PROCEEDINGS OF SCIENCE

Accuracy of COVID-19 evolution models for different forecast horizons

Saveliy Zavertyaev,[1,2,*] **Ivan Moloshnikov,**[1] **Alexander Naumov,**[1] **Roman Rybka**[1,3] **and Alexander Sboev**[1,3,4]

[1] *NRC "Kurchatov Institute",*
Akademika Kurchatova sq., 1, Moscow, Russian Federation

[2] *Moscow Institute of Physics and Technology "MIPT",*
Institutskiy Pereulok, 9, Dolgoprudny, Moscow Oblast, 141701, Russian Federation

[3] *Russian Technological University "MIREA",*
Vernadsky av., 78, Moscow, 119454, Russian Federation

[4] *NRNU "MEPhi",*
Kashira Hwy, 31, Moscow, Russian Federation

E-mail: sag111@mail.ru

Currently, the statistics on COVID-19 for many regions are accumulated for the time span of over than two years, which facilitates the use of data-driven algorithms, such as neural networks, for prediction of the disease's further development. This article provides a comparative analysis of various forecasting models of COVID-19 dynamics. The forecasting is performed for the period from 07/20/2020 to 05/05/2022 using statistical data from the regions of the Russian Federation and the USA. The forecast target is defined as the sum of confirmed cases over the forecast horizon. Models based on the Exponential Smoothing (ES) method and deep learning methods based on Long Short-Term Memory (LSTM) units were considered. The training data set included the data from all regions available in the full data set. The MAPE metric was used for model comparison, the evaluation of the effectiveness of LSTM in the learning process was carried out using cross-validation on the mean squared error (MSE) metric. The comparisons were made with the models from various literature sources, as well as with the baseline model "tomorrow as today" (for which the sum of cases over the forecast horizon is supposed to be equal to the current case number multiplied by the forecast horizon length). It was shown that on small horizons (up to 28 days) the "tomorrow as today" model and ES algorithms show better accuracy than LSTM. In turn, on longer horizons (28 days or more), the preference should be given to the more complex LSTM-based model.

**** The 6th International Workshop on Deep Learning in Computational Physics (DLCP2022) ****
**** 6-8 July 2022 ****
**** JINR, Dubna, Russia ****

*Speaker

PoS(DLCP2022)025

Introduction Since the beginning of the coronavirus pandemic, several predictive models have been proposed to determine the number of new cases [1]. Along with generally accepted approaches motivated by modeling of the spread of the virus, such as SIR [2], SEIR [3] and their modifications [4, 5], various methods for analyzing one-dimensional data were applied to forecasting, such as exponential smoothing [6], ARIMA [7] and others. A number of works demonstrated the effectiveness of neural network models based on Long Short-Term Memory layers (LSTM) [1, 6, 8]. A recent study [9] showed that the effectiveness of the existing methods applied to the COVID-19 dynamics forecasting is comparable to the estimates based on the "tomorrow as today" method, which indicates the complexity of the task and the need for further improvement of predictive methods. In our previous works [5, 7] it was shown that: 1) the accuracy of machine learning models strongly depends on the number of training samples, 2) the usage of the models of this type is inefficient on short prediction horizons (up to 28 days). Due to the fact that retrospective data for 2 years of the development of the pandemic are currently available, including details for individual regions and countries, this work is aimed at creating a model that takes into account the newly-accumulated data sets. Here we evaluate our best-performing models from the last year [5, 7] on this year's data. In an ongoing work, we compare the "tomorrow as today" (Dummy) baseline model, Exponential Smoothing (ES), and LSTM-based neural networks (NN) on the new data and show that LSTM with the current data amount can outperform the baseline and ES in accuracy. The paper explores the possibility of improving the accuracy of a neural network based on LSTM by the use of combined training samples, including data from different countries (in this work, the USA and the Russian Federation), and a representation of input samples divided into even and odd days. The Data section describes how the data was processed, how the models were trained and tested. The Methods section describes which models and hyperparameters were used. The Experiments section shows the results obtained. The Discussion section is dedicated to the conclusions that can be drawn from the experimental results. The Conclusion section gathers the discussion and experiments into a short summary. Supplementary Materials contain models' accuracies grouped by folds and countries.

