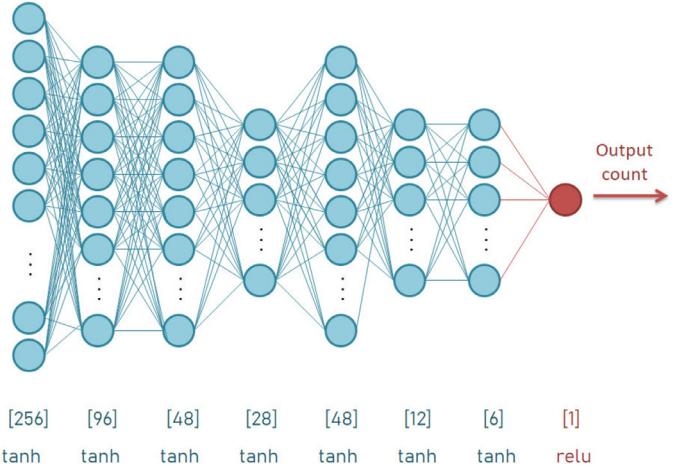


Fig. 2. Developed ANN architecture.

Table 2
Comparison of Accuracy
for tested training algo-
rithms.

Algorithm	Accuracy
Adadelta	0%
SGD	9.80%
Adagrad	40.10%
Ftrl	40.10%
Adamax	87.47%
Adam	87.53%
RMSprop	87.65%
NAdam	87.73%

3.2. Looking for the best optimization algorithm

To find the best solution for our network training we have done some tests using most known optimization algorithms. After analysis we have selected one with the best accuracy and the fastest learning. Results are shown in Table 2. Conclusions are as follows:

- Adadelta - gave us the worst results of actually 0% accuracy,
- SGD - the network was slowly learning however it did not reach any high accuracy,
- Adagrad - loss function was decreasing very slowly however the final accuracy was much higher than for SGD,
- Ftrl - similar to Adagrad,
- Adamax - loss function decreased fast however high accuracy was reached very late during the training process,
- Adam - high accuracy was reached much faster than in Adamax and was even little bit higher,
- RMSprop - network reached around 87.65% accuracy and made it half the time of NAdam algorithm,
- NAdam - final accuracy was the highest however it was reached much later than in RMSprop.

Results of our tests are presented in Fig. 5. In the end we decided to use NAdam, because it gave us the best accuracy overall.

3.3. Network training by using NAdam

In order to get the best possible accuracy of our model, based on previously conducted tests, we have used an improved version of Adaptive Moment Estimation Algorithm - NAdam. Because of its high performance and small requirement of computing power it's widely used in Artificial Intelligence research. To adjust our network's prediction even more, we have also used a learning rate

