Two machine-learning approaches for short-term COVID-19 hospitalization forecasting in Slovakia

Veronika Kurilová
Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia
Department of ophthalmology of Slovak Medical University and University hospital in Bratislava, Antolská 11, 85107 Bratislava, Slovakia

Martin Huba
Swiss Re, Management AG
Mlynské Nivy 12, 811 09 Slovakia

Jozef Goga, Miloš Oravec, Jarmila Pavlovičová
Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia

Nora Majtánová
Department of ophthalmology of Slovak Medical University and University hospital in Bratislava, Antolská 11, 85107 Bratislava, Slovakia
Faculty of Medicine, Slovak Medical University,
Limbová 12, 833 03 Bratislava, Slovakia

Abstract. *COVID-19 is a life-threatening novel respiratory virus-borne disease, which was discovered in December 2019 in Wuhan and subsequently spread globally. Monitoring and predicting COVID-19 epidemic data is crucial to control pandemic outbreaks. Machine learning-based methods, including deep learning, are promising approaches to predict COVID-19 data such as new cases, infected patients, and deaths. Our study focused on short-term COVID-19 hospitalizations forecasting using two machine learning approaches— ensemble time-series method and multilayer perceptron (MLP) feedforward network method. Both methods make predictions based on hospitalization, polymerase chain reaction (PCR), and antigen (Ag) test data, which were collected between October 2020 and June 2021 in Slovakia for our study. The ensemble time-series method was more sensitive in the beginning of experimental period but failed when the number of hospitalizations began to drop. The MLP method was ineffective in the beginning because of lack of training data but improved when more robust data was available; this method is promising for monitoring the third wave of pandemic in Slovakia.*

1 Introduction

The first patients with the novel coronavirus SARS-CoV-2, were hospitalized in Wuhan, China in December 2019 [1]. In January 2020 more cases were reported throughout China and abroad [2]. The most sensitive diagnostic method currently available for COVID-19 testing is the polymerase chain reaction (PCR) test [3]. However, for effective screening, frequent repetition and fast reporting is more important than sensitivity [4], which makes rapid antigen (Ag) tests or loop-mediated isothermal amplification tests (LAMP) tests advantageous in COVID-19 diagnostics. There are several studies on the relationship between symptom onset, positive PCR testing, and hospitalization. The fifth day post infection is a typically when symptom onset occurs, and most infected people test false negative before this day. The decrease in probability of false negativity is noted four days before symptoms onset, from 100% four days before onset to 67% one day before onset. On the day of onset, the probability of getting a false negative result is less than 40% [5]. From zero to four days after symptom onset, PCR tests from nasopharyngeal swabs are positive [6] in most infected individuals, with a peak in the first week after onset [7]. The median time between symptom onset and hospitalization is 5 days [6], and the median number of days from symptom onset to death was 14 and less in patients aged over 70 [8]. Despite these observations, which demonstrate the importance of time in COVID-19 infections, predicting the number of hospitalized patients from positive tests and average hospitalization period is not straightforward and depends on personal and regional factors. A nationwide cohort study reported that 20% of all PCR-positive cases result in hospitalizations, and the proportion increases with age and multimorbidity [9]. In another study [10], stronger hospitalization risk is associated with men aged ≥ 75 years with comorbidities, particularly cardiovascular disease, diabetes chronic kidney disease, hyperlipidemia and obesity than in other groups.

Predicting COVID-19 epidemic data and monitoring epidemiological changes of the virus spread are crucial for controlling pandemic outbreaks [1]. Machine learning methods, including deep learning, show promise in predicting COVID-19 epidemic data such as new cases, infected patients, and mortality. A multilayer perceptron (MLP) artificial neural network was used in [11] to create a worldwide model for predicting the maximum number of infected patients in a location from available data in time. The MLP was shown to have slightly better performance for analyzing contributing factors for COVID-19 spread and deaths than the radial basis function in [12]. The authors of [13] analyzed continuous variable quantum neural networks and quantum backpropagating MLP models for predicting COVID-19 cases in India and the USA. Both methods showed better performance than classical artificial neural networks. MLP and adaptive network-based fuzzy inference systems showed promising results in outbreak predictions [14]. Hybrid machine learning prediction models of adaptive network-based fuzzy inference system and MLP-imperialist competitive algorithms were used in [15] to predict cases and mortality rate. Machine learning models, such as the linear regression, linear regressor polynomial, support vector regressor, random forest regressor, decision tree, and auto-regressive moving average were used to predict outbreaks [16] [17] [18] with the highest accuracy being shown by the auto regressive moving-average [16] and random-forest approaches [18]. A random forest model was also proposed to predict mortality in the first 20–84 hours following hospitalization [19].

Deep learning forecasting methods for new or, new and recovered cases using recurrent neural networks were proposed in [20] and [21]. The authors tested the long short-term memory (LSTM), bidirectional LSTM, gated recurrent units, variational autoencoder, and convolutional LSTM; the variational autoencoder showed the best performance among all of these [20]. The convolutional LSTM outperformed other models in predicting new cases for a one-month period [21].

However, the development and subsequent practical application of models based on deep learning requires high computing power [22], which aggravates the financial disparities between different universities [23]. At least a partial solution towards the democratization of research in the field of deep learning is to use commercial cloud computing platforms, which allow the direct purchase of necessary computing power [24]. This approach has proven to be effective and is beginning to be applied in fields such as information retrieval [25], flexible maintenance [26], and time-series prediction [27], [28].

Prediction of COVID-19 hospitalizations in Slovakia based on linear regression was firstly conducted by [29]. The problem of preventing the spread of the disease is complex, and multidisciplinary approaches, including artificial intelligence methods, are required.

Our paper focuses on short-term COVID-19 hospitalizations forecasting in Slovakia with a machine learning ensemble model implemented on the MS Azure [30] cloud computing platform and an MLP feedforward network implemented locally using in MATLAB [31]. The automated machine learning (AutoML) [32] approach enables acceleration of the development and deployment of machine learning models without extensive programming knowledge, making it suitable and user-friendly for epidemiologists and data analytics professionals.

The goal of hospitalization predictions is to aid the preparedness of hospitals and health-care professionals to admit all patients required hospitalizations and provide them proper healthcare without rescheduling planned elective care; it also aids the redistribution of hospitalized patients among different regions and districts as needed.

2 Methods

We propose two different machine learning methods: the first based on ensemble learning and the second based on the MLP method. The first was developed and tested in real time during the second wave of pandemic in Slovakia. The second was tested retrospectively.

2.1 Data acquisition

We used the dataset provided by The Institute for Healthcare Analysis [33], publicly available on Github. The dataset includes COVID-19 statistics in Slovakia, i.e., the daily number of positive and total PCR and antigen (Ag) tests; number of hospitalized patients including daily hospital admissions and discharges; vaccination statistics, etc. Available hospitalizations data are divided by districts and regions because of reporting from every hospital every day.