1. Data

In this work, historical data on confirmed cases of disease, death, and recovery published in the JHU [10] and [11] projects was used. For the USA and Russia those projects contain detailed information (daily changes in the cumulative values of confirmed cases, recovered patients and deaths) about individual regions (51 states of America and 86 constituent entities of Russia), motivating the choice of those countries for the conducted experiments. Only the data on confirmed cases (Confirmed) were used in the work. The data set was pre-processed in the following way:

- A series of daily values was created (Confirmed_daily)(Daily difference of the cumulative Confirmed cases);
- The values were normalized per 100 thousand population for the series of daily (Confirmed_daily) and cumulative (Confirmed) confirmed cases of the disease (as defined in Eq. 1).

$$V_{\text{norm}} = \frac{V \cdot 100000}{P} \tag{1}$$

Here V is the value of the current day, P is the population of the region according to the data from the [11] web-site for Russia and the data from the web-site [12] for the USA.

We used accumulative cross validation to run the experiments and to compare different models. To do this, the entire data set was divided into 6 time intervals, with the subsequent formation of 5 pairs of training and test parts (folds) from them. The splitting was done in a data-accumulating manner in the training part, as shown in Figure 1. The dates of splitting into subsets for the formation of folds were fixed for all the considered regions of the countries: July 22, 2020; December 2, 2020; April 14, 2021; August 25, 2021; January 5, 2022.

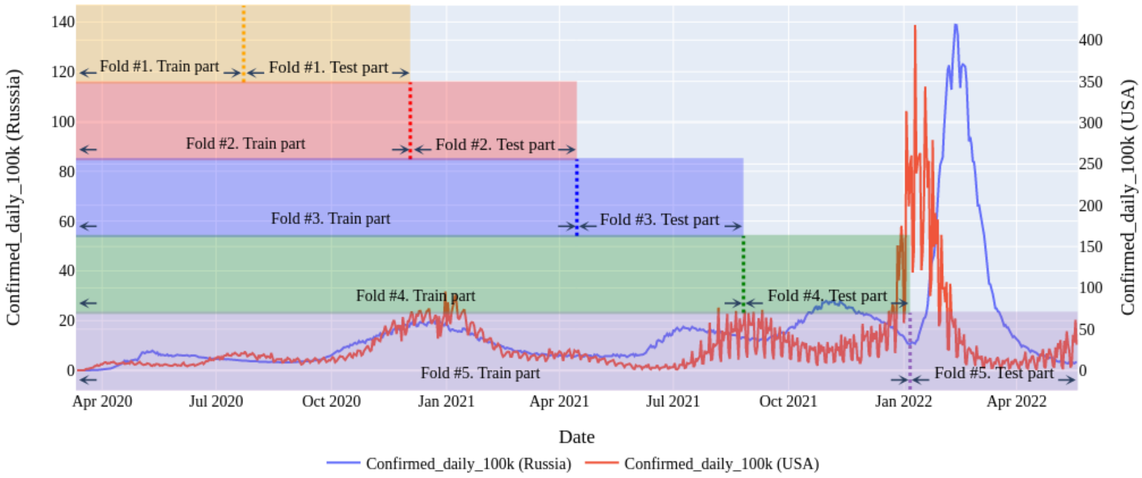

Figure 1: An example of splitting the data into folds with accumulation of the training part. The data and the dates are presented for Russia and the USA.

2. Target Series

We considered new cases for the selected period (FH - Forecast Horizon) as the prediction target. 3 FH were considered: 14, 28 and 42 days.

Since the number of examples for different horizons differs from the number of total cumulative values (there are fewer examples for FH days, see Fig. 2 and Eq. 2) and fold dates are fixed, the number of examples in the test part of the last fold was different for different FH.

$$\text{TargetValue} = \text{Confirmed}(\text{t}_3) - \text{Confirmed}(t_2), \quad t_3 = t_2 + \text{FH} \tag{2}$$

3. Methods

3.1 Dummy models

In the current work, the "tomorrow as today" (Dummy) model was chosen as the baseline. This model takes the last day from a known sample and multiplies it by the duration of the Forecast Horizon.