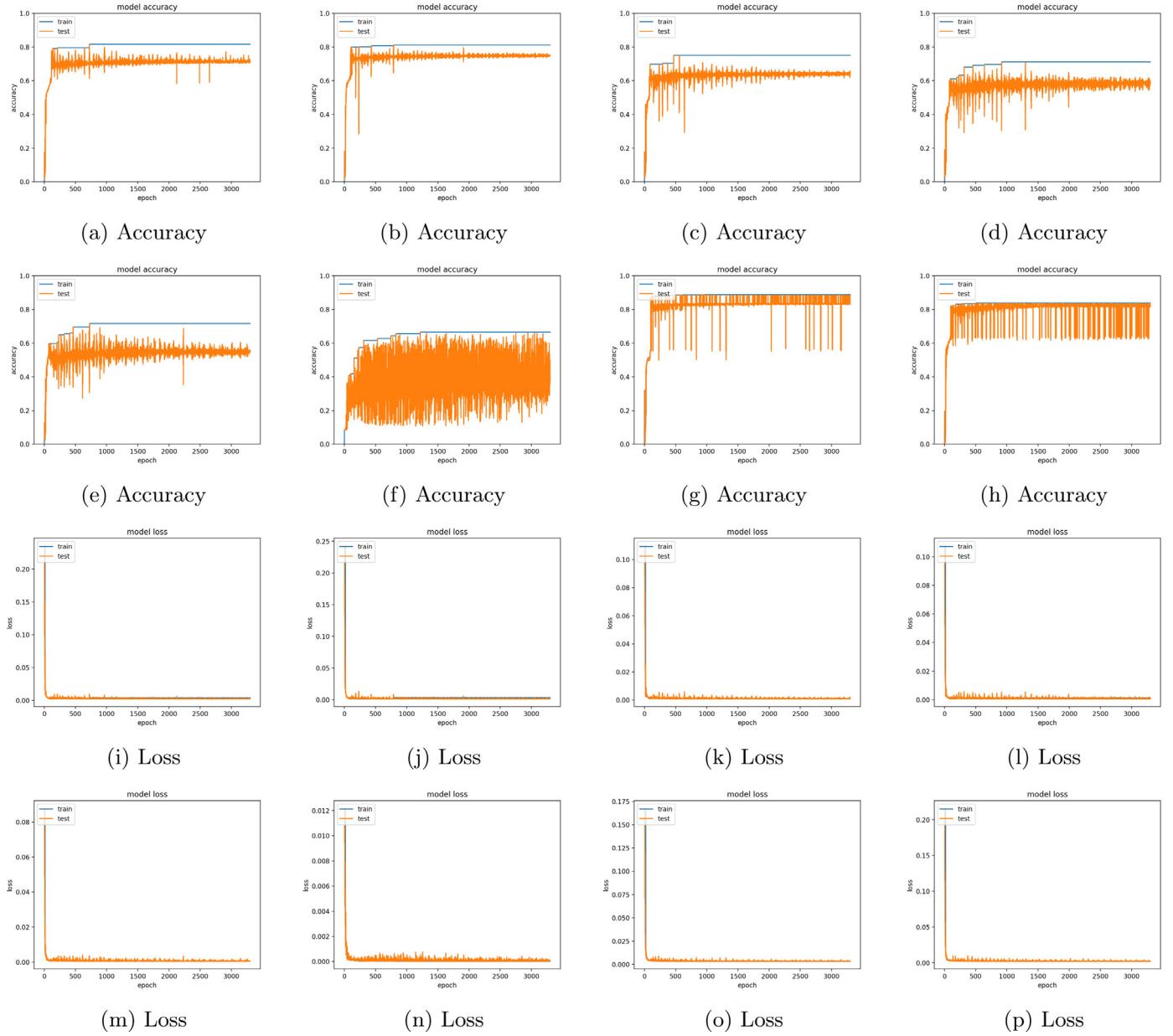


Fig. 3. Test results of using ANN on different parts of the research dataset.

decay to speed up rough adaptation to the cases curve and then, using much smaller steps, slowly polish the final model.

NAdam formula can be described as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (5)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (6)$$

where g is current gradient value of the error function and β is constant values called hyper-parameter. Values m_t and t_m are used for calculation of the correlations marked as \hat{m}_t and \hat{v}_t according to equations:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (7)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \quad (8)$$

Using correlation calculated above, the final formula for changing weights in our ANN architecture was defined as follows:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t, \quad (9)$$

where ϵ is a constant small value and η is a learning rate (in our case with value 0.0004).

We applied NAG to Adam using equations below:

$$w_t = w_{t-1} - \eta \frac{\beta_1 m_{t-1}}{\beta_2 v_{t-1} + (1 - \beta_2) g_t^2 + \epsilon} - \eta \frac{(1 - \beta_1) g_t}{\sqrt{\beta_2 v_{t-1} + (1 - \beta_2) g_t^2 + \epsilon}} \quad (10)$$

We modify standard Adam update rule, because the first term no longer depends on g_t we need to change expressions for w_t and \hat{m}_t :

$$\hat{m}_t = (1 - \beta_1) g_t + \beta_{1,t+1} m_t \quad (11)$$

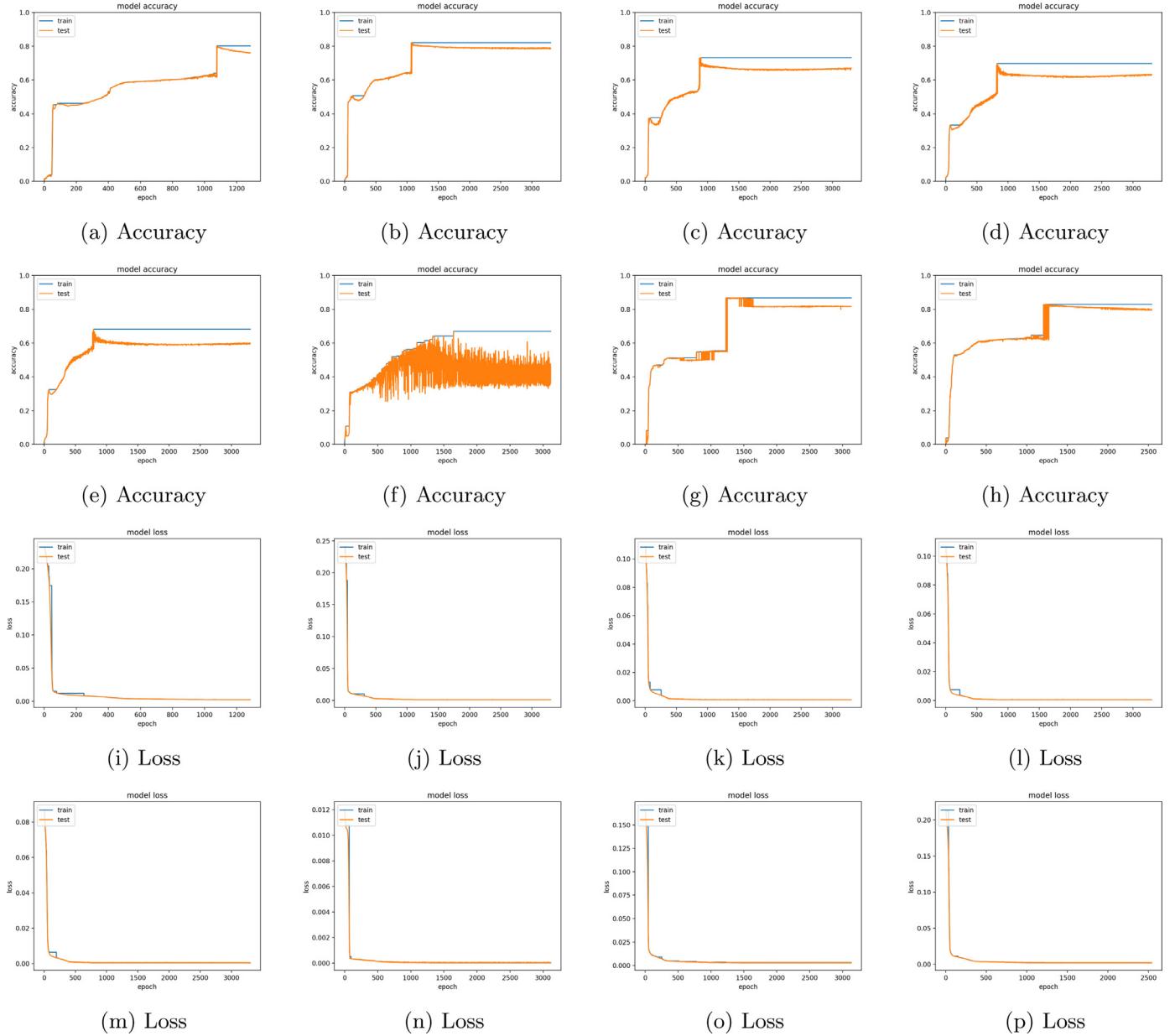


Fig. 4. Test results of using RNN on different parts of the research dataset.

