

Approaches for forecasting of socioeconomic impacts to the spread of COVID-19 with territorial differences of Russian regions

Dmitry Devyatkin^{1*}, Yulia Otmakhova², and Natalia Usenko³

¹Federal Research Center “Computer Science and Control” RAS, 119333, Moscow, Russia

²Central Economic and Mathematics Institute RAS, 117418 Moscow, Russia

³Technologies for Systems Analysis LLC, 117312 Moscow, Russia

Abstract. The COVID-19 pandemic has brought severe demographical, socioeconomic, and territorial impacts. Those challenges require the world community to develop both response measures and anticipation of new threats. Therefore, creating the modern tools to forecast various indicators of the impact intensity pandemic becomes important and relevant for consideration and evaluation of interregional differences. This paper presents deep neural network models to predict a viral pandemic's effects in the regional cluster of Moscow and its neighbors. They are based on recurrent and Transformer-like architectures and utilize the attention mechanism to consider the features of the neighbor regions and dependencies between various indicators. These models are trained on heterogeneous data, including daily cases and deaths, the diseased age structure, transport, and hospital availability of the regions. The experimental evaluation shows that the demographic and healthcare features can significantly improve the accuracy of economic impact prediction. We also revealed that the neighboring regions' data helps predict the outburst's healthcare and economic impact. Namely, that data helps to improve accuracy for both the number of infected and the unemployment rate. The impact forecasting would help to develop strategies to reduce inter-territorial inequality due to the pandemic.

Keywords: COVID-19 pandemic, socioeconomic and interterritorial impacts, recurrent neural network, Long-Short Term Memory, forecasting, region

1 Introduction

The COVID-19 pandemic has become an unprecedented challenge for the modern world and caused severe demographic, economic, and social consequences: the fall of the world economy, the decline in the quality of life, and the well-being of the world's population [1, 2, 3]. The COVID-19 pandemic has had a significant impact on the Russian economy: many companies were forced to suspend or close temporarily; there was a decrease in

* Corresponding author: devyatkin@isa.ru

aggregate demand; the share of the unemployed has increased; the winners turned out to be those that have been able to satisfy the rush demand of the population for food and essential goods, as well as protective equipment. Sectors and segments of the economy that take advantage of modern IT technologies are distinguished by flexibility, mobility, and creativity received a significant impetus.

The effects of a pandemic can be compared in quality and scale with threats to national and international security and require the world community to develop a countermeasures strategy, including reducing socioeconomic effects [4]. This reduction is becoming a primary economic problem, and large countries with a multi-regional economy have to consider interregional differences while solving it. It becomes crucial to forming a strategy to eliminate regional imbalances, as all aspects of the functioning of each region change. The head of the OECD [5] emphasizes that pandemic has a strong regional impact, examines the territorial effect of the crisis in its various dimensions: health, economic, social, and fiscal. The report proposes an ideology of multi-level governance and points for policymakers to consider as they build more resilient regions. This politic could help territories recover and adapt to the impact of economic, financial, environmental, political, and social shocks. On examples of responses by national and subnational governments, the authors of the report offer takeaways on managing COVID-19's territorial impact.

The destabilization of the economic situation in the country's regions due to the epidemic makes it necessary to solve the problem of predicting the socio-economic consequences at the regional level, considering territorial differences. The paper [6] presents the spatial aspect analysis of the distribution of pandemic and its effects using modern methods of intelligent analysis of geospatial information, which can reveal the spatial and temporal features of the infection spread and identify cluster areas with a high risk. Therefore, we conclude that all kinds of features should be considered together, and the deep neural network framework suits that well.

Following these ideas, in this paper, we set the research questions below.

- Which neural network models are helpful to predict the economic impact of the pandemic (unemployment rate)?
- Can interregional dependences improve the healthcare and economic impact prediction?

The paper is organized as follows. Section 1 provides a short review of state-of-the-art coronavirus impact prediction. Section 2 contains the description of all the data sources we integrate to train the models. Section 3 presents the forecasting models; namely, they describe the Non-Linear Heterogeneous Autoregressive Model (NARX), recurrent network and Transformer models with region-level attentions. Section 4 provides the experimental results on the historical data from March 2020 to March 2021. The last section concludes that the demographic and healthcare indicators can improve economic impact prediction accuracy, but not vice-versa. We have also revealed that the neighboring regions often influence the healthcare and economic impact.

