pHealth 2020
B. Blobel et al. (Eds.)

doi:10.3233/SHTI200653

Modelling of COVID-19 Morbidity in Russia

Georgy KOPANITSA[a, 1], Oleg METSKER[b], Alexey YAKOVLEV[b], Alexey FEDORENKO[b] and Nadezhda ZVARTAU[b]
[a] *ITMO University, Saint-Petersburg, Russia*
[b] *Almazov National Medical Research Centre, Saint-Petersburg, Russia*

Abstract. The outbreak of COVID-19 has led to a crucial change in ordinary healthcare approaches. In comparison with emergencies re-allocation of resources for a long period of time is required and the peak utilization of the resources is also hard to predict. Furthermore, the epidemic models do not provide reliable information of the development of the pandemic's development, so it creates a high load on the healthcare systems with unforeseen duration. To predict morbidity of the novel COVID-19, we used records covering the time period from 01-03-2020 to 25-05-2020 and include sophisticated information of the morbidity in Russia. Total of 45238 patients were analyzed. The predictive model was developed as a combination of Holt and Holt-Winter models with Gradient boosting Regression. As we can see from the table 2, the models demonstrated a very good performance on the test data set. The forecast is quite reliable, however, due to the many uncertainties, only a real-world data can prove the correctness of the forecast.

Keywords. COVID-19, machine learning, morbidity, forecast, Russia.

Introduction

Large-scale natural disasters and pandemics usually result in a deficiency of vital medical resources especially in highly inhabited areas. Thus, it is critical to optimize medical resource allocation to provide the quality of care in emergency situation and to maintain the high level of services for the routine medical needs [1]. As the medical resources include funding, supplies, equipment, facilities, and personnel, the change in allocation especially for a longer period can potentially affects the whole healthcare system. Re-allocation of resources is a common practice in emergency situations [2] when resources temporally must be shifted to a certain groups of diseases [3]. Normally, this happens temporary and doesn't last long or at least for a known or well predictable amount of time. Although this is an emergency measure it doesn't have long term effects on the health care system on a national or international level. The unpredictable increase in the load on medical institutions, as well as the necessary anti-epidemic measures, causes delays in the provision of emergency medical care to patients, which leads to worsening or even death of patients who have not received prompt medical care [4]. With COVID-19 outbreak the situation is different. In comparison to the emergencies re-allocation of resources is required for a longer period and the peak utilization of the resources is also

[1] Corresponding Author. Georgy Kopanitsa, ITMO University, Saint-Petersburg, 192034, Russia; Email: georgy.kopanitsa@gmail.com.

hard to predict. Furthermore, the epidemic models do not provide reliable information about the development of the pandemic, so it creates a high load on the healthcare systems with unforeseen duration.

Neural network-based models have already shown good results for predicting the COVID19 pandemic around the world [9]. In this study we develop predictive models for morbidity for COVID19 patients in Russia.

1. Methods

The data from the Russian government resource (https://digital.gov.ru/) and John Hopkins University (https://www.jhu.edu/) was used to train predictive models.

To find the most efficient model we made a grid search experiment. The dataset was split into training (70% random selection) and testing (30% random selection). Each experiment ran in the setting of stratified 5-fold cross-validation i.e., random 80% of training dataset was used for training and random 20% of training dataset for testing. Target class ratios in the folds were preserved. Mean Absolute Error (MAE) was used as a performance metric. After determining the optimal dataset and model parameters, we performed a validation with the testing dataset. We used a series of classification models available within scikit-learn as a pool of Random Forest, Gradient Boost and Voting regressors for the selection of the best predictive methods to be applied within the proposed scheme. The prediction models for New Cases per day, Total Cases, Total recovered and Mortality per day were implemented using the most efficient regressor in combination with Holt and Holt-Winter models. The models were evaluated using Root mean squared error (RMSE) and MAE.

2. Results and Discussion

Table 1 presents the performance of the regressors during the grid search. The predictive model was developed as combination of Holt and Holt-Winter models with Gradient boosting Regression.

Table 1. Regressors' performance

Regressor	New Cases	Total cases	Total recovered	Mortality per day
Random Forest	15.26	10.2	29.34	11.42
Gradient Boosting	79.9	0.01	23.06	0.0078
Voting Regressor	404.56	5.71	43.23	5.713

The predictive model was developed as combination of Holt and Holt-Winter models with Gradient boosting Regression. Figures 1 to 3 present a forecast for New Cases per day, Total Cases, Mortality per day and recovery per day. The following tests were performed on the example of the Saint Petersburg region.