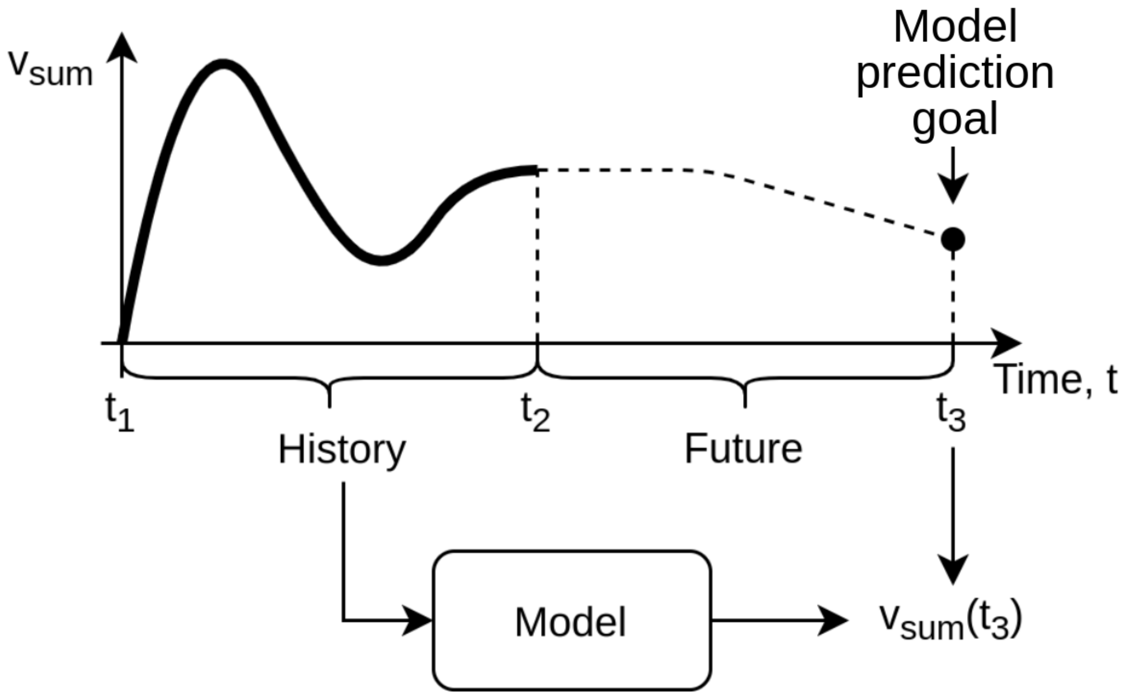

Figure 2: Splitting the data into History and Future parts

3.2 Statistical approach

The Holt-Winters exponential smoothing method from [13] with damped additive trend, box-cox normalization and no seasonality was used, which proved itself efficient on the data from the first year of the pandemic [5, 7]. The output of the prediction consisted of the new daily cases for the FH days ahead. Confirmed_daily values from the training part were used as input, and the number of new cases for the FH was used as the prediction target. The predicted values were subsequently summed up (this process is referred to further as Exponential Smoothing (ES));

3.3 Neural Network models

In this paper, 2 models based on LSTM cells were considered, see Fig. 3 and Fig. 4. The historical data (daily values of Confirmed_daily_100k series) for 28 days was used as the models' input, while the output corresponded to the sum of cases over the FH.

The main difference between the two neural networks considered in this paper is in the input data received: NN receives historical data for 28 days, while NN_odd additionally splits it into even and odd days and learns from all three series separately and independently.

Each of these models also had 3 versions (USA, Ru, USA + Ru), depending on the data (regions) of which country it was trained and tested on: separately on the data from the USA, separately on the data from the Russian Federation, or using the information from both countries.