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\beta_{2t}} + \epsilon} \quad (12)$$

4. Results

We will start discussion of our prediction from classic measures used in statistics. These measure will serve as some point in the discussion to present further accuracy of our predictor and final predictions for countries and regions.

4.1. Statistical modeling

In a statistical sense the shape of prediction line for confirmed cases can be affected by several factors, such as the number of tests carried out in country or region, restrictions imposed by state authorities, and also by behavior of people in a given area.

Sample Moving Average (SMA) of detected cases is calculated as

$$SMA_i = \frac{\sum_{i=1}^n c_i}{n} \quad (13)$$

where n is number of assumed factors, c_i is i th value.

Exponential Moving Average (EMA) of detected cases is calculated as

$$EMA_i = Y_i \cdot a + EMA_{i-1} \quad (14)$$

where $EMA_0 = Y_0$ and Y is a value from the i th value of the set taken into account, a - is $\frac{2}{n-1}$ where n is the number of values of the set.

Linear Trend line is built for the number of detected cases as

$$TL = a \cdot x + b \quad (15)$$

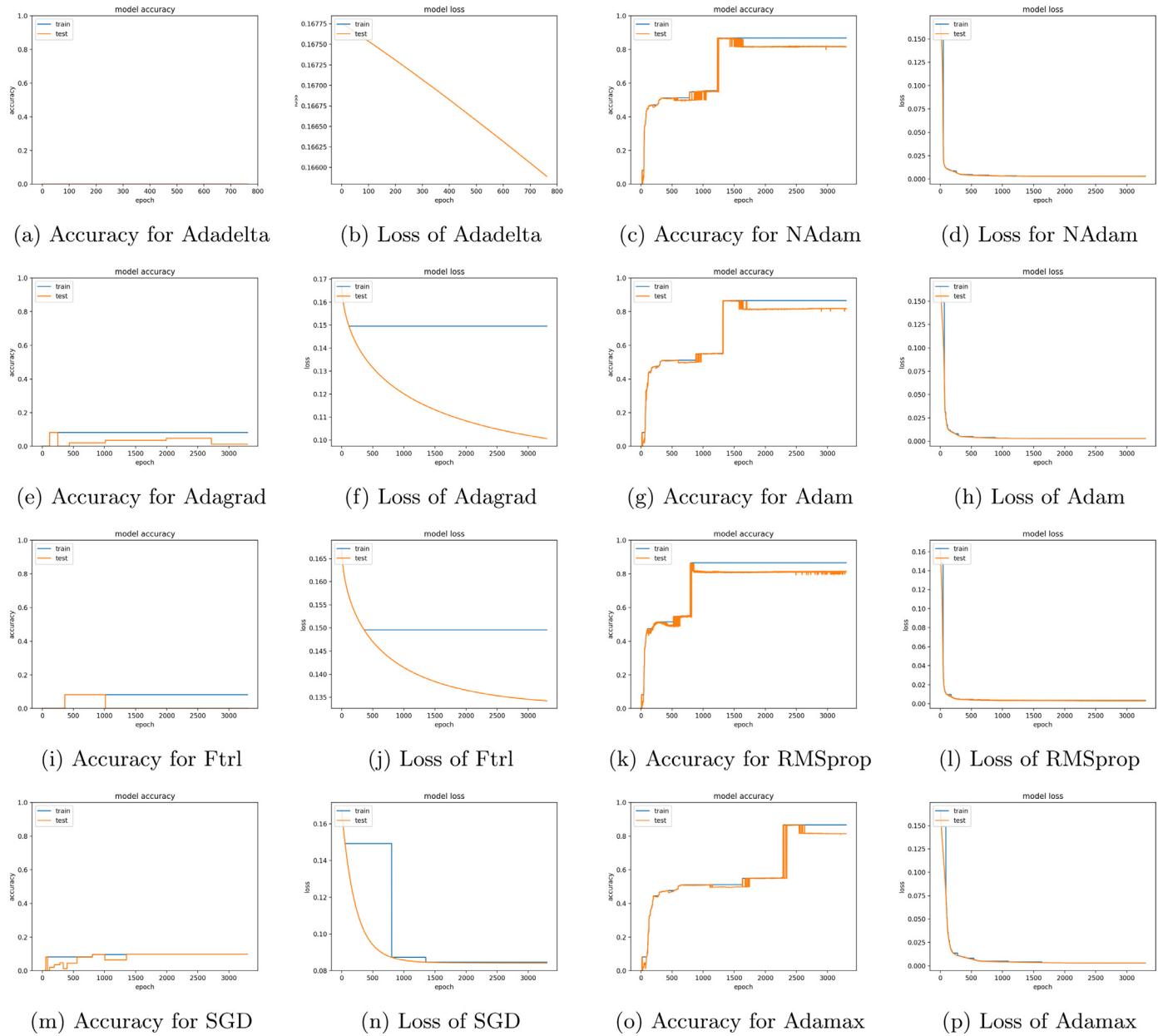


Fig. 5. Results of using different optimization algorithms. On this plots we can see the loss and accuracy during training for the World.

where X_i is i th value of the set, \bar{X} is arithmetic mean of the set, Y_i is i th value of the set, \bar{Y} is arithmetic mean of the set. Directional coefficient a of the trend line, and free expression b are calculated as

$$\left\{ \begin{array}{l} a = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ b = \bar{Y} - a \cdot \bar{X} \end{array} \right. \quad (16)$$

4.2. Prediction results

In Tables 3 and 4 we can see the final accuracy for selected regions and countries using different error margins in our predictor.