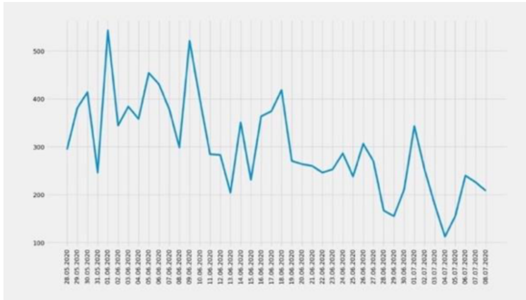

Figure 1. Forecast of new cases

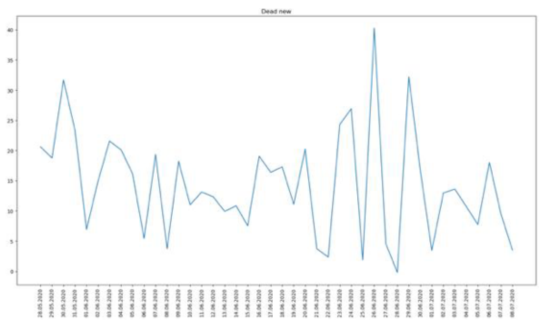

Figure 2. Forecast of new mortality cases per day

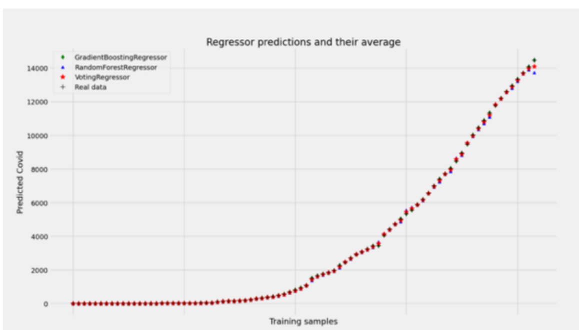

Figure 3. Forecast of recoveries per day

Table 2 presents the performance of the implemented models on the test dataset.

Table 2. Gradient boosting performance

Metrics	New Cases	Total cases	Total recovered	Mortality per day
RMSE	0.01	0.21	0.18	2.04
MAE	81.1	0.01	24.32	0.01

As we can see from the table 2, the models demonstrate a very good performance on the test data set. The performance results are consistent with the models from other countries [7,8]. The forecast is quite reliable, however, due to the many uncertainties, only a real-world data can prove the correctness of the forecast. The main result of the study is the prediction of the number of new cases, the total number of cases, the number of deaths and recoveries by day, with high accuracy, trying to predict daily fluctuations.

Existing models, including those used at the regional and federal level to plan resource requirements (beds, equipment, personnel, consumables, funding) estimate indicative maximum requirements within a sufficiently broad time frame (here we can refer to WHO and Cornell University calculations). At the same time, fluctuations with a significant amplitude within a day, which fit well into the general smoothed trend, can lead to serious overstrain of the health system, short-term acute shortage of resources, fraught with short but significant peaks in the number of lethal outcomes, the risk of forming centers of infection, which may already have a significant impact on the situation. In situations where, operational reserves cannot be maintained and idle capacity is minimal, short-term forecasting can provide additional opportunities to optimize the use of available resources.

3. Conclusion and Future Work

The aim of this study was to design and develop technologies, methods and models of health processes under long-term external influences, including pandemics, to optimize the development of regional health systems.

Acknowledgement

The research was supported by Russian Science Foundation (project No. 17-15-01177). The work of Georgy Kopanitsa was financially supported by the Government of the Russian Federation through the ITMO fellowship and professorship program.

References

[1] Zhuang J, Bier VM. Balancing terrorism and natural disasters-defensive strategy with endogenous attacker effort. Oper. Res. 2007; 55,5:976-991. doi:10.1287/opre.1070.0434.
[2] Nuti S, Vainieri M, Bonini A. Disinvestment for re-allocation: A process to identify priorities in healthcare. Health Policy (New York) 2010; 95,2-3:137-43. doi:10.1016/j.healthpol.2009.11.011.
[3] Bhattacharjee P, Ray PK. Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections. Comput Ind Eng. Dec. 2014; 78: 299-312. doi:10.1016/j.cie.2014.04.016.
[4] Bostick NA, Subbarao I, Burkle FM, Hsu EB, Armstrong JH, James JJ. Disaster triage systems for large-scale catastrophic events. Disaster Med. Public Health Prep. 2008 Sep; 2 Suppl 1: S35-9. doi:10.1097/DMP.0b013e3181825a2b.
[5] Rao ASRS, Vazquez JA. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. Infect. Control Hosp. Epidemiol. 2020 Mar 3; 1-5. doi:10.1017/ice.2020.61.
[6] Tárnok A. Machine Learning, COVID-19 (2019-nCoV), and multi-OMICS. Cytom. Part A. 2020 Mar; 97,3: 215-216. doi:10.1002/cyto.a.23990.
[7] Santosh KC. AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. J. Med. Syst. 2020; 44,5: 93. doi:10.1007/s10916-020-01562-1.
[8] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons and Fractals. May 2020; 134: 109761. doi:10.1016/j.chaos.2020.109761.