Model hyperparameters were optimized using the [14] to minimize Mean absolute percentage error (MAPE, see Eq. 3) of the 2nd fold for the Moscow region. Both models had 8 neurons with hyperbolic tangent (tanh) activation function in the fully-connected layer, batch size equal to 200, learning rate equal to 0.003 and Adam optimizer. An LSTM layer with 2 neurons was used for the

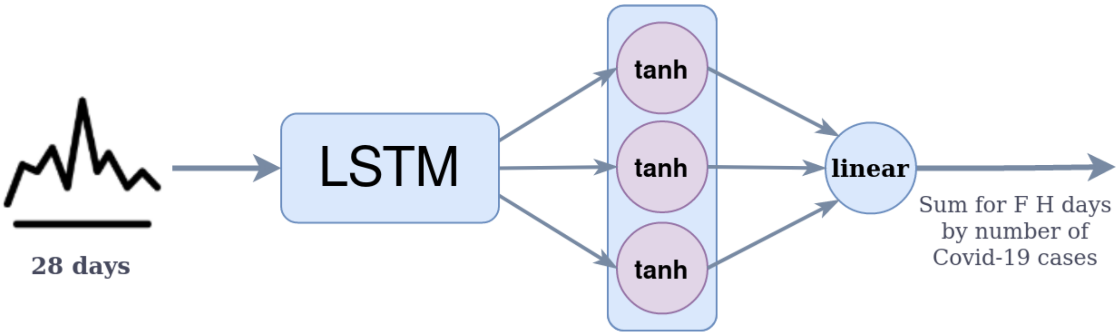

Figure 3: A neural network designed to forecast the sum of confirmed Covid-19 cases in FH (NN)

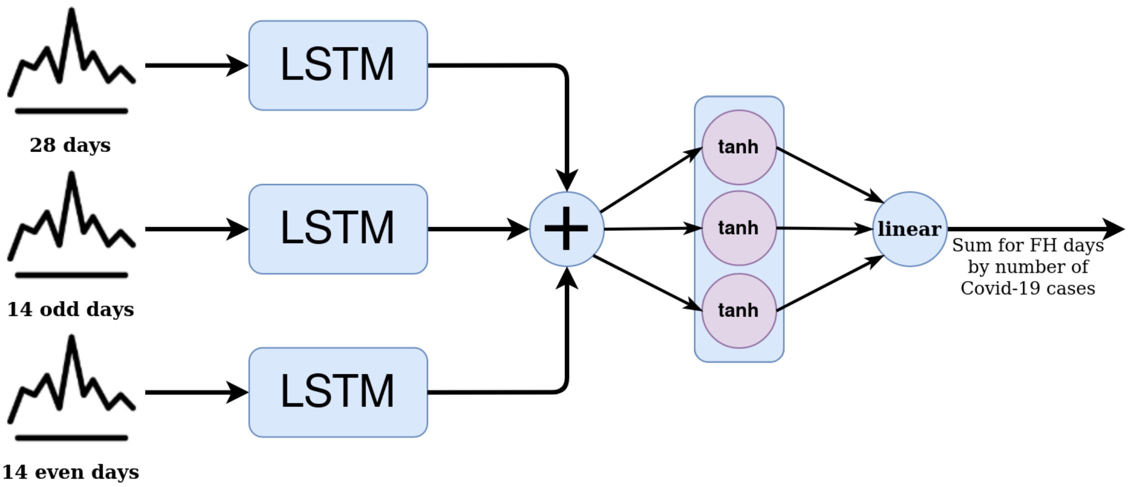

Figure 4: A neural network with even/odd input splitting (NN_odd)

NN model. Regarding the NN_odd model, an LSTM layer with 4 neurons was utilized to process the complete history, with 2 neurons for even days and 2 for odd ones. The dropout rate was set to 0.01.

4. Experiments

The described approaches were used to determine the level of accuracy on different Forecast Horizons. As mentioned above, MAPE was used as the target metric:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|, \tag{3}$$

Here n is the number of days in the testing part, A_t and F_t are the true and predicted values of the time series for the day t (the total increase in the number of cases from the prediction date to the forecast date was used). Tables 3, 4 and 5, 6 present the results for the USA and Russia averaged over folds and horizons. The values in the table were obtained by averaging the country-wide MAPE scores for each individual region.