2.2 Preparing dataset for machine learning time-series ensemble method

The model training time-series data were the daily numbers of PCR and Ag positive tests; daily percentage of positive PCR; and Ag tests from total PCR and Ag tests. All training data were filtered by a seven-day simple moving average (SMA) filter. We assumed that testing PCR or Ag positive leads to a hospitalization time of four and seven days on average, respectively. These time shifts were inspired by [29], where the average time for positive PCR and Ag test hospitalization was taken as three and seven days, respectively. We performed these time shifts with time-series training data; the schematic of a sample data preparation is shown on Fig. 1. The distributions of the variables in the dataset and correlations between all these variables with hospitalizations are provided in the Appendix. The last PCR and Ag test data from the dataset were used as inputs for predicting hospitalizations from time-series data. The ensemble method was trained on the entire time-series dataset with 1/14-day shifts until the first day of prediction without splitting into time periods. This means that a separate training dataset was used for every 14 days, ending exactly before prediction, allowing the continuation of this time series using the data from the previous days for forecast computations. The method predicts for the next 14 days using the rest of time shifted data (Fig. 1). Because of the time shifting, we can use the last four days from PCR testing and last seven days from the Ag testing as future values and consider these as prediction inputs. In this way, we created input data, allocating the last sample representing 14 days, as a test set. We used 5-fold rolling origin cross validation (ROCV) with a fixed starting point on the remaining training data. Shifting the data by 1/5 thus created cross-validation folds, which ensured that there was no data leakage.

2.3 Preparing dataset for MLP method

The MLP trains with the input and output data. It needs to have a robust dataset to be well trained. Prediction was performed for 14 days using input data from 14 days before first prediction day. Training input data were moved by one day to the right until the end of the dataset. As in ensemble method, the inputs were created from daily PCR and Ag test data. In addition, hospitalizations in last 14 days were used. To enlarge the dataset, data from all regions of Slovakia (8 regions) and their summaries were used for

Dataset

PCR	PCR%	Ag	Ag%	Hosp
1486	0.136	48	0.070	350
1582	0.142	44	0.073	375
1580	0.143	36	0.061	413
1477	0.147	31	0.052	469
1502	0.149	35	0.052	481
1615	0.152	29	0.051	482
1586	0.156	31	0.053	560
1659	0.162	46	0.049	622
1790	0.169	59	0.045	668
2001	0.176	69	0.052	674
2066	0.182	74	0.060	765
2094	0.187	97	0.064	891
2192	0.192	125	0.066	973
2343	0.193	151	0.072	1025
2454	0.195	165	0.069	1087
2409	0.185	218	0.062	1150
2300	0.172	348	0.056	1237
2382	0.154	396	0.050	1320
2500	0.146	553	0.042	1360
2547	0.152	534	0.033	1376
2429	0.146	514	0.024	1405
2285	0.146	499	0.028	1392
2286	0.146	476	0.027	1444
2213	0.151	399	0.027	1441
				?

Training

PCR	PCR%	Ag	Ag%	Hosp
1477	0.147	48	0.070	622
1502	0.149	44	0.073	668
1615	0.152	36	0.061	674
1586	0.156	31	0.052	765
1659	0.162	35	0.052	891
1790	0.169	29	0.051	973
2001	0.176	31	0.053	1025
2066	0.182	46	0.049	1087
2094	0.187	59	0.045	1150
2192	0.192	69	0.052	1237
2343	0.193	74	0.060	1320
2454	0.195	97	0.064	1360
2409	0.185	125	0.066	1376
2300	0.172	151	0.072	1405
2382	0.154	165	0.069	1392
2500	0.146	218	0.062	1444
2547	0.152	348	0.056	1441

Prediction input

PCR	PCR%	Ag	Ag%
1790	0.146	29	0.050
2001	0.146	31	0.042
2066	0.146	46	0.033
2094	0.151	59	0.024
AUTO	AUTO	69	0.028
AUTO	AUTO	74	0.027
AUTO	AUTO	97	0.027
AUTO	AUTO	AUTO	AUTO
AUTO	AUTO	AUTO	AUTO
AUTO	AUTO	AUTO	AUTO
AUTO	AUTO	AUTO	AUTO
AUTO	AUTO	AUTO	AUTO
AUTO	AUTO	AUTO	AUTO
AUTO	AUTO	AUTO	AUTO

→ **Hosp**: **14-DAYS PREDICTION**

Fig. 1. Preparing the training dataset for machine learning time-series ensemble method using time shifts. The example dataset (left) is divided into the training part (upper right) and prediction input part (bottom right) using time shifts. Missing days in the prediction input (labeled as "AUTO") are automatically filled by MS Azure. The entire process results in 14 days of hospitalization forecasting (labeled as "14-DAYS PREDICTION"). The example is shown without normalization for better visualization. Following data are mentioned in the figure: number and percent of positive polymerase chase reaction tests (PCR and PCR%), number and percent of positive antigen tests (Ag and Ag%), number of hospitalized patients (Hosp).

training. This dataset was nine times larger than one with only summary data.

For example, in one of our models, trained from 11/10/2020 till 02/02/2020 (i.e., 85 days) we obtained 58 inputs and their outputs (14 × 58 inputs together with 14 × 58 outputs moved by one day) once from every region and once from the whole of Slovakia, giving 58 × 9 = 522 input data and their outputs for training. We applied a seven-day moving average filter to the input data for preprocessing and normalization to simplify the function fit. All data were normalized to a scale of 0 to 1, with 1 corresponding to 1.5 times the maximum value in positive or hospitalized inputs and 100% as the maximum if inputs are in percent. The dataset was divided randomly and 80% assigned as training data and 20% as validation data. The last 14 days before prediction period were used as the test data to get the final 14 days prediction. The dataset was prepared in MATLAB 2020b.

2.4 Time-series ensemble method

The time-series machine learning ensemble method was trained in the MS Azure cloud machine learning module, which is a cloud-based machine learning service with a user interface. This allows users without programming knowledge to train; all that needs to be done is to upload the training dataset, choose the built-in method and desired features, and click to start. Azure trains huge number of models by default and compares them. After considering several models (such as decision tree, random forest, AutoArima, ProphetModel, ElasticNet, GradientBoosting, and LassoLars) we chose the voting ensemble model because of smallest normalized root mean squared error (RMSE) on prediction gained by this model compared to other available machine learning models. Best performance of this model among all available models was confirmed in most of our experiments. The architecture of Azure Voting ensemble model is shown in Fig. 2. The model consists of six soft voting base regressors: three gradient boosting regressors, one random forest regressor and two decision tree regressors, each with different parameters. The voting ensemble model considers the predictions of every regressor, which are weighted and averaged, the final prediction being the weighted average from all regressors in the ensemble model. The hyperparameters used are listed in Table 1.

After training the model, the online endpoint must be created. An online endpoint is an HTTPS endpoint which is called by the user to obtain output of trained model. It contains deployments to receive data and send responses in real time. Access to the endpoint is deployed through a Python script in combination with the prediction input data.

2.5 MLP method

For the 14-day prediction, we used a standard MLP with two hidden layers. We used 15 neurons with the hyperbolic tangent activation function in each hidden layers. The output layer with 14 neurons represents a multi-step prediction of hospitalization with a linear activation function at the output.

The architecture of our MLP network is shown in Fig. 3. Training and validation data are randomly divided using early-stopping with six validation checks. The Levenberg–Marquardt training algorithm was used, and performance measured using the mean squared error.

2.6 Evaluating the results

The results of both methods were evaluated with RMSE and mean absolute percentage error (MAPE) using the standard formulas

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(F_t - A_t)^2}{n}}, \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|, \tag{2}$$

where n is number of fitted points, A_t is the actual value, and F_t is forecast value.

Table 1. Hyperparameters used in Azure Soft Voting Ensemble model.