The accuracy was computed as an arithmetic mean of the sum of accuracy per day for the given time interval. High accuracy with the lowest error margin means that the network was very precise in predicting the value of cases and medium accuracy for the low error margin with high accuracy with the highest error margin mean that the network was predicting correctly the shape of the growth curve. Even in not all predicted values were exact from analyzed charts we can see that our predictor works well reaching high efficiency both for regions and countries.

Accuracy plots are shown in Figs. 9 and 10. For each of charts X-axis contains dates and Y-axis shows accuracy for that day with the error margin of 0.2 for selected region/country. As we can see when there are no active cases in some regions developed Neural Network correctly predicts number of cases. However in the beginning of period we can see fluctuations since to reach good performance the network needs proper amount of data, what was



Fig. 6. Graphs showing various statistical methods in comparison to real number of detected cases. Yellow - number of detected cases, Green (dotted line) - Sample Moving Average (SMA), Magenta - Exponential Moving Average (EMA), Blue - linear trend line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hard to reach in that time. When cases were recorded and data was available from various services developed network predictor quickly adapts to the new situation and the accuracy peaks up to almost 100%.

In Figs. 7 and 8 we can see the comparison between real and our predicted values for the total cases count plot. As we can see the network correctly predicts the main trend of cases growth,

however for some days the results are a little bit unstable. We conclude this is caused by rapid changes up and down in recorded cases what is normal situation in case of virus detection. Usually it happens when there are some additional variables changing the speed of the spread. On the other hand it is also visible that our predictor learns such situation and returns to correct trend easily.



Fig. 7. Sample predicted COVID-19 cases growth curve in selected countries.

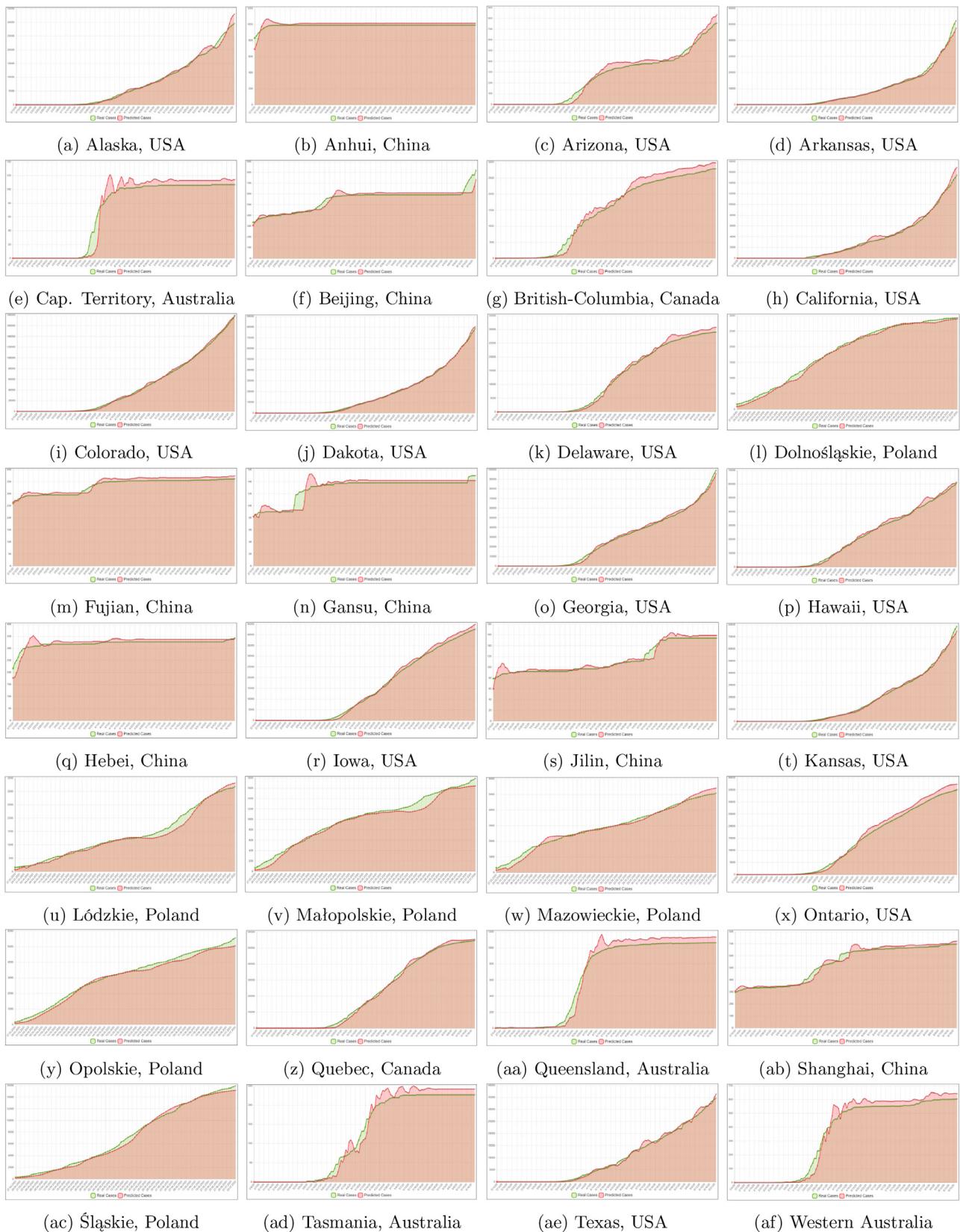


Fig. 8. Sample predicted COVID-19 cases growth curve in selected regions.