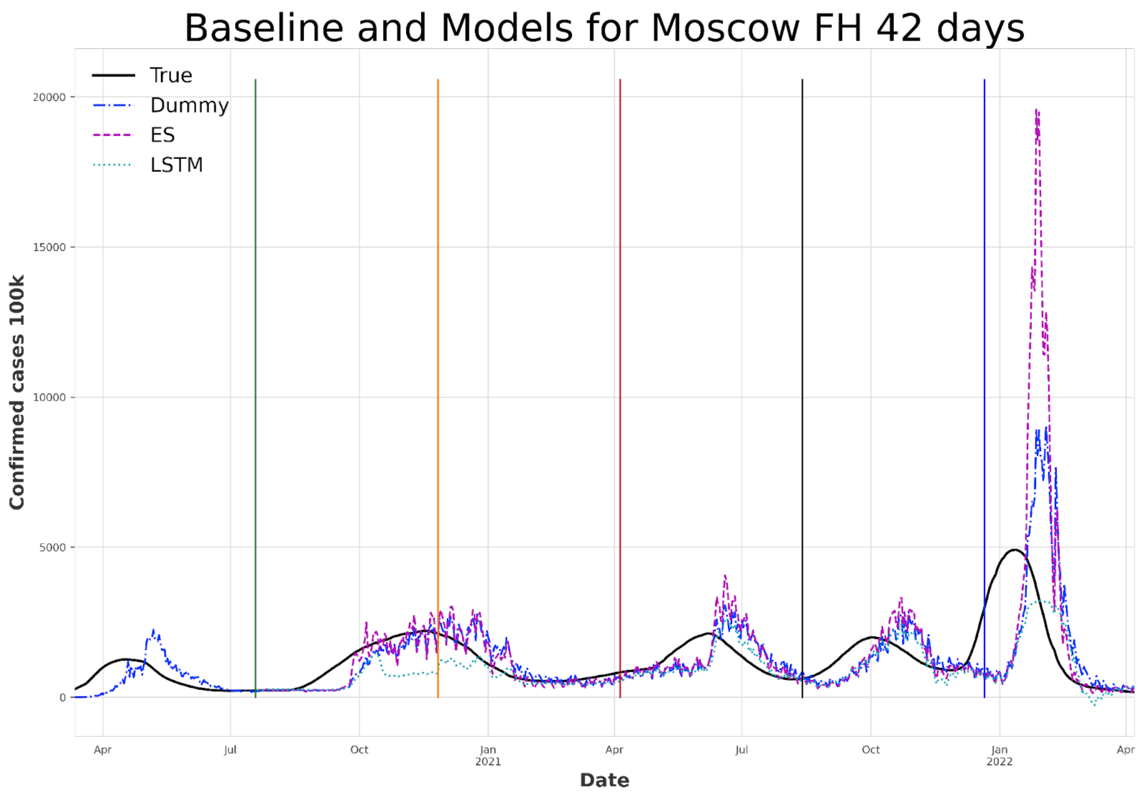

Figure 5: A prediction example for different models for Moscow region (vertical lines are fold borders)

Table 1: MAPE with averaging over folds and all regions of the USA and Russia excluding countries as a whole.

Country	Dummy	ES	NN (one country)	NN (USA+Rus)	NN_odd (one country)	NN_odd (USA+Rus)
Average (Russia)	32,9	32,7	32,6	38,4	32,7	41,2
Average (USA)	73,1	53,3	63,2	48,6	49,7	45,5
Average (both)	48	40,4	44	42,2	39	42,8

5. Discussion

The proposed neural network model and the ES method demonstrate comparable accuracies on average for all prediction horizons for Russian regions: 32.6 and 32.7 MAPE (see Table 1). Considering prediction horizons with lengths equal to 28 and 42 days, the neural network model trained on the regions of Russia with splitting into even/odd days (NN_odd(Rus)) demonstrates accuracies of 30.4 and 40.7. These results outperform ES by 1.5 and 6.9 MAPE and Dummy by 2.6 and 5.2 MAPE, respectively (see Table 6). The analysis of Tables 4 and 3 (see Supplementary Material) for the USA shows that splitting the data into even/odd days and combining the data from both countries in the training data improves the accuracy compared to Dummy by 17.7 MAPE. The

Table 2: Comparison with the literature sources [15, 16].