2 Related work

There are plenty of studies considering the COVID19 forecasting, although they often focus only on the demographical impact. Besides, the accuracy of the obtained predictions has room for improvements [7]. When restrictions are based on poor forecasts, the economy and social harm can overweight the positive outcomes. Ioannidis, with colleagues, claims that several principles should be used to achieve better accuracy. Namely, they suggest focusing more on modeling distributions rather than point estimates, considering multiple dimensions of impact, and continuously reappraising models based on their validated performance. Simple approaches like linear regression or Holt and Winter's models miss

inter-dependences of different features and effects [8]. On the other hand, most of the existing research is devoted to forecast brief periods. For example, in [9], the researchers investigate the impact of the pandemic on the financial movements of stock indexes. The method integrates the stationary wavelet transform and bidirectional long short-term memory recurrent networks. Firstly, they apply the stationary wavelet transform to decompose the data into approximation and detail coefficients. After the decomposition, data of stock market indexes along with COVID-19 confirmed cases were considered as inputs to predict future price movement. The experiments show the method achieved fair results in terms of a five-day Crude Oil price forecast.

Paper [10] shows that different region features affect the spread of coronavirus for different periods. Among these features, there is a high population density in cities, proximity to the largest megacities, a higher proportion of the most active and frequently traveling part of the population, and intensive connections within the community with other regions and countries. Study [11] presents the main research results on the social aspect of the pandemic in the context of social heterogeneity and differences in regions. The paper emphasizes that spatial, social transformations inevitably become a condition for the effectiveness of many social constraints. Paper [12] examines the territorial aspects of the pandemic impact on economic development and industrial production, including budget revenues and expenditures of the regions.

In this work, we consider several deep neural network models, which can process heterogeneous inputs simultaneously, and, at the same time, catch long-term dependencies between steps of the analyzed time series. Those models can also be modified with a quantile loss to provide interval estimates instead of point ones.

3 Features and Data

We combine several data sources to predict the impact of the coronavirus pandemic. The list below contains aliases for the data sources that are used in the results.

- Summary data of regional headquarters for monitoring the situation with coronavirus from mid-March 2020 to March 2021 for particular Russia's regions. The data includes per-day values for the following indicators: number of infected, number of deaths, number of recovered.
- Detail data of the Moscow coronavirus headquarter. It contains per-day values (March 2020 to January 2021) for the number of infected with different ages: before 18, 18-45, 46-65, 66-79, > 80.
- The normalized score of transport availability of Russian regions obtained with data and methodology provided in [13].
- The unemployment rate for Russian regions in 2020-2021 (per-quarter data, as provided by Russian Federal State Statistics Service).
- Hospital bed availability for different regions in 2016 (as provided by Federal State Statistics Service).

The dataset can be downloaded from [14].

4 Models for Forecasting of Socioeconomic Impacts

We consider the impact forecasting as a time-series analysis problem. We have tested three different models to deal with the problem. The NARX-based model [15] is a feed-forward network (Fig. 1a), which has linear and non-linear layers and predicts the future values for the outputs $Y(t, t+1, \dots)$ based on the inputs X and output values Y from the previous time-steps $(t-1, \dots, t-k)$. The proposed model has several outputs: number of infected Y_t , number of

deaths Y_D , and unemployed Y_{UE} (all are the daily cases). As far as we design the model to predict the impacts three months ahead, the model returns the outputs for the whole forecast period at once to avoid error accumulation. We have combined and tested various sets of features from the dataset to form the inputs X . Population mobility leads to spreading infection between neighbor regions. We use the attention mechanism to consider long-term dependencies between features. The model has three separate attention layers, one for each output. Although the attention mechanism helps to focus on the important features, we limited the model input by the regional cluster of Moscow to reduce the number of parameters. First, we build an input embeddings with the dense network. Then we evaluate attention and apply the global pooling layers to build generalized attended representations of the inputs (*Pooling I*, *Pooling D*, *Pooling UE*), and eventually, we process them with another multi-layer feed-forward network.