Model	Hyperparameters
Gradient boosting regressor 1	Loss: least squares regression Learning_rate: 0.01; N_estimators: 600; Subsample: 0.95 Criterion: "friedman_mse"; Min_samples_split: 0.007532; Min_samples_leaf: 0.006152; Max_depth: 5; Max_features: 0.9; Validation_fraction: 0.1; Tolerance: 0.0001
Random forest regressor	Bootstrap: True; N_estimators: 200; Subsample: 0.95; Criterion: "mse"; Min_samples_split: 0.001281; Min_samples_leaf: 0.001953; Max_depth: None; Max_features: 0.4
Gradient boosting regressor 2	Loss: least squares regression; Learning_rate: 0.1; N_estimators: 600; Subsample: 0.45; Criterion: "mse"; Min_samples_split: 0.052854; Min_samples_leaf: 0.023458; Max_depth: 3; Max_features: 0.1; Validation_fraction: 0.1; Tolerance: 0.0001
Decision tree regressor 1	Criterion: "mse"; Min_samples_split: 0.003709; Min_samples_leaf: 0.007595; Max_depth: None; Max_features: 0.9 Splitter: "best"
Gradient boosting regressor 3	Loss: "hubel"; Alpha: 0.9; Learning_rate: 0.01; N_estimators: 400; Subsample: 0.35; Criterion: "mse"; Min_samples_split: 0.008992; Min_samples_leaf: 0.013218; Max_depth: 6; Max_features: 0.9; Validation_fraction: 0.1; Tolerance: 0.0001
Decision tree regressor 2	Criterion: "friedman_mse"; Min_samples_split: 0.007532; Min_samples_leaf: 0.009524; Max_depth: None; Max_features: None; Splitter: "best"
Weights (w1–w6):	0.400,0.0667,0.0667,0.0667,0.2667,0.1333

3 Results

The results of the time-series machine learning ensemble method are shown in Table 2. The ensemble method performed well in first three predictions in February when hospitalizations had risen and in March when it predicted the peak of the second wave in Slovakia. When comparing with real hospitalizations cases, the RMSE and MAPE values were in the range of 61.62–91.49 and 1.43–2.19, respectively. The method failed to predict the drop in hospitalizations from April (RMSE: 440.19 and MAPE: 16.21) and later (Fig. 4). The importance of the variables in the ensemble model is listed in Fig. 5. Highest importance showed percent of positive antigen test, following with positive antigen tests. Correlations of these variables in this model are shown in Appendix section.

In contrast, the MLP prediction method was quite inaccurate during the first months of this year (RMSE in range of 114.81–328.64, MAPE in range 2.76–12.94). The prediction accuracy improved with time; more training samples led to better results. Table 3 gives the average RSME and MAPE values for all regions and for the whole of Slovakia. Values for all of Slovakia were computed as a summary of all regions. The best regional and summary results were obtained in May.

Comparisons of the performances of both the proposed methods in three random time periods are shown in Table 2 and Fig. 4.

Predictions of MLP method in all regions and the whole of Slovakia in random time periods in May improved compared to those predicted earlier (Fig. 6, Table 3-marked bold).

Table 2. Results of time series ensemble and MLP predictions for the whole of Slovakia. The MLP predictions were made directly for the whole Slovakia (and not as a summary of all regions predictions). Time periods are from the year 2021.

Time period	Metrics			
	Ensemble RMSE	Ensemble MAPE	MLP RMSE	MLP MAPE
3.2.–16.2.	61.62	1.43	328.64	9.31
17.2.–2.3.	91.49	2.19	-	-
5.3.–18.3.	82.71	1.93	114.81	2.76
3.–16.4.	440.19	16.21	311.44	12.94

Table 3. Results of MLP method: average RMSE and MAPE in all regions and entire Slovakia (obtained as summary of all regional predictions). Metrics were computed from a seven-day moving average of hospitalizations. Time periods are from the year 2021.

Time period	Metrics			
	Average RMSE regions	RMSE whole country	Average MAPE regions	MAPE whole country
3.2.–16.2.	72.22	308.98	15.57	8.88
5.3.–18.3.	41.67	140.71	8.77	3.31
3.4.–16.4.	34.69	350.91	14.68	14.35
1.5.–14.5.	**17.81**	**66.95**	**17.77**	**5.94**

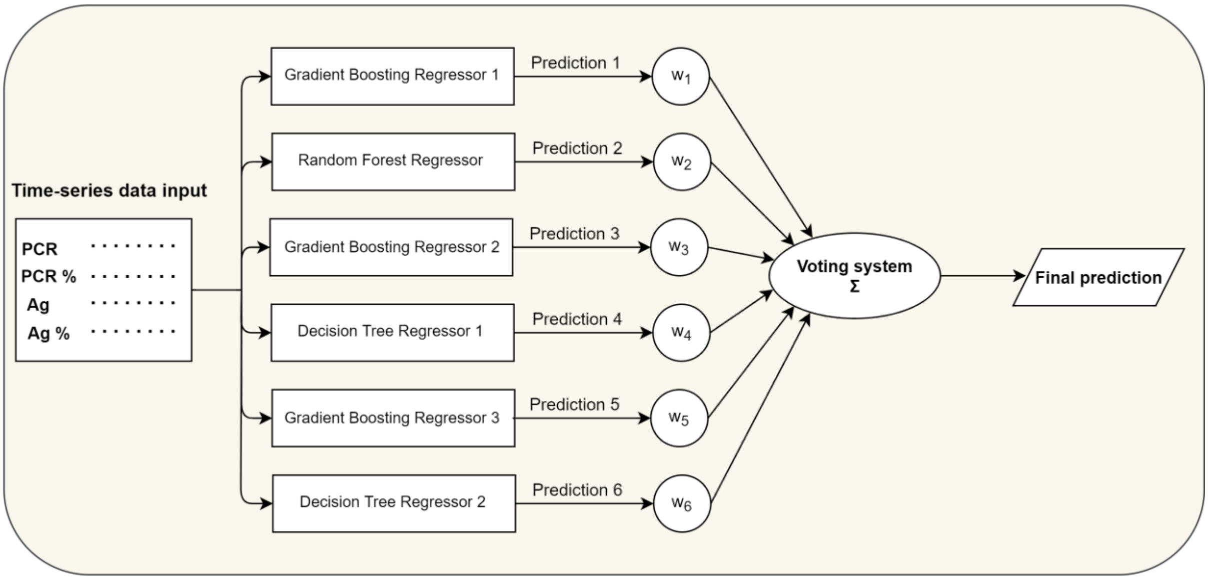


Fig. 2. Architecture of our implemented voting ensemble time-series model in MS Azure. The time-series ensemble model consists of six base regressors whose predictions are weighted with weights (w_1-w_6) and enter the voting system. In the figure are mentioned following data: number and percent of positive polymerase chase reaction tests (PCR and PCR %), number and percent of positive antigen tests (Ag and Ag %).

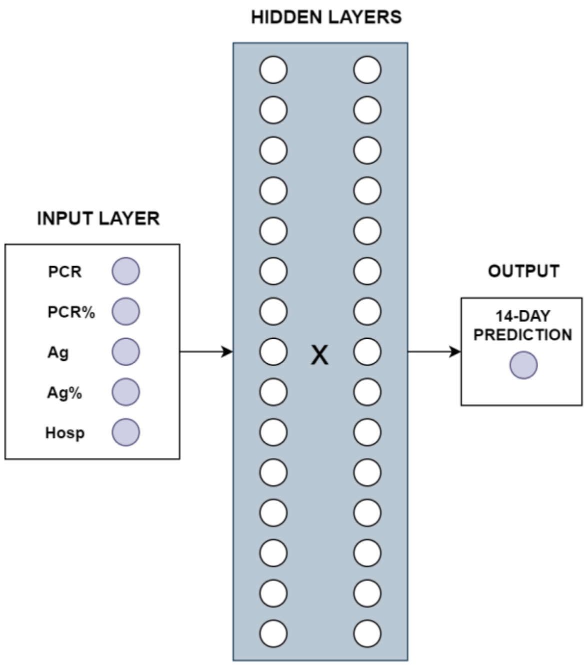


Fig. 3. Architecture of MLP feedforward network. Input layer consists of five time-series inputs, following two hidden layers each containing 15 neurons and the 14-day time-series prediction as the output. Following data are mentioned in the figure: number and percent of positive polymerase chase reaction tests (PCR and PCR%), number and percent of positive antigen tests (Ag and Ag%), number of hospitalized patients (Hosp).