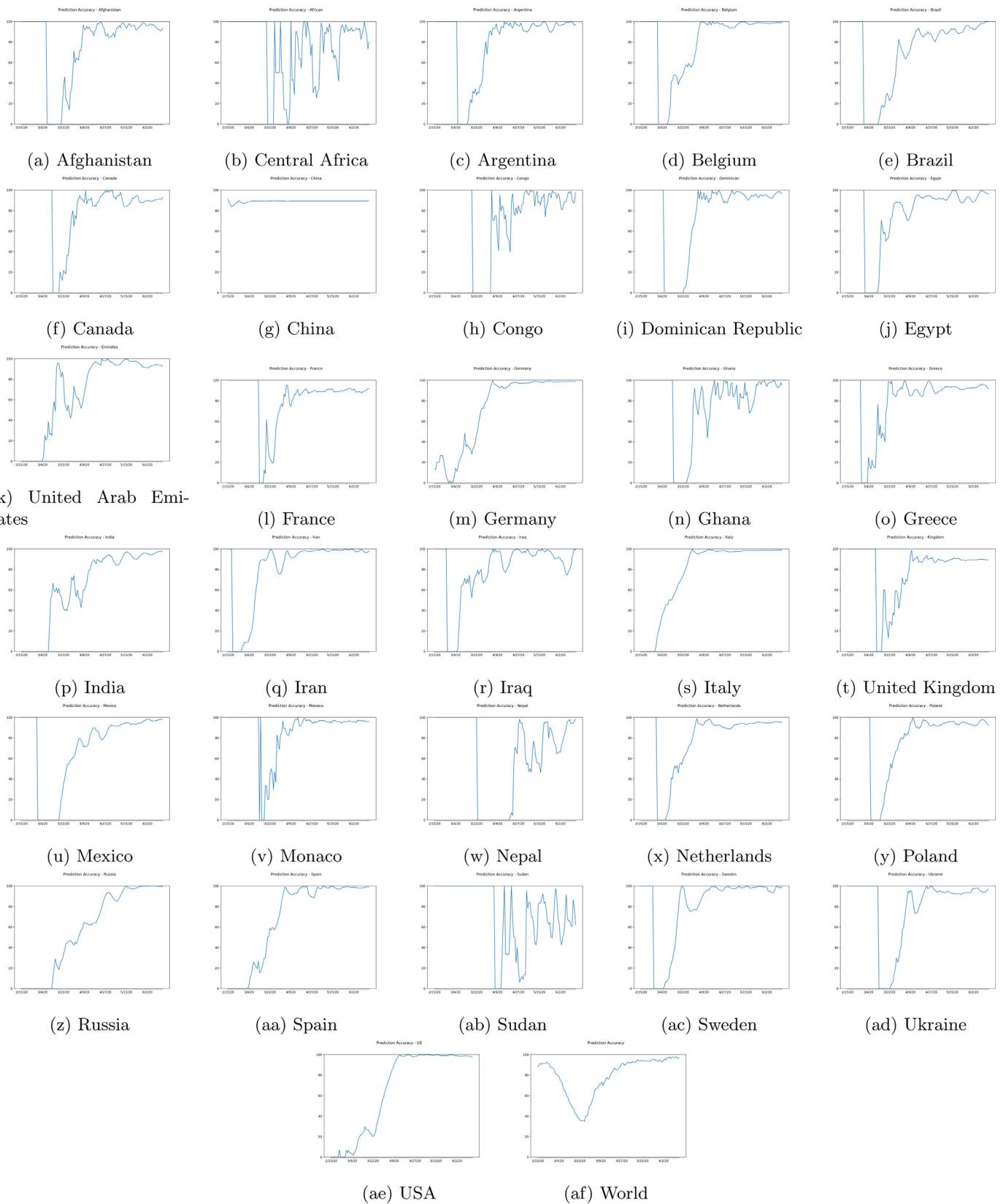


Fig. 9. Accuracy per day for selected countries.