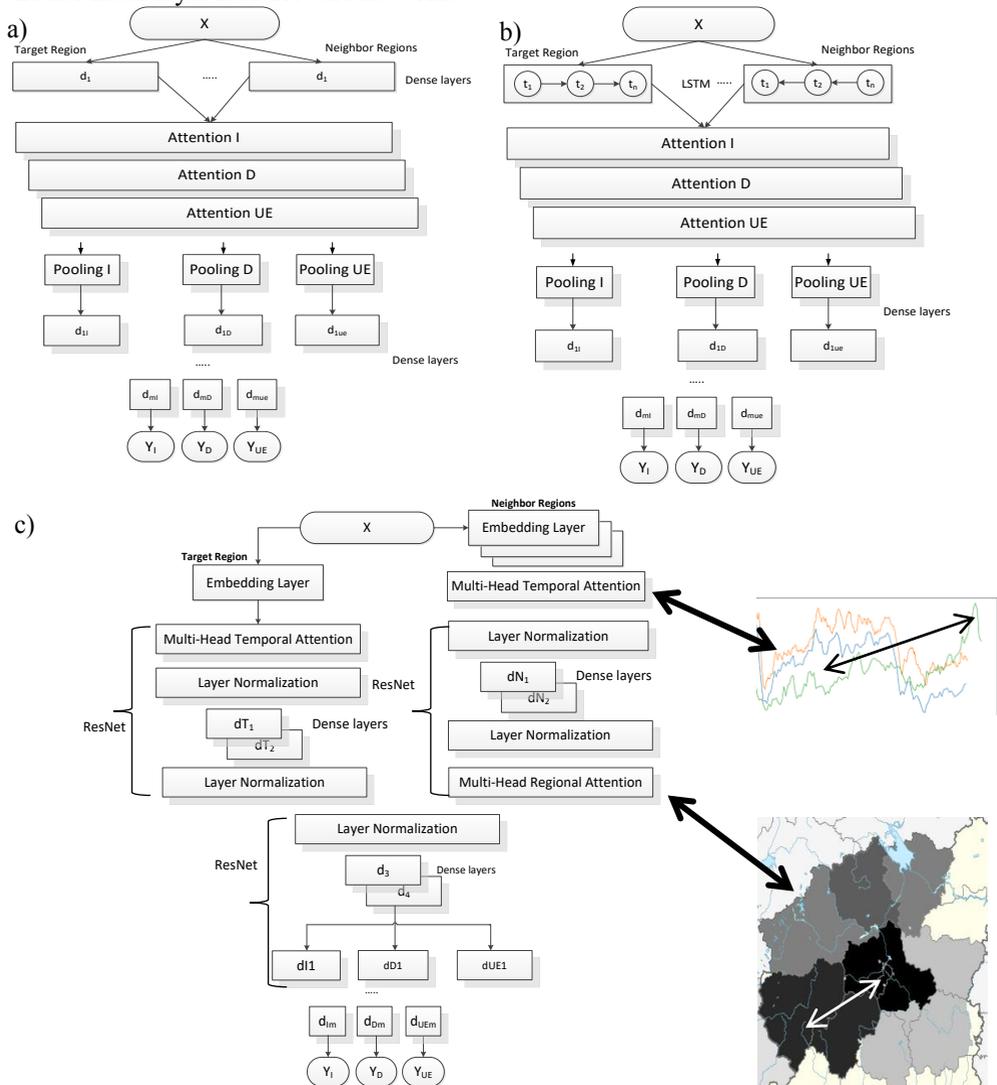


Fig. 1. Models to forecast a virus pandemic impact: a) NARX with attention, b) LSTM with attention, c) Transformer-like with temporal and spatial attention.

The Long-Short Term Memory (LSTM-based) model considers the whole time-series to predict the future values. We use the same inputs and outputs as in the NARX model, except that we do not consider past output values as the input (Fig 1 b). The first layer of the model is a recurrent layer, which builds a single embedding for each input time series. Then that embedding is processed with a deep feed-forward network. Similarly to the NARX-based model, we apply the attention layers to consider the dependencies between the cluster's regions. The model architecture after the recurrent layer is the same as for the NARX model.