4 Discussion

We proposed two machine learning approaches for short-term hospitalization forecasting in Slovakia. The first approach is time-series ensemble method and the second an MLP neural network.

The ensemble method performed well at the beginning of the experimental period, with the best RMSE being 61.62, but failed when hospitalizations decreased. This could be due to lack of training data— the method was trained with data from the whole of Slovakia only from November 2020. In that time hospitalizations had risen, and the data from the period when cases were decreasing could not be learned. Surprisingly, the peak of the second wave, which followed the decrease in hospitalizations was predicted successfully with this approach. After that, no successful predictions with decreasing of hospitalizations were made. In addition, predicting hospitalizations with cloud-based user-friendly built-in services could make this solution accessible to non-programmers and easier to implement.

Using the MLP method, the initial predictions were inaccurate. Its performance improved with time with accurate results being obtained from the time period in May. This success was also observed in regional predictions— with an average RMSE of 17.81. The RMSE for all regional summaries was 66.95, which we consider as best result for the MLP method for whole Slovakia. We assume that The improvement of the MLP results with time are due to the increase in training dataset size.

We propose that regional predictions with RMSE lower than 20 and for all of Slovakia with RMSE lower than 100 can be valuable in practice.

We took 14 days as our forecasting periods; however, shorter prediction periods are expected to give better results. As prediction period increases, the discrepancy between predicted and real numbers rises. In addition, predicting in shorter time periods in the MLP method leads to more robust dataset, which may lead to even better results. This can be a promising direction for further investigation.

Using only positive tests and previous hospitalizations as inputs may not be sufficient in the future. This experiment was done during the second wave of pandemic in Slovakia, when the vaccination status was not an important factor, and therefore, we did not notice any sudden change in the age distribution of positive tests. As vaccination begins, new input variables would be necessary, such as the percentage of vaccinated individuals in the population or in the elderly and the daily mean age of tested positive. This would be especially relevant during the third wave, when due to vaccination, the age distribution among positive tested and hospitalized may differ. We hope that with new input variables and more robust data, these methods can adapt to such changes.

Forecasting COVID-19 hospitalizations is critical for monitoring pandemic outbreaks and provide healthcare without compromising on elective care. Redistribution of patients among district and regions can be considered based on such predictions if there is a shortage of hospital beds. Our machine learning forecasting approaches are promising

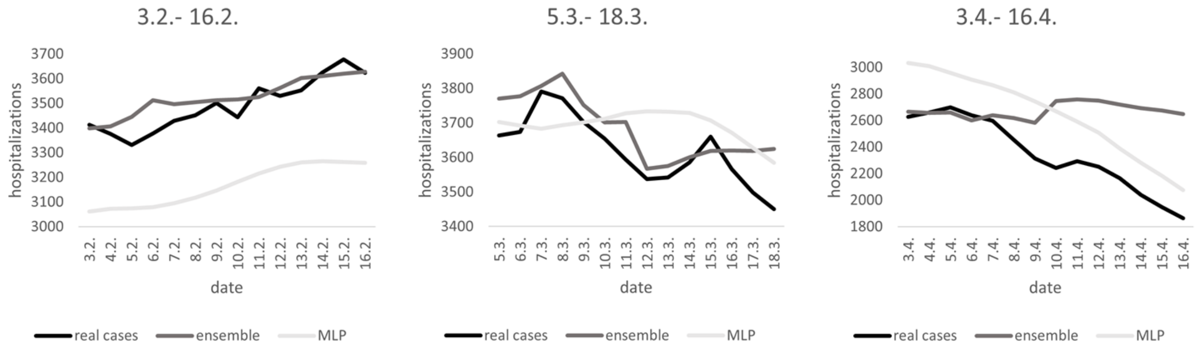


Fig. 4. Comparison of time-series ensemble and MLP method hospitalization predictions in three random time periods.

when sufficient training data is available. Augmenting the training dataset using data from all regions, (as in our MLP method) increases the accuracy of predictions, which gives hope for forecasting hospitalizations in the coming third wave of COVID-19.

Author contributions

V.K. developed the experimental premise, design, and procedures. V.K., and M.H. conducted the research, trained the ensemble method and networks, and analyzed the data. V.K. and J.G. processed the figures and analyzed the data. V. K. and J.G. prepared the manuscript. All authors interpreted the results, contributed to manuscript revision, and approved the submitted version.

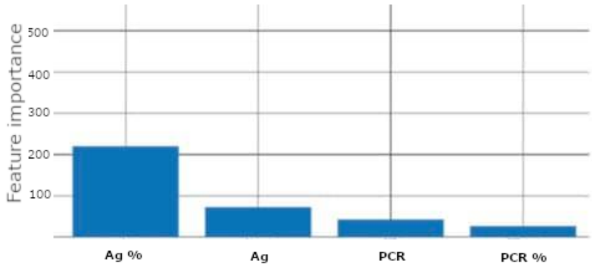


Fig. 5. Importance of variables in proposed ensemble model from March 2021, obtained from model analysis in MS Azure. Following data are mentioned in the figure: number and percent of positive polymerase chase reaction tests (PCR and PCR %), number and percent of positive antigen tests (Ag and Ag %). Feature importance is computed in MS Azure using permutation feature importance inspired by [34].