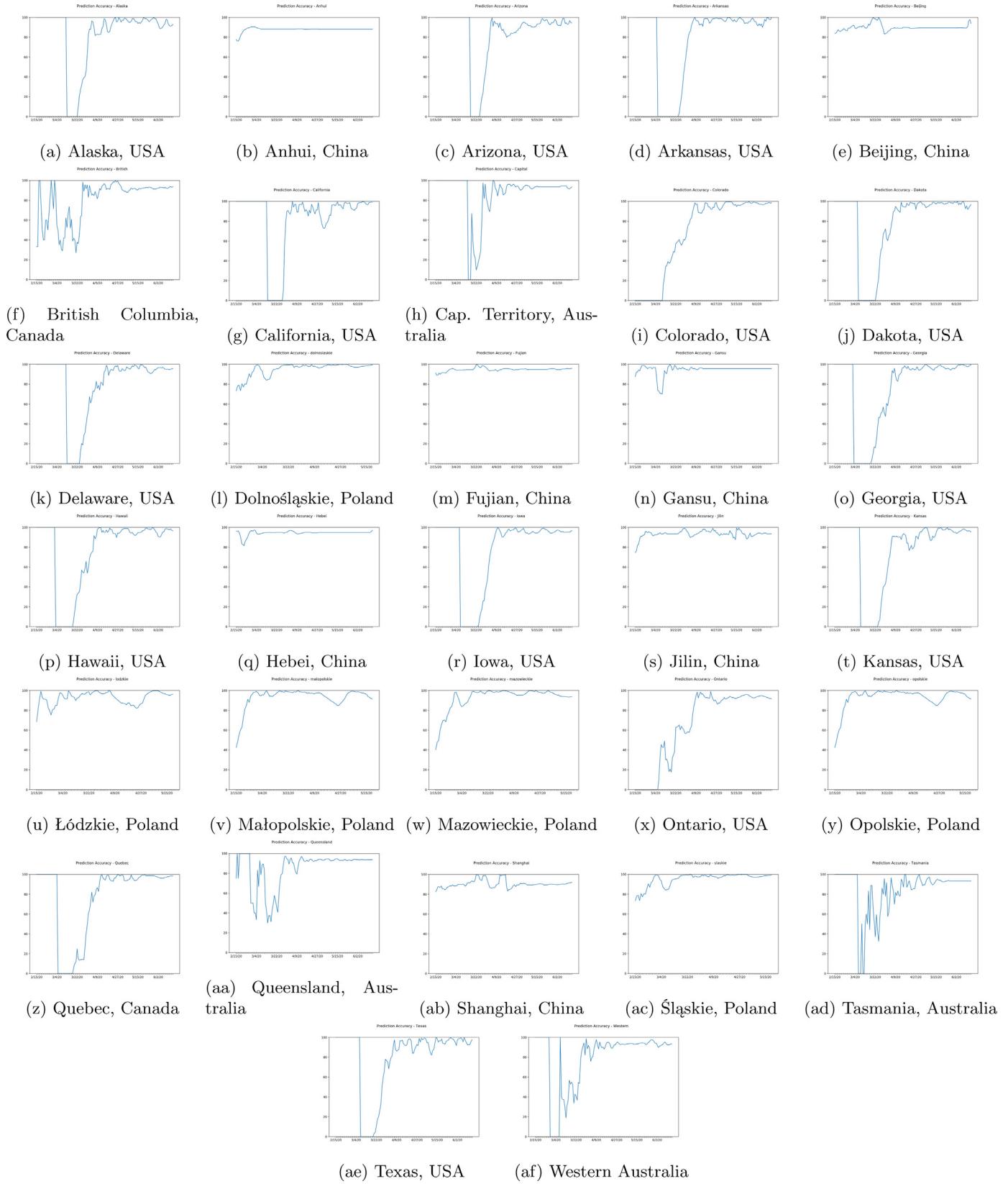


Fig. 10. Accuracy per day for selected regions.

Table 3

Sample accuracy of our model with different error margins in selected countries.

Error Margin	Afghanistan	Argentina	Belgium	Brazil	Canada	China	Congo	Dominican
0.05	80.58%	83.44%	82.78%	74.68%	85.95%	94.28%	80.76%	79.9%
0.1	81.92%	84.46%	83.47%	77.08%	88.17%	99.08%	82.4%	80.92%
0.15	82.72%	84.91%	84.15%	78.51%	89.08%	99.98%	83.64%	81.27%
0.2	83.2%	85.34%	84.9%	79.44%	89.51%	100.0%	84.43%	81.54%
0.25	83.72%	85.73%	85.71%	80.24%	89.9%	100.0%	85.03%	81.82%
0.3	84.26%	86.17%	86.6%	80.98%	90.22%	100.0%	85.4%	82.11%
0.35	84.75%	86.58%	87.57%	81.56%	90.54%	100.0%	85.67%	82.37%
0.4	85.18%	87.04%	88.65%	82.03%	90.8%	100.0%	85.98%	82.56%
0.45	85.55%	87.5%	89.64%	82.49%	91.08%	100.0%	86.31%	82.77%
0.5	85.96%	88.06%	90.44%	83.03%	91.41%	100.0%	86.56%	82.98%
Error Margin	Egypt	United Arab Emirates	France	Germany	Ghana	Greece	India	Iran
0.05	82.6%	71.18%	85.65%	74.33%	81.23%	84.51%	66.61%	83.02%
0.1	84.83%	72.65%	88.65%	75.25%	82.95%	86.68%	68.85%	84.05%
0.15	85.92%	73.65%	89.74%	76.08%	84.41%	87.58%	70.4%	84.66%
0.2	86.74%	74.63%	90.33%	76.94%	85.32%	88.0%	71.65%	85.17%
0.25	87.46%	75.62%	90.88%	77.8%	85.94%	88.42%	72.93%	85.55%
0.3	88.05%	76.68%	91.3%	78.66%	86.44%	88.83%	74.26%	85.81%
0.35	88.45%	77.8%	91.67%	79.53%	86.77%	89.25%	75.65%	86.07%
0.4	88.88%	78.94%	92.05%	80.49%	87.02%	89.72%	77.08%	86.33%
0.45	89.19%	79.96%	92.38%	81.56%	87.18%	90.26%	78.18%	86.58%
0.5	89.4%	80.75%	92.73%	82.82%	87.37%	90.84%	79.02%	86.88%
Error Margin	Iraq	Italy	Mexico	Monaco	Nepal	Netherlands	Poland	Central African Republic
0.05	83.27%	78.89%	74.47%	90.25%	66.03%	83.58%	83.92%	78.63%
0.1	85.18%	79.62%	76.31%	91.05%	67.33%	85.06%	85.35%	80.7%
0.15	86.72%	80.39%	77.64%	91.66%	68.56%	85.77%	86.1%	82.1%
0.2	88.2%	81.18%	78.83%	92.06%	69.73%	86.48%	86.77%	83.31%
0.25	89.4%	81.99%	79.76%	92.4%	70.86%	87.22%	87.3%	84.42%
0.3	90.27%	82.81%	80.49%	92.74%	71.92%	87.94%	87.8%	85.49%
0.35	90.8%	83.6%	81.07%	93.08%	72.89%	88.67%	88.27%	86.5%
0.4	91.03%	84.39%	81.66%	93.46%	73.63%	89.36%	88.72%	87.35%
0.45	91.18%	85.19%	82.11%	93.91%	74.31%	90.05%	89.14%	88.18%
0.5	91.33%	85.98%	82.4%	94.44%	74.67%	90.54%	89.52%	89.03%
Error Margin	Russia	Spain	Sudan	Sweden	Ukraine	United Kingdom	US	World
0.05	60.01%	71.0%	76.12%	83.1%	82.69%	83.55%	67.35%	98.47%
0.1	61.58%	71.99%	78.12%	84.0%	83.75%	86.59%	68.06%	99.98%
0.15	62.98%	72.63%	80.02%	84.78%	84.36%	87.66%	68.76%	100.0%
0.2	64.25%	73.28%	81.67%	85.51%	84.91%	88.36%	69.48%	100.0%
0.25	65.63%	73.99%	83.24%	86.04%	85.4%	89.11%	70.2%	100.0%
0.3	67.13%	74.76%	84.84%	86.27%	85.75%	89.89%	70.95%	100.0%
0.35	68.74%	75.56%	86.29%	86.54%	86.07%	90.62%	71.7%	100.0%
0.4	70.04%	76.43%	87.32%	86.79%	86.41%	91.34%	72.51%	100.0%
0.45	71.2%	77.1%	88.22%	87.08%	86.72%	91.86%	73.38%	100.0%
0.5	72.43%	77.75%	89.21%	87.41%	87.03%	92.36%	74.3%	100.0%