The recurrent networks have a gradient vanishing problem, which can be an obstacle for a mid-and long-time series analysis. The Transformer-like model (Fig. 1c.) uses positional encoding and multi-head attention mechanisms, which tackle the problem [16]. That architecture was originally introduced for Natural Language Processing, but then it was generalized and modified to perform time-series predictions [17]. The model has the same inputs and outputs as the previous one. First, the model generates embeddings for the inputs; it uses positional encoding to add temporal information. Second, it applies the multi-head attention to catch the dependencies between past time steps and passes the results through the residual network with normalization. Eventually, the model uses multi-head attention to reveal dependencies between the series from different regions and generates the outputs.

As far as the forecasting error is expected to be distributed normally, we apply quadratic loss to train all the models. We also use a dropout mechanism to regularize the networks. All the network hyper-parameters (number of network cells at each layer, number of layers, dropout level) have been fine-tuned with the 3-fold cross-validation.

We have also implemented a standard statistical ARIMA model as a baseline to predict the infection spread, the deaths, and the number of unemployed, although we trained independent models for each target variable because of limitations of the ARIMA.

5 Experiment Results and Discussion

All the models digest past data from 2020 and predict the impacts between January and March 2021. It is worth noting that we forecast per-day deltas of the effects instead of absolute values. A 3-fold cross-validation procedure is used to obtain the scores so that at each step, a part of the regions was considered only to train, and another one was used to test. Because all the territories differ in population number and density, they also differ in infection spreading. Therefore, the pure MSE (mean squared error) score is not very useful to assess the models. In the experiments, we added the normalized MSE (NMSE), which is scale-independent. First, we tested all possible combinations of the feature sets from Section 2 and revealed that all the features except the "Hospital bed availability" help increase accuracy. The latter can be related to the fact that the 2016 data is outdated and not helpful. Besides, we found that the healthcare-related indicators (Infection rate, Death rate) help to predict the unemployment rate, but not vice-versa. Table 1 shows the results for each model (trained with all the features). Namely, it presents the average of the scores obtained for the selected Russian regions.

Table 1. Average performance scores of the pandemic impact prediction models.

Model	Infected		Deaths		Unemployed	
	MSE	NMSE	MSE	NMSE	MSE	NMSE
ARIMA	53012.00±130.44	0.05±0.01	7.14± 0.57	0.07±0.10	35141.71± 2187.30	0.06±0.01
LSTM	52651.00±139.27	0.04±0.01	7.00±0.65	0.02±0.09	12941.80±643.40	0.02±0.01
NARX	53086.14±660.5	0.05±0.02	6.20±0.45	0.01±0.07	12545.57±192.22	0.01±0.00
Transformer	9012.25±188.0	0.01±0.01	7.70±0.73	0.09±0.09	11587.71±870.30	0.01±0.01

The experiments show that the Transformer-based model achieves the best accuracy for the Infection and Unemployed prediction tasks, confirming the importance of considering interregional and long-term dependencies. Although the scores for the death rate were lower, this can be related to overfitting because the Transformer-based model has the most significant number of parameters. Table 2 shows the detailed scores for the considered regions.

Table 2. Per-region prediction scores for the most populated Russian regions (Transformer model).

Region	Infected		Deaths		Unemployed	
	MSE	NMSE	MSE	NMSE	MSE	NMSE
Moscow	37469.15	0.004	17.30	0.005	50694.10	0.009
Moscow Region	10764.13	0.005	15.74	0.104	25653.85	0.003
Tver Region	140.10	0.009	6.98	0.142	2072.14	0.011
Yaroslavl Region	154.02	0.021	3.55	0.072	636.90	0.002
Kaluga Region	73.28	0.015	7.16	0.146	137.16	0.001
Tula Region	77.25	0.016	2.09	0.042	1272.71	0.012
Vladimir Region	101.85	0.007	6.30	0.127	636.30	0.002
Smolensk Region	120.97	0.005	7.80	0.159	2078.09	0.012

The table confirms that all the impacts can be reliably predicted except the Death rate. That indicator is discrete by nature, mercifully relatively low for most regions; therefore, it can be predicted accurately only for the regions with a high population. However, that type of model requires the largest amount of data to be trained reliably, including data on the number of doctors per thousand inhabitants in the region, the level of doctors and medical personnel's qualifications, and the technical equipment of hospitals and ambulance facilities, etc. The use of such data could increase the accuracy of the model mortality forecast.