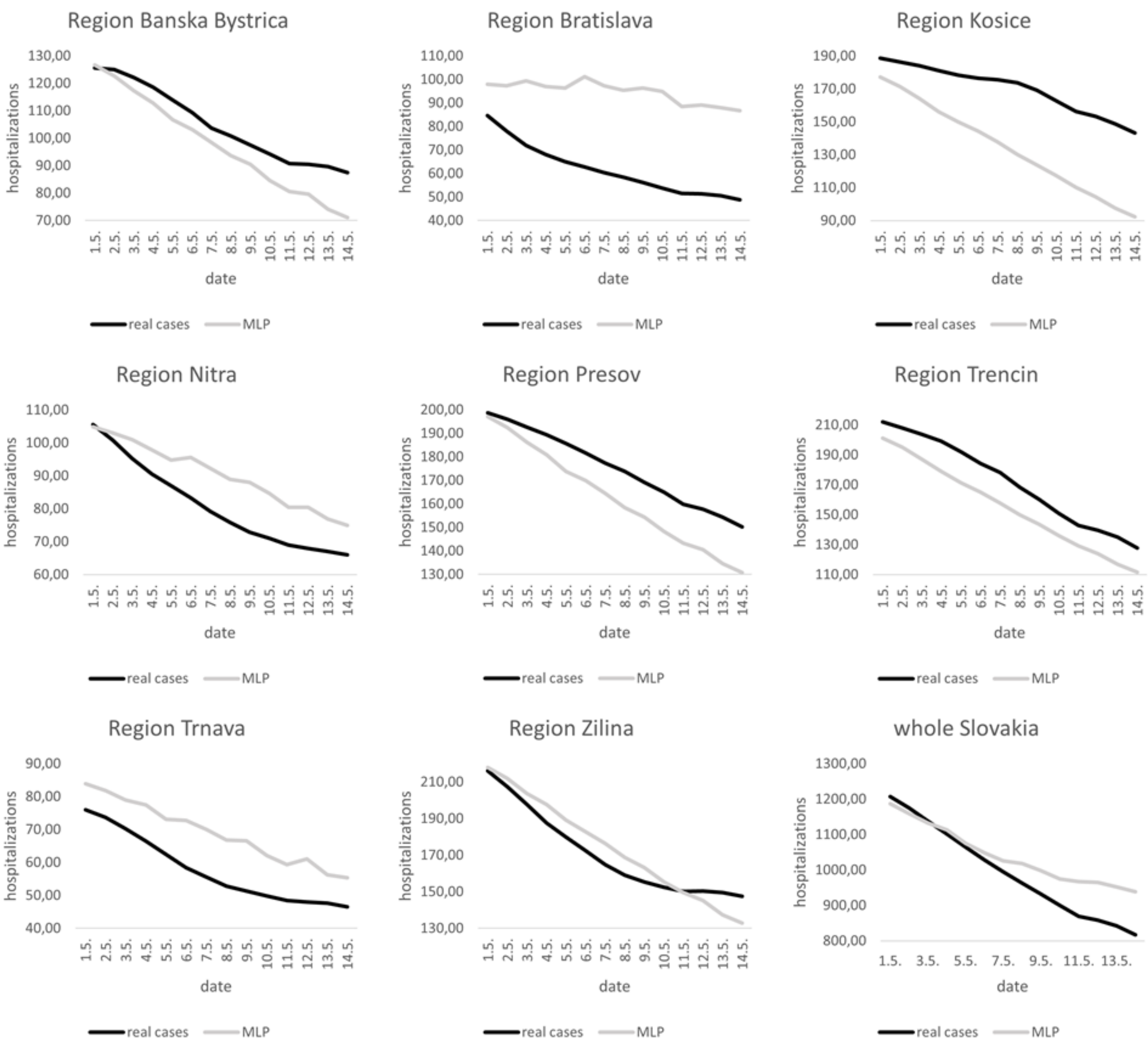


Fig. 6. Results of MLP method in random time periods in May in all Slovakian regions and the whole of Slovakia. Hospitalization data were filtered by a seven-day moving average filter.

Appendix

Fig. 7 Basic data overview: distribution of number and percent of positive polymerase chain reaction tests (PCR, PCR perc.), number and percent of positive antigen tests (Ag, Ag perc.) and hospitalizations in our dataset. Graphs were created in MS Azure.