Model	2020 year (Month-Day)						2021 year (Month-Day)		
	11-15	11-22	11-29	12-06	12-13	12-20	12-27	01-03	01-10
Google SEIR	27,9	43,5	43,1	26,6	32,9	38	28,6	31,8	93,9
NN (USA)	48,1	43,3	42	29,3	30,6	39,2	51,2	25,9	24,8
NN (USA+Rus)	53,1	54,2	46,8	32,7	33,8	43,9	56,5	26,7	26,3
NN_odd (USA)	49,9	48,2	33,5	28,2	27,3	31,6	48,8	32,8	14,8
NN_odd (USA+Rus)	50,2	46	27,5	24,3	20,8	19,9	34,9	22,8	21,6
CovidHub baseline	29,8	29,8	31,2	32	32	20,8	16,6	17,8	62,3

resulting accuracy is 45.5 MAPE, which is higher than ES and Dummy by 7.8 and 27.6 MAPE, respectively. On prediction horizons of 28 and 42 days, the difference from ES is equal to 9.5 and 6.5 MAPE, and the difference from Dummy is equal to 28.6 and 31.3 MAPE in favor of the neural network model NN_odd(USA+Rus) (see Table 4).

Neural networks and LSTM-based neural networks in particular require large datasets to demonstrate good performance. The effect of the dataset size on the models' performance can be seen in Tables 3, 5 from Supplementary Material section (on Fold#0).

Additionally, a comparison to publicly-available accuracies for the USA [15, 16] was carried out. For comparison the following two models were chosen:

- Google-SEIR(Suspended-Exposed-Infected-Recovered)-like population dynamics model [15] that includes such factors as mobility, census, healthcare supply etc.
- COVIDhub-baseline: Extrapolation with the condition that the median prediction at all future horizons is equal to the most recently observed incidence.

The obtained results (see Table 2) demonstrate that the proposed NN_odd(USA+Rus) model is more effective than the Google-SEIR model (by 10.9 MAPE) and is comparable to the COVIDhub-baseline: 29.8 vs 30.3 MAPE.

6. Conclusion

In this paper, the feasibility of combining data at the regional level into a single training set that is used to train a predictive neural network model for determining the number of cases was investigated. The conducted study presents the results of comparative analysis of the proposed model with regards to a number of previously-proposed approaches and the baseline model "tomorrow as today". It is shown that the proposed neural network outperforms the baseline methods (Exponential Smoothing and "tomorrow as today") by 1.6-9.5 MAPE on the task of forecasting the dynamics of the COVID-19 disease for 28 and 42 days. In future, the proposed approach will be tested on the extended dataset containing more countries.

7. Acknowledgements

This study was supported by the Russian Foundation for Basic Research project №20-04-60528 and carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC “Kurchatov Institute”, http://ckp.nrcki.ru/.

8. Supplementary Material

Table 3: Average MAPE by folds (averaged across all US states)

Fold#	Dummy	ES	NN (USA)	NN (USA+Rus)	NN_odd (USA)	NN_odd (USA+Rus)
0	38,6	34,4	40,6	38,8	48,3	40,5
1	47,6	40,2	27,2	25,4	36,8	23,9
2	68,2	58,5	133,8	48,5	56,2	52,2
3	68,6	60,9	30,6	31,2	40,5	30,9
4	142,3	72,6	83,6	99,1	66,6	80,1
Av.	73,1	53,3	63,2	48,6	49,7	45,5

Table 4: Average MAPE by horizons (averaged over all US states)

Horizon (days)	Dummy	ES	NN (USA)	NN (USA+Rus)	NN_odd (USA)	NN_odd (USA+Rus)
14	60,1	44,9	37,1	40,5	35,2	37,5
28	72,1	53	49,8	47,3	46,5	43,5
42	86,9	62,1	102,5	58,1	67,3	55,6

Table 5: Average MAPE by folds (averaged over all regions of the Russian Federation)

Fold#	Dummy	ES	NN (Rus)	NN (USA+Rus)	NN_odd (Rus)	NN_odd (USA+Rus)
0	22,1	22,4	30,5	39,3	30,8	41,4
1	22,9	20,4	34,9	42,1	32,1	48,3
2	20,7	21,4	20,8	23,7	21,7	34,3
3	20,8	20,7	15,7	18,1	15,5	18,1
4	78	78,3	60,8	68,7	63,3	63,9
Av.	32,9	32,7	32,6	38,4	32,7	41,2

