medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Predicting community mortality risk due to CoVID-19 using machine learning and development of a prediction tool

Ashis Kumar Das[1#], Shiba Mishra[2], Saji Saraswathy Gopalan[1]

1. The World Bank, Washington DC, USA

2. Credit Suisse Private Limited, Pune, India

Corresponding author:
Ashis Kumar Das, MBBS, MPH, PhD. Email: adas8@worldbank.org

Abstract

Background: The recent pandemic of CoVID-19 has emerged as a threat to global health security. There are a very few prognostic models on CoVID-19 using machine learning.

Objectives: To predict mortality among confirmed CoVID-19 patients in South Korea using machine learning and deploy the best performing algorithm as an open-source online prediction tool for decision-making.

Materials and methods: Mortality for confirmed CoVID-19 patients (n=3,022) between January 20, 2020 and April 07, 2020 was predicted using five machine learning algorithms (logistic regression, support vector machine, K nearest neighbor, random forest and gradient boosting). Performance of the algorithms was compared, and the best performing algorithm was deployed as an online prediction tool.

Results: The gradient boosting algorithm was the best performer in terms of discrimination (area under ROC curve=0.966), calibration (Matthews Correlation Coefficient=0.656; Brier Score=0.013) and predictive ability (accuracy=0.987). The best performer algorithm (gradient boosting) was deployed as the online CoVID-19 Community Mortality Risk Prediction tool named CoCoMoRP (https://ashis-das.shinyapps.io/CoCoMoRP/).

Conclusions: We describe the framework for the rapid development and deployment of an open-source machine learning tool to predict mortality risk among CoVID-19 confirmed patients using publicly available surveillance data. This tool can be utilized by potential stakeholders such as health providers and policy makers to triage patients at the community level in addition to other approaches.

Keywords: CoVID-19; artificial intelligence; modelling; machine learning; mortality risk prediction

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1. Introduction

A novel coronavirus disease 2019 (CoVID-19) originated from Wuhan in China was reported to the World Health Organization in December of 2019.[1] Ever since, this novel coronavirus has spread to almost all major nations in the world resulting in a major pandemic. As of April 18, 2020, it has contributed to more than 2 million confirmed cases and about 150,000 deaths.[2] The first CoVID-19 case was diagnosed in South Korea on January 20, 2020. According to the Korea Centers for Disease Control and Prevention (KCDC), there have been 10,653 confirmed cases and 232 deaths due to CoVID-19 as of April 18, 2020.[3]

In the field of healthcare, accurate prognosis is essential for efficient management of patients while prioritizing care to the more needy. In order to aid in prognosis, several prediction models have been developed using various methods and tools including machine learning.[4,5] Machine learning is a field of artificial intelligence where computers simulate the processes of human intelligence and can synthesize complex information from huge data sources in a short period of time.[6] Though there have been a few prediction tools on CoVID-19, only a handful have utilized machine learning.[7] To the best of our knowledge, by far there is no publicly available CoVID-19 prognosis prediction model from the general population of confirmed cases using machine learning. We attempt to apply machine learning on the publicly available CoVID-19 data at the community level from South Korea to predict mortality.

Our study had two objectives, (1) predict mortality among confirmed CoVID-19 patients in South Korea using machine learning, and (2) deploy the best performing algorithm as an open-source online prediction tool for decision-making.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2. Material and methods

2.1 Patients

Patients for this study were selected from the data shared by Korea Centers for Disease Control and Prevention (KCDC).[3] The timeframe of this study was from the beginning of the detection of the first case (January 20, 2020) through April 07, 2020. In the dataset, there were a total of 3,128 patients. Our inclusion criteria were confirmed CoVID-19 cases with availability of socio-demographic, exposure and diagnosis confirmation features along with the outcome. We excluded patients those had missing features – sex (n=94) and age (n=12), and thus, 3,022 patients were included in the final analysis.

2.2 Outcome variable

The outcome variable was mortality and it had a binary distribution – "yes" if the patient died, or "no" otherwise.

2.3 Predictors

The predictors were individual patient level socio-demographic and exposure features. They were age group, sex, province, date of diagnosis, and exposure. There were ten age groups as follows below 10 years, 10-19 years, 20-29 years, 30-39 years, 40-49 years, 50-59 years, 60-69 years, 70-79 years, 80-89 years, 90 years and above. Patients represented all 17 provinces of South Korea. Dates of confirmation of CoVID-19 status were converted to weeks and they were 20-26 Jan 2020, 27 Jan-02 Feb 2020, 03-09 Feb 2020, 10-16 Feb 2020, 17-23 Feb 2020, 24 Feb-01 Mar 2020, 02-08 Mar 2020, 09-15 Mar 2020, 16-22 Mar 2020, 23-29 Mar 2020, 30 Mar-07 Apr 2020. Patients were exposed in several settings, such as nursing home, hospital, religious

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

gathering, call center, community center, shelter and apartment, gym facility, overseas inflow, contact with patients and others.