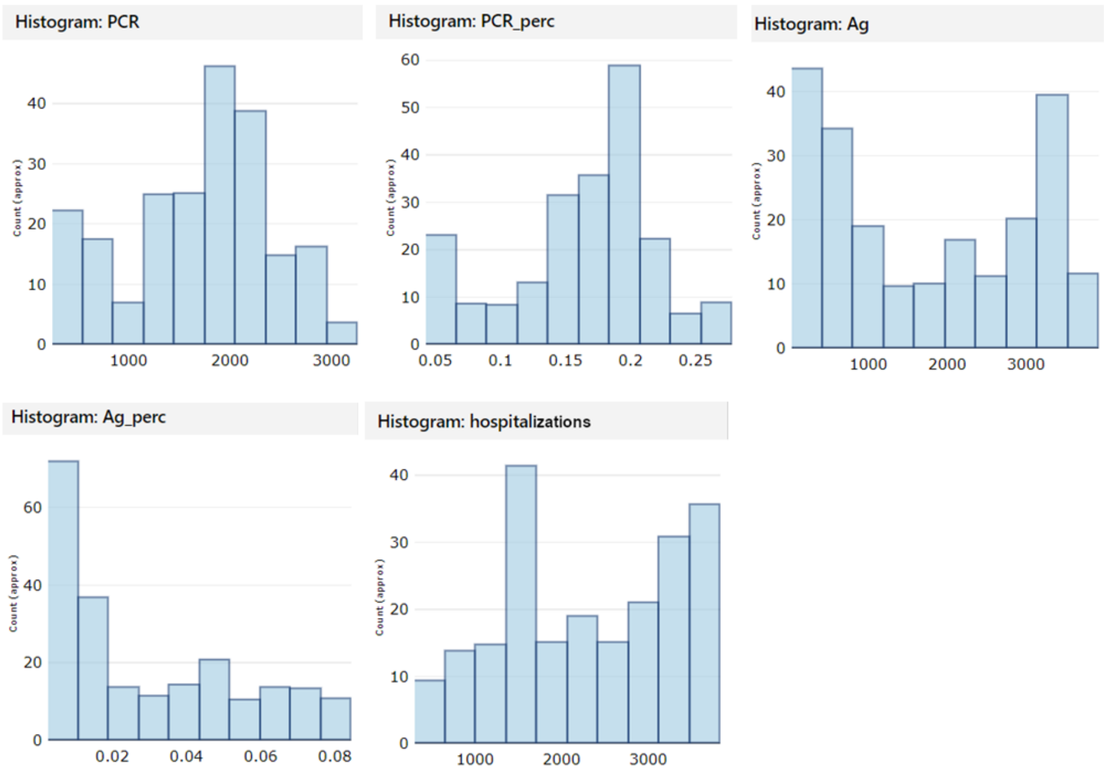
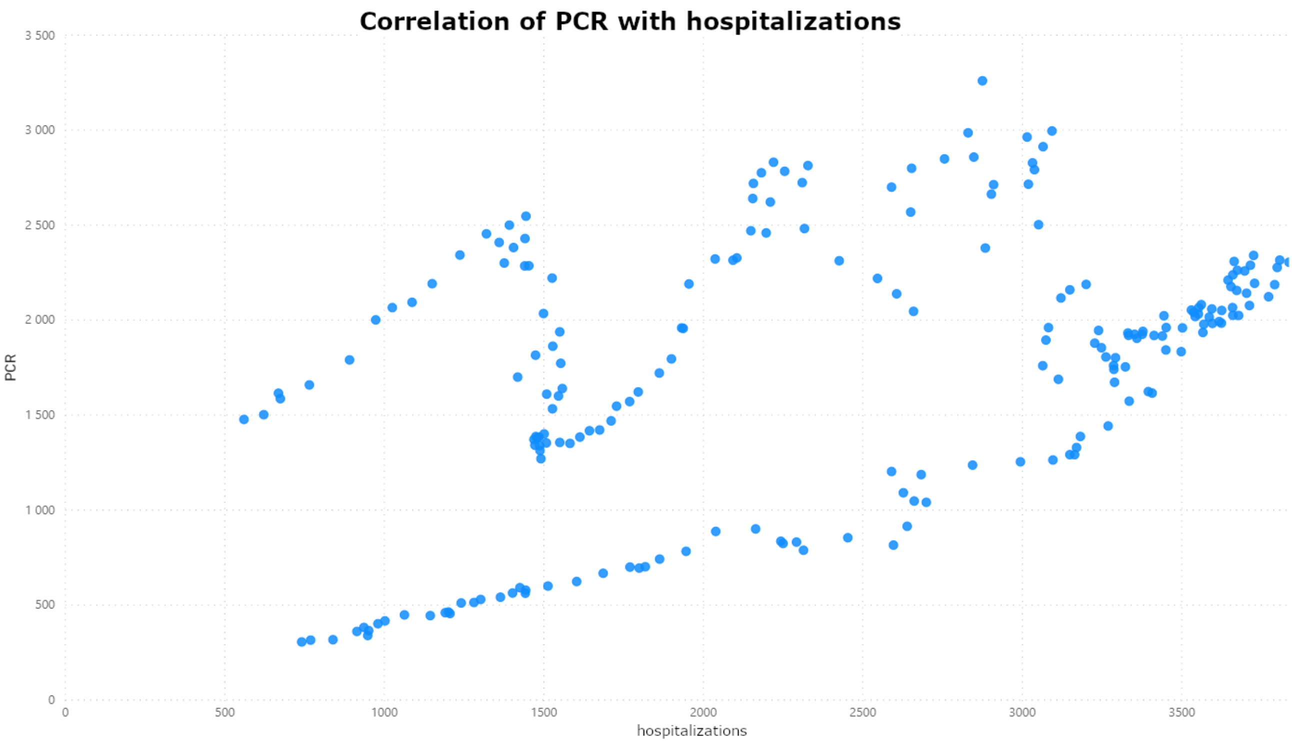
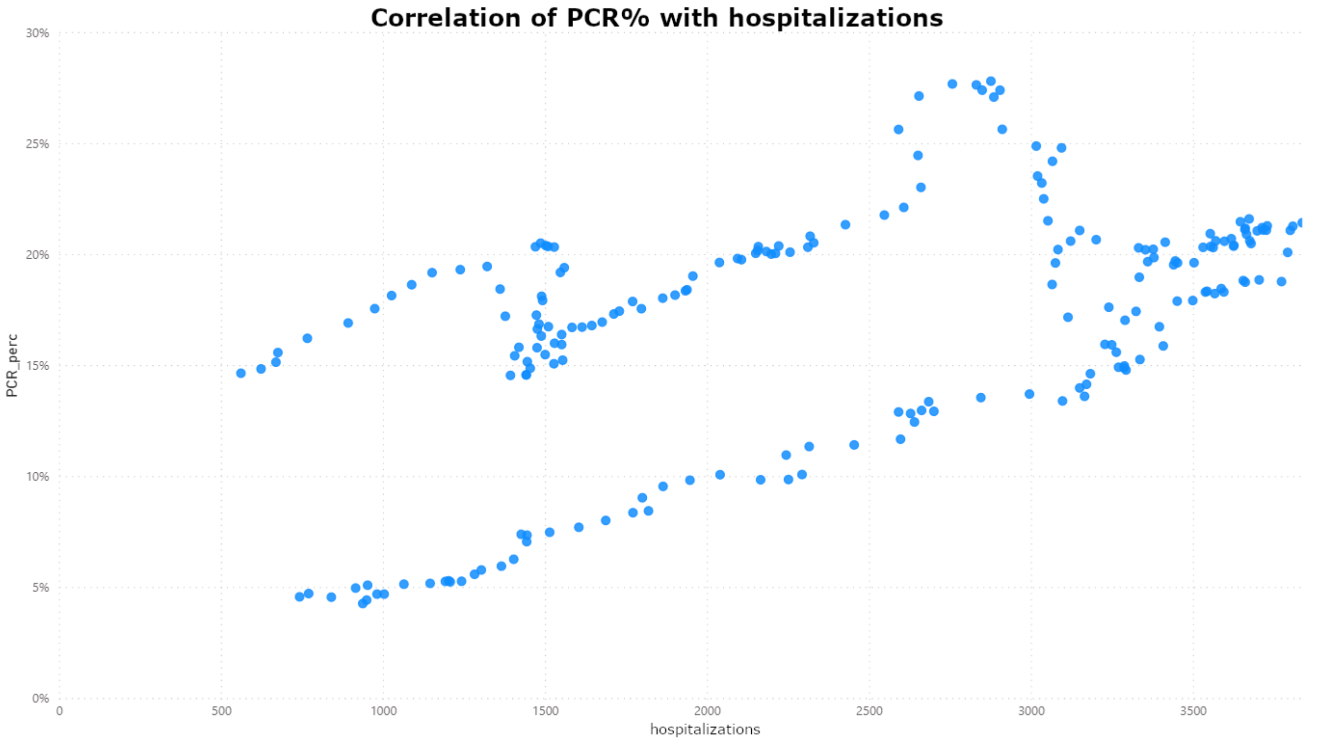
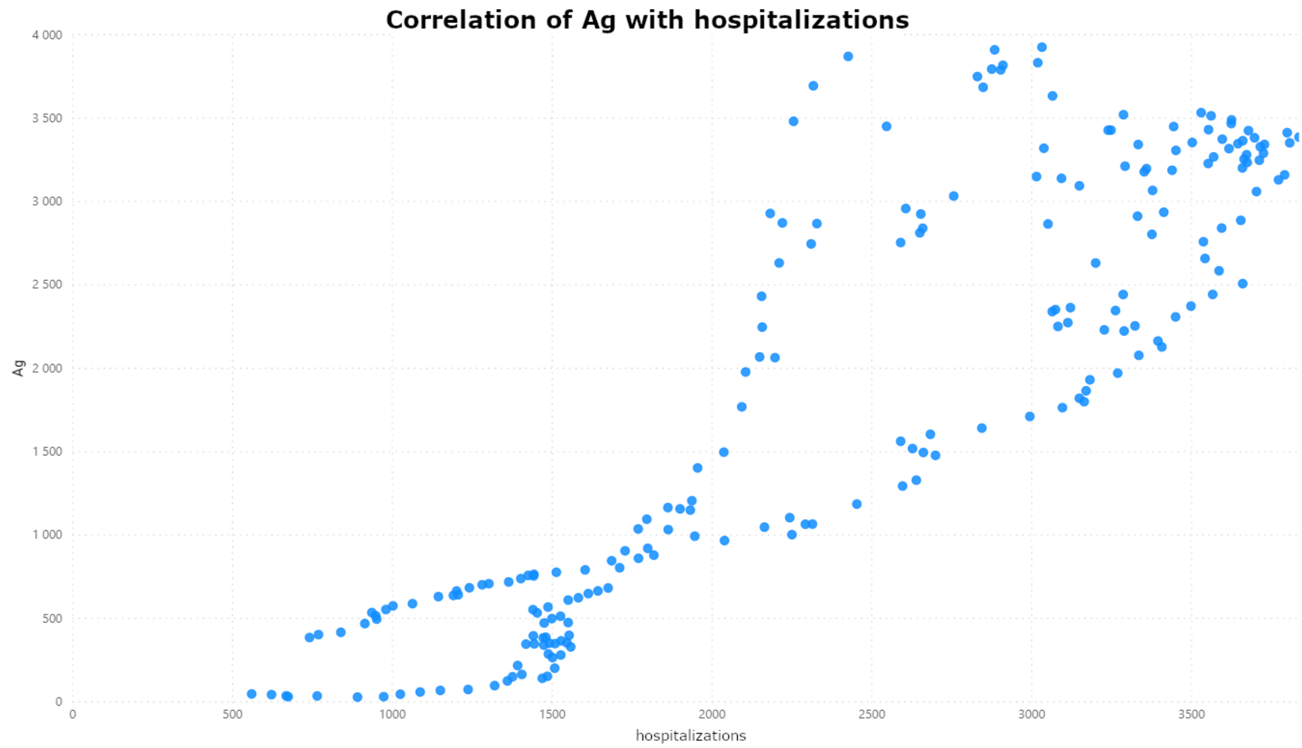


Fig. 8, 9, 10, 11 Basic data overview: correlation of number and percent of positive polymerase chain reaction tests (PCR, PCR%.), number and percent of positive antigen tests (Ag, Ag%) with hospitalizations.