Analysis of statistical predictions and our proposed Neural Network shows differences (see Fig. 6). For our model it is necessary to have enough data from countries with COVID-19 affection to train predictor. The increment curve itself is however, much more adapted to the real data. Statistical approach gives results when in the beginning we don't have much information from other locations. Disadvantage of our model is long training time, necessary adjustment of all network parameters as well as the fact that the full slip of predicted values with sudden changes in the value of the curve is not compensated. On the other hand, as shown in our experiments, when new data is coming regularly to well defined Neural Network architecture, such composed predictor trains well for high efficiency of predictions what compensates all possible difficulties to develop it in the beginning.

In Fig. 11 we can see how the system predicts situation in the World. Such map gives a prediction model for various continents. By analyzing such presentation we can see which continents may suffer from growth in recent time. Our system also gives marks to countries which report decreasing trend so that the estimation is much useful.



Fig. 11. World map from our prediction system with marked growth values. Red color represents prediction of increasing trend, while green decreasing trend. The size of each circle represents the power of predicted change. The bigger the size the bigger growth predictor calculates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4

Sample accuracy of our model with different error margins in selected regions.

Error Margin	Alaska	Anhui	Arizona	Arkansas	Australian Capital Territory	Beijing	British Columbia	California
0.05	85.5%	93.22%	86.62%	79.78%	89.97%	94.86%	84.26%	85.09%
0.1	86.67%	98.06%	88.19%	80.34%	91.41%	99.05%	87.01%	86.44%
0.15	87.31%	99.74%	88.93%	80.62%	91.94%	99.96%	88.39%	87.03%
0.2	87.65%	99.92%	89.3%	80.84%	92.32%	100.0%	89.51%	87.37%
0.25	87.87%	100.0%	89.51%	81.04%	92.64%	100.0%	90.7%	87.58%
0.3	88.12% Figure 6	100.0%	89.76%	81.26%	92.94%	100.0%	91.91%	87.66%
0.35	88.4%	100.0%	90.01%	81.46%	93.2%	100.0%	93.19%	87.72%
0.4	88.69%	100.0%	90.22%	81.69%	93.46%	100.0%	94.42%	87.79%
0.45	89.01%	100.0%	90.44%	81.88%	93.68%	100.0%	95.58%	87.86%
0.5	89.38%	100.0%	90.69%	82.09%	93.95%	100.0%	96.8%	87.87%
Error Margin	Colorado	Dakota	Delaware	Dolnośląskie	Fujian	Gansu	Georgia	Hawaii
0.05	69.75%	81.25%	83.14%	98.07%	99.4%	98.6%	78.56%	79.39%
0.1	70.86%	81.96%	84.06%	98.81%	99.99%	99.05%	79.24%	80.22%
0.15	71.66%	82.41%	84.72%	99.43%	100.0%	99.34%	79.77%	80.71%
0.2	72.45%	82.89%	85.39%	99.81%	100.0%	99.57%	80.25%	81.25%
0.25	73.25%	83.39%	85.96%	99.96%	100.0%	99.82%	80.74%	81.85%
0.3	74.07%	83.88%	86.37%	100.0%	100.0%	100.0%	81.3%	82.4%
0.35	74.98%	84.3%	86.74%	100.0%	100.0%	100.0%	81.88%	82.94%
0.4	75.94%	84.59%	87.06%	100.0%	100.0%	100.0%	82.48%	83.5%
0.45	76.62%	84.83%	87.37%	100.0%	100.0%	100.0%	83.04%	84.02%
0.5	77.23%	85.05%	87.68%	100.0%	100.0%	100.0%	83.58%	84.35%
Error Margin	Hebei	Iowa	Jilin	Kansas	Łódzkie	Małopolskie	Mazowieckie	Ontario
0.05	99.2%	80.27%	98.02%	79.66%	96.97%	96.86%	96.33%	68.71%
0.1	99.78%	80.8%	99.61%	81.0%	98.52%	97.82%	97.15%	70.55%
0.15	99.95%	81.13%	99.83%	81.71%	99.47%	98.33%	97.76%	71.48%
0.2	100.0%	81.41%	99.94%	82.26%	99.83%	98.58%	98.19%	72.38%
0.25	100.0%	81.68%	100.0%	82.62%	99.93%	98.82%	98.58%	73.37%
0.3	100.0%	81.96%	100.0%	82.94%	99.98%	99.05%	98.95%	74.45%
0.35	100.0%	82.21%	100.0%	83.24%	100.0%	99.3%	99.2%	75.7%
0.4	100.0%	82.5%	100.0%	83.55%	100.0%	99.53%	99.41%	76.79%
0.45	100.0%	82.77%	100.0%	83.87%	100.0%	99.71%	99.59%	77.55%
0.5	100.0%	83.09%	100.0%	84.22%	100.0%	99.84%	99.79%	78.26%
Error Margin	Opolskie	Quebec	Queensland	Shanghai	Śląskie	Tasmania	Texas	Western Australia
0.05	96.86%	78.31%	88.91%	95.17%	98.07%	87.69%	82.18%	83.03%
0.1	97.82%	78.94%	90.87%	99.05%	98.81%	90.01%	83.39%	84.93%
0.15	98.33%	79.47%	91.86%	99.96%	99.43%	91.4%	84.15%	85.75%
0.2	98.58%	79.93%	92.75%	100.0%	99.81%	92.45%	84.68%	86.47%
0.25	98.82%	80.31%	93.53%	100.0%	99.96%	93.29%	85.07%	87.13%
0.3	99.05%	80.68%	94.23%	100.0%	100.0%	94.09%	85.36%	87.85%
0.35	99.3%	81.02%	95.01%	100.0%	100.0%	94.84%	85.63%	88.64%
0.4	99.53%	81.36%	95.89%	100.0%	100.0%	95.59%	85.87%	89.53%
0.45	99.71%	81.67%	96.84%	100.0%	100.0%	96.23%	86.08%	90.52%
0.5	99.84%	82.05%	97.88%	100.0%	100.0%	96.84%	86.33%	91.32%