Table 6: Average MAPE by horizons (averaged over all regions of the Russian Federation)

Horizon (days)	Dummy	ES	NN (Rus)	NN (USA+Rus)	NN_odd (Rus)	NN_odd (USA+Rus)
14	19,8	18,5	24,6	26,7	26,9	32,1
28	33	31,9	31,6	38	30,4	38,5
42	45,9	47,6	41,5	50,4	40,7	53

References

[1] S. Prajapati, A. Swaraj, R. Lalwani, A. Narwal, K. Verma, G. Singh et al., *Comparison of traditional and hybrid time series models for forecasting covid-19 cases*, *arXiv preprint arXiv:2105.03266* (2021) .

[2] I. Cooper, A. Mondal and C.G. Antonopoulos, *A SIR model assumption for the spread of COVID-19 in different communities*, *Chaos, Solitons & Fractals* **139** (2020) 110057.

[3] S. He, Y. Peng and K. Sun, *Seir modeling of the covid-19 and its dynamics*, *Nonlinear dynamics* **101** (2020) 1667.

[4] N.A. Kudryashov, M.A. Chmykhov and M. Vigdorowitsch, *Comparison of some covid-19 data with solutions of the sir-model*, in *AIP Conference Proceedings*, vol. 2425, p. 340009, AIP Publishing LLC, 2022.

[5] A. Sboev, N. Kudryshov, I. Moloshnikov, A. Serenko, A. Naumov and R. Rybka, *Baseline accuracy of forecasting covid-19 cases in moscow region on a year in retrospect using basic statistical and machine learning methods*, in *Journal of Physics: Conference Series*, vol. 2036, p. 012029, IOP Publishing, 2021.

[6] A.G. Sboev, N.A. Kudryshov, I.A. Moloshnikov, S.V. Zavertyaev, A.V. Naumov and R.B. Rybka, *Strict baselines for covid-19 forecasting and ml perspective for usa and russia*, *arXiv preprint arXiv:2207.07689* (2022) .

[7] A. Naumov, I. Moloshnikov, A. Serenko, A. Sboev and R. Rybka, *Baseline accuracies of forecasting covid-19 cases in russian regions on a year in retrospect using basic statistical and machine learning methods*, *Procedia Computer Science* **193** (2021) 276.

[8] K. ArunKumar, D.V. Kalaga, C.M.S. Kumar, M. Kawaji and T.M. Brenza, *Comparative analysis of gated recurrent units (gru), long short-term memory (lstm) cells, autoregressive integrated moving average (arima), seasonal autoregressive integrated moving average (sarima) for forecasting covid-19 trends*, *Alexandria Engineering Journal* **61** (2022) 7585.

[9] A. Chharia, G. Jeevan, R.A. Jha, M. Liu, J.M. Berman and C.A. Glorioso, *Accuracy of us cdc covid-19 forecasting models*, *medRxiv* (2022) .

[10] E. Dong, H. Du and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time*, *The Lancet infectious diseases* **20** (2020) 533.

PoS(DLCP2022)025

[11] "Yandex datalens." [Electronic resource]. https://datalens.yandex/covid19, accessed 05.08.2022.

[12] "American community survey." [Electronic resource]. https://www.census.gov/programs-surveys/acs, accessed 05.08.2022.

[13] "Statsmodels library." [Electronic resource]. https://www.statsmodels.org/stable, accessed 05.08.2022.

[14] "Ray tune library." [Electronic resource]. https://www.ray.io/ray-tune, accessed 05.08.2022.

[15] S.Ö. Arık, J. Shor, R. Sinha, J. Yoon, J.R. Ledsam, L.T. Le et al., *A prospective evaluation of ai-augmented epidemiology to forecast covid-19 in the usa and japan*, *NPJ digital medicine* **4** (2021) 1.

[16] "European covid-19 forecast hub." [Electronic resource]. https://covid19forecasthub.eu, accessed 05.08.2022.

PoS(DLCP2022)025