Correlation of PCR% with hospitalizations
30%
25%
20%
15%
10%
5%
0%
PCR_perc
0
500
1000
1500
2000
2500
3000
3500
hospitalizations


Correlation of Ag with hospitalizations
4 000
3 500
3 000
2 500
2 000
1 500
1 000
500
0
Ag
0
500
1000
1500
2000
2500
3000
3500
hospitalizations

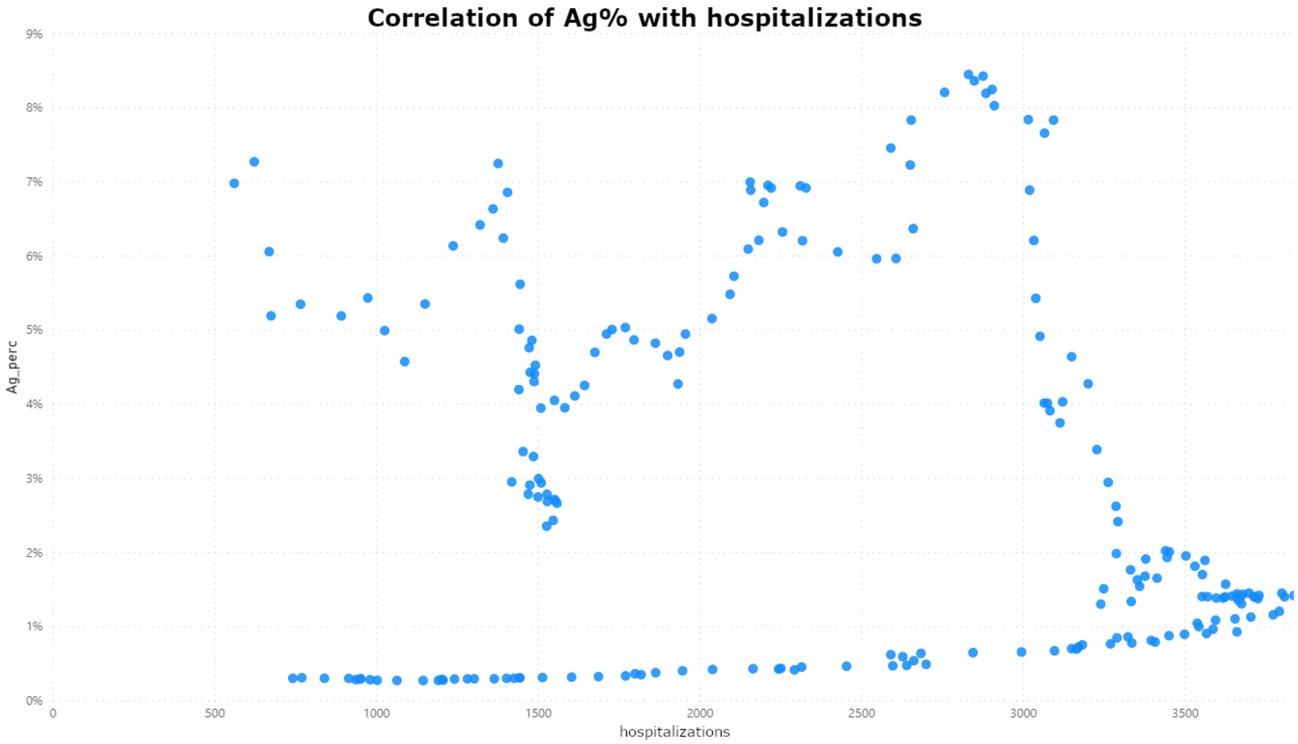
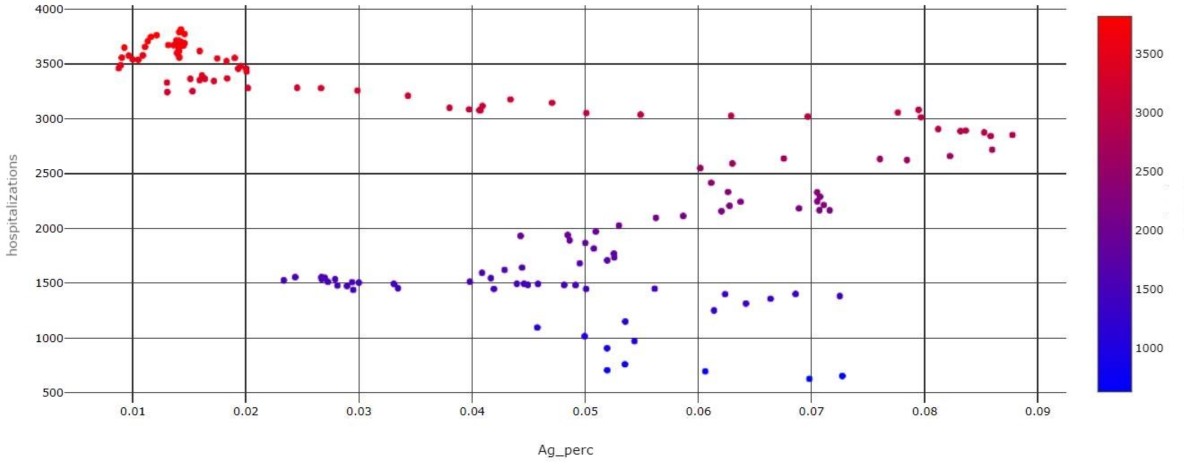
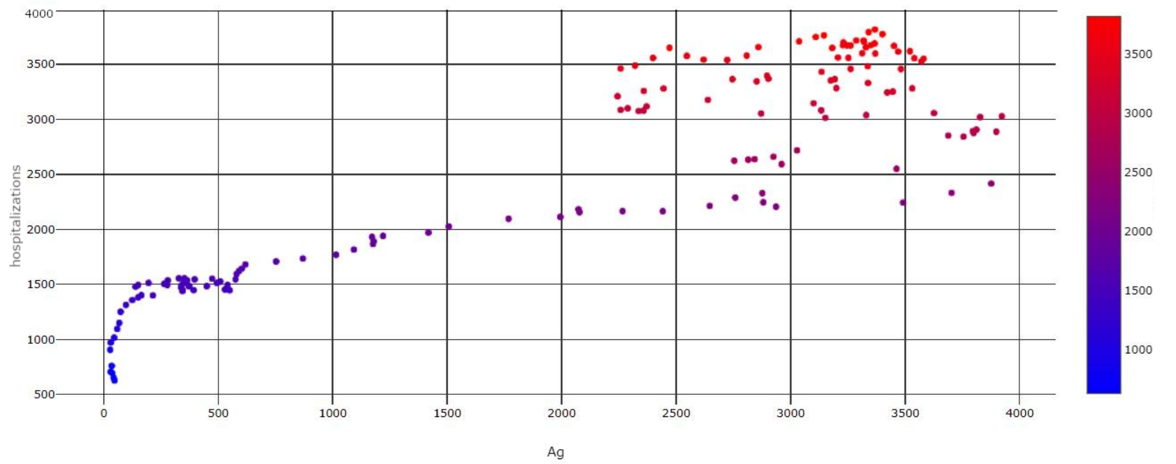


Fig. 12, 13 Importance of Ag perc. and Ag in Ensemble model. Graphs were created in MS Azure:





References

[1] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *J Autoimmun*, vol. 109, p. 102433, May 2020, doi: 10.1016/j.jaut.2020.102433.