5. Conclusions

In our article we have shown both, advantages and disadvantages of solutions related to prediction of cases for infected with the COVID-19 virus. We have presented why, according to us, such models benefit from the use of neural network. One should also take into account many factors influencing the shape of the curve of the increase in infections, among others: behavior of the population in a given region, behavior of governments of given countries as well as access to knowledge and medical equipment.

The network we have developed has unified architecture what makes it easy to work as predictor, since we don't need to change it for countries or regions. Our predictor achieves very high accuracy for most of regions which is around 87.70%. We can conclude that by using individual predictors for each possible region or country, the predictions could be increased. Our research also show the problem of insufficient data at the beginning of the situation. with increasing amount od data from governments our predictor was gaining on accuracy during following days. In future works, we want to expand precision with devoted predictors and increase network efficiency when there is a small amount of data available by using statistical models as some compensation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Bárcena M, Oostergetel GT, Bartelink W, Faas FG, Verkleij A, Rottier PJ, et al. Cryo-electron tomography of mouse hepatitis virus: insights into the structure of the coronavirion. *Proc Natl Acad Sci* 2009;106(2):582–7.
- [2] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579(7798):270–3.
- [3] Casella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, evaluation and treatment coronavirus (COVID-19). *StatPearls [internet]*. StatPearls Publishing; 2020.
- [4] Li L-Q, Huang T, Wang Y-Q, Wang Z-P, Liang Y, Huang T-B, et al. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. *J Med Virol* 2020;92(6):577–83.
- [5] Ardabili S.F., Mosavi A., Ghamisi P., Ferdinand F., Varkonyi-Koczy A.R., Reuter U., et al. COVID-19 outbreak prediction with machine learning. Available at SSRN 35801882020..
- [6] Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model* 2020.

- [7] Ivorra B, Ferrández MR, Vela-Pérez M, Ramos A. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun Nonlinear Sci Numer Simul* 2020;88.
- [8] Biswas K, Khaleque A, Sen P. COVID-19 spread: reproduction of data and prediction using a sir model on euclidean network. *arXiv:200307063* 2020;.
- [9] Kuniya T. Prediction of the epidemic peak of coronavirus disease in Japan, 2020. *J Clin Med* 2020;9(3):789.
- [10] Ciufolini I, Paolozzi A. Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a gauss error function and Monte Carlo simulations. *Eur Phys J Plus* 2020;135(4):355.
- [11] He S, Tang S, Rong L. A discrete stochastic model of the COVID-19 outbreak: forecast and control. *Math Biosci Eng* 2020;17:2792–804.
- [12] Pandey G, Chaudhary P, Gupta R, Pal S.. SEIR and regression model based COVID-19 outbreak predictions in India. *arXiv:200400958* 2020;.
- [13] Hu Z, Ge Q, Jin L, Xiong M.. Artificial intelligence forecasting of COVID-19 in China. *arXiv:200207112* 2020;.
- [14] Jewell NP, Lewnard JA, Jewell BL. Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *JAMA* 2020;323(19):1893–4.
- [15] Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar T. A model based study on the dynamics of COVID-19: prediction and control. *Chaos Solitons Fractals* 2020;109889.
- [16] Qin L, Sun Q, Wang Y, Wu K-F, Chen M, Shia B-C, et al. Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. *Int J Environ Res Public Health* 2020;17(7):2365.
- [17] Wangping J, Ke H, Yang S, Wenzhe C, Shengshu W, Shanshan Y, et al. Extended sir prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Front Med* 2020;7:169.
- [18] Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the diamond princess cruise ship: a data-driven analysis. *Int J Infect Dis* 2020;93:201–4.
- [19] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals* 2020;134:109761.
- [20] Yang Z, Zeng Z, Wang K, Wong S-S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020;12(3):165.
- [21] Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS ONE* 2020;15(3):e0231236.