6 Conclusion

Space is an essential factor in political, economic, and socio-cultural differentiation. The coronavirus pandemic has exacerbated the economic lag of weaker regions from more developed ones in many countries. In the regions of Russia, in conditions of different population density, the level of business activity, the state of health care systems, and the sectoral specifics of regional economies, there are significant regional imbalances. It is shown that the coronavirus had an ambiguous effect on the degree of economic, social consequences in the provinces, significantly increased intra-industry and inter-territorial divergence.

In this study, we have proposed and tested several models to predict the healthcare and economic impacts of the coronavirus pandemic. The experiments show that healthcare-related features are helpful to predict the economic impact of the pandemic, but not vice-versa.

In general, the test results for the selected Russian regions show that ARIMA, LSTM, NARX, and Transformer models can reliably forecast the impacts, both in terms of the extent of the pandemic spread and consequences. We also revealed that the Transformer-based model, which can digest interregional dependencies, shows the best accuracy for two of three types of impact. However, that type of model requires the largest amount of data to be trained reliably. The presented study results can be used by federal and regional authorities to adjust the policy and measures connected with the coronavirus for each specific region of Russia. Forecast indicators on the spread of infection in neighboring regions can contribute to the formation of interregional interaction programs. The forecast

of regional unemployment indicators can be the basis for forming budgets for temporal measures of state social assistance to the population.

In the future, for forecasting and assessing the consequences, we propose to use a broader range of socio-economic indicators of the regions (per capita income, price index, the volume of industrial production, trade turnover, etc.).

Acknowledgments: The reported study was funded by RFBR, project number 20-04-60188.

References

1. M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, R. Agha, *International journal of surgery*, **78**, 185-193 (2020).
2. I. Chakraborty, P. Maity, *Science of the Total Environment*, **728**, 138882 (2020).
3. A. Tretuakov, K. Danilova, *Theory and practice of World Science*, **6**, 23-32 (2020).
4. C. Gros, R. Valenti, L. Schneider, K. Valenti, D. Gros “*Strategies for controlling the medical and socio-economic costs of the Corona pandemic*” ArXiv preprint arXiv: **2004.00493** (2020).
5. D. Allain-Dupré, I. Chatry, V. Michalun, A. Moisisio “*The territorial impact of COVID-19: Managing the crisis across levels of government*”. *OECD* (2020).
6. Y. Xiong, Y. Wang, F. Chen, M. Zhu. *International journal of environmental research and public health*, **17(11)**, 3903 (2020).
7. J. P. A. Ioannidis, S. Cripps, M. A. Tanner, *International journal of forecasting* (2020).
8. S. Kumar et al. *Forecasting major impacts of COVID-19 pandemic on country-driven sectors: challenges, lessons, and future roadmap*. In *Personal and Ubiquitous Computing*, 1-24 (2021).
9. D. Štifanić et al., *Complexity*, **2020** (2020).
10. S. P. Zemtsov, V. L. Baburin, *Regional Research of Russia*, **10(3)**, 273-290 (2020).
11. S. Gordeev, *Socium and authority*, **5(85)**, 32-50 (2020).
12. O. Dranko, O. Loginovskiy, S. Prikazchikov *Proceedings of South-Ural State University*, **21(1)**, 90-101 (2021).
13. P. Lavrinenko, A. Romashina, P. Stepanov, P. Chistyakov, *Studies on Russian Economic Development*, **6(30)**, 692-699 (2019).
14. COVID-19 demographic and economy impact dataset. URL: [https://github.com/masterdoors/COVID_data_\(2020\)](https://github.com/masterdoors/COVID_data_(2020)).
15. T. Lin et al. *IEEE Transactions on Neural Networks*, **7(6)**, 1329-1338 (1996).
16. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, I. Polosukhin *Attention is all you need*. arXiv preprint arXiv:**1706.03762** (2017).
17. S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y. X. Wang, X. Yan, *Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting*. arXiv preprint arXiv:**1907.00235** (2019).