[2] A. Lee, "Wuhan novel coronavirus (COVID-19): why global control is challenging?," *Public Health*, vol. 179, pp. A1–A2, Feb. 2020, doi: 10.1016/j.puhe.2020.02.001.

[3] P. B. van Kasteren *et al.*, "Comparison of seven commercial RT-PCR diagnostic kits for COVID-19," *J Clin Virol*, vol. 128, p. 104412, Jul. 2020, doi: 10.1016/j.jcv.2020.104412.

[4] D. B. Larremore *et al.*, "Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening," *Science Advances*, vol. 7, no. 1, p. eabd5393, Jan. 2021, doi: 10.1126/sciadv.abd5393.

[5] L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, and J. Lessler, "Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure," *Ann Intern Med*, pp. M20-1495, May 2020, doi: 10.7326/M20-1495.

[6] S. Mallett *et al.*, "At what times during infection is SARS-CoV-2 detectable and no longer detectable using RT-PCR-based tests? A systematic review of individual participant data," *BMC Med*, vol. 18, no. 1, p. 346, Nov. 2020, doi: 10.1186/s12916-020-01810-8.

[7] N. Sethuraman, S. S. Jeremiah, and A. Ryo, "Interpreting Diagnostic Tests for SARS-CoV-2," *JAMA*, vol. 323, no. 22, p. 2249, Jun. 2020, doi: 10.1001/jama.2020.8259.

[8] W. Wang, J. Tang, and F. Wei, "Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China," *Journal of Medical Virology*, vol. 92, no. 4, pp. 441–447, 2020, doi: 10.1002/jmv.25689.

[9] M. Reilev *et al.*, "Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort," *International Journal of Epidemiology*, vol. 49, no. 5, pp. 1468–1481, Oct. 2020, doi: 10.1093/ije/dyaa140.

[10] C. M. Petrilli *et al.*, "Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study," *BMJ*, vol. 369, p. m1966, May 2020, doi: 10.1136/bmj.m1966.

[11] Z. Car, S. Baressi Šegota, N. Anđelić, I. Lorencin, and V. Mrzljak, "Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron," *Computational and Mathematical Methods in Medicine*, vol. 2020, p. e5714714, May 2020, doi: 10.1155/2020/5714714.

[12] S. Ibrahim, "Performance Evaluation of Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF): COVID-19 Spread and Death Contributing Factors," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 625–631, Sep. 2020, doi: 10.30534/ijatcse/2020/8791.42020.

[13] P. Kairon and S. Bhattacharyya, "Comparative study of variational quantum circuit and quantum backpropagation multilayer perceptron for COVID-19 outbreak predictions," *arXiv:2008.07617 [quant-ph]*, Aug. 2020, Accessed: Jun. 28, 2021. [Online]. Available: http://arxiv.org/abs/2008.07617

[14] S. F. Ardabili *et al.*, "COVID-19 Outbreak Prediction with Machine Learning," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3580188, Apr. 2020. doi: 10.2139/ssrn.3580188.

[15] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, "COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach," *Mathematics*, vol. 8, no. 6, Art. no. 6, Jun. 2020, doi: 10.3390/math8060890.

[16] A. Khakharia *et al.*, "Outbreak Prediction of COVID-19 Patients for Dense and Populated Countries Using Machine Learning," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3655759, Jul. 2020. doi: 10.2139/ssrn.3655759.

[17] R. Tamhane and S. Mulge, *Prediction of COVID-19 Outbreak using Machine Learning*. 2020.

[18] R. Majhi, R. Thangeda, R. P. Sugasi, and N. Kumar, "Analysis and prediction of COVID-19 trajectory: A machine learning approach," *Journal of Public Affairs*, vol. n/a, no. n/a, p. e2537, doi: 10.1002/pa.2537.

[19] P. Parchure *et al.*, "Development and validation of a machine learning-based prediction model for near-term in-hospital mortality among patients with COVID-19," *BMJ Support Palliat Care*, p. bmjspcare-2020-002602, Sep. 2020, doi: 10.1136/bmjspcare-2020-002602.

[20] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study," *Chaos, Solitons & Fractals*, vol. 140, p. 110121, Nov. 2020, doi: 10.1016/j.chaos.2020.110121.

[21] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study," *Chaos, Solitons & Fractals*, vol. 140, p. 110227, Nov. 2020, doi: 10.1016/j.chaos.2020.110227.

[22] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The Computational Limits of Deep Learning," *arXiv:2007.05558 [cs, stat]*, Jul. 2020, Accessed: Aug. 01, 2021. [Online]. Available: http://arxiv.org/abs/2007.05558

[23] N. Ahmed and M. Wahed, "The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research," *arXiv:2010.15581 [cs]*, Oct. 2020, Accessed: Aug. 01, 2021. [Online]. Available: http://arxiv.org/abs/2010.15581

[24] C. Kotas, T. Naughton, and N. Imam, "A comparison of Amazon Web Services and Microsoft Azure cloud platforms for high performance computing," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2018, pp. 1–4. doi: 10.1109/ICCE.2018.8326349.

[25] H. Patel, R. Mankaveettil, R. K. Rajan, N. Challa, R. Maryala, and S. Parsha, "Comparison of AWS and AZURE for COVID-19 Information Retrieval," in *springerprofessional.de*, pp. 304–320. Accessed: Aug. 01, 2021. [Online]. Available: https://www.springerprofessional.de/comparison-of-aws-and-azure-for-covid-19-information-retrieval/19380504

[26] A. Milad, N. I. M. Yusoff, S. A. Majeed, A. N. H. Ibrahim, M. A. Hassan, and A. S. B. Ali, "Using an Azure Machine Learning Approach for Flexible Pavement Maintenance," in *2020 16th IEEE International Colloquium on Signal Processing Its Applications (CSPA)*, Feb. 2020, pp. 146–150. doi: 10.1109/CSPA48992.2020.9068684.

[27] K.-P. Kim and S.-W. Song, "A Study on Prediction of Business Status Based on Machine Learning," *Korea Journal of Artificial Intelligence*, vol. 6, no. 2, pp. 23–27, 2018, doi: 10.24225/kjai.2018.6.2.23.

[28] P. H. Borghi, O. Zakordonets, and J. P. Teixeira, "A COVID-19 time series forecasting model based on MLP ANN," *Procedia Computer Science*, vol. 181, pp. 940–947, Jan. 2021, doi: 10.1016/j.procs.2021.01.250.

[29] V. Nosal, K. Bodova, R. Kollar, and M. Suster, *Determinants of Covid-19 hospitalizations in Slovakia*. 2020. doi: 10.13140/RG.2.2.18148.40325.

[30] J. Barnes, *Azure machine learning. Microsoft Azure Essentials.* Microsoft, 2015.

[31] *Matlab, S.* The MathWorks, Natick, MA., 2012.

[32] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artificial Intelligence in Medicine*, vol. 104, p. 101822, Apr. 2020, doi: 10.1016/j.artmed.2020.101822.

[33] M. Misik, "covid-19-data." Institute for Healthcare Analyses. [Online]. Available: https://github.com/Institut-Zdravotnych-Analyz

[34] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.