2.4 Statistical Methods

2.4.1 Descriptive Analysis

We performed descriptive analyses of the predictors by respective sub-groups and present the results as numbers and proportions. Potential correlations between predictors were tested with Pearson's correlation coefficient.

2.4.2 Predictive Analysis

We applied machine learning algorithms to predict mortality among CoVID-19 confirmed cases. Machine learning is a branch of artificial intelligence where computer systems can learn from available data and identify patterns with minimal human intervention.[8] Typically, in machine learning several algorithms are tested on data and performance metrics are used to select the best performing algorithm. We tested five commonly used supervised machine learning algorithms in healthcare research (logistic regression, support vector machine, K neighbor classification, random forest and gradient boosting) to compare algorithm performance efficiency. Logistic regression is best suited for a binary or categorical output. It tries to describe the relationship between the output and predictor variables.[9] In support vector machine (SVM) algorithm, the data is classified into two classes based on the output variable over a hyperplane.[9] The algorithm tries to increase the distance between the hyperplane and the most proximal two data points in each class. K Nearest Neighbors (KNN) is a non-parametric approach that decides the output classification by the majority class among its neighbors.[10] The number of neighbors can be altered to arrive at the best fitting KNN model. For our model, we selected 20 nearest

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

neighbors. Random forest algorithm uses a combination of decision trees.[11] Decision trees are generated by recursively partitioning the predictors. New attributes are sequentially fitted to predict the output. We used an ensemble of 501 decision trees with the trees extended up to a maximum depth of 10. Gradient boosting algorithm uses a combination of decision trees.[12] Each decision tree dynamically learns from its precursor and passes on the improved function to the following. Finally, the weighted combination of these trees provides the prediction.

2.4.3 Evaluation of the performance of the algorithms

We split the data into training (80 percent) and validation cohorts (20 percent). Initially, the algorithms were trained on the training cohort and then were validated on the validation cohort for determining predictions. The data was passed through a 10-fold cross validation where the data was split into training and validation cohorts at 80/20 ratio randomly ten times. The final prediction came out of the cross-validated estimate. As our data was imbalanced (only 2% output were with the condition against 98% without), we applied an oversampling technique called synthetic minority oversampling technique (SMOTE) to enhance the learning on the test data.[13,14]

The performance of the algorithms were evaluated for discrimination, calibration and overall performance. Discrimination is the abillity of the algorithm to separate out patients with the mortality risk from those without, where as calibration is the agreement between observed and predicted risk of mortality. An ideal model should have the best of both discrimination and calibration. We tested discriminaiton with area under the receiver operating characteristics curve (AUC) and calibration with accuracy and Matthews correlation coefficient. A receiver operator characteristic (ROC) curve plots the true positive rate on y-axis against the false positive rate on x-axis.[15] AUC is score that measures the area under the ROC curve and it ranges from 0.50 to

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1.0 with higher values meaning higher discrimination. Accuracy is a measure of correct classification of death cases as death and survived cases as survived.[15] Matthews correlation coefficient (MCC) is a measure that takes into account all four predictive classes – true positive, true negative, false positive and false negative.[16] It is considered a better measure than accuracy for unbalanced data. Brier score simultaneously account for discrimnation and calibration.[15] A smaller Brier score indicates better performance. In addition, the gradient boosting algorithm was used to estimate the relative contributions of the predictors and draw the variable importance plot.[17]

The statistical analyses were performed using Stata Version 15 (StataCorp LLC. College Station, TX), Python programming language Version 3.7 (Python Software Foundation, Wilmington, DE, USA) and R programming language Version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). The web application was built using the Shiny package for R and deployed with Shiny server.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3. Results

3.1 Patient profile

The profile of the patients is presented in Table 1. Out of 3,022 confirmed patients, a slightly more than half were females (56.3%). Among the age groups, the maximum patients were from 20-29 years (23.9%), followed by 50-59 years (18.6%), 40-49 years (14.3%), 30-39 years (12.9%) and 60-69 years (12.1%). Gyeongsangbuk-do (38.8%), Gyeonggi-do (19.7%) and Seoul (15.8%) provinces together presented the maximum patients. Considering the source/mode of infection, the largest group had unknown mode (44%) followed by direct contact with patients (27.8%) and from overseas (13.7%). According to this available data source, 60 percent of the patients were confirmed of their diagnosis between 24 February and 15 March of 2020. There were 61 deaths accounting for 2 percent of the patients.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. Sample characteristics (N=3,022)

Variable	Number	Proportion (%)
Sex		
Female	1,700	56.3
Male	1,322	43.8
Age group (years)		
Below 10	43	1.4
10-19	126	4.2
20-29	721	23.9
30-39	389	12.9
40-49	432	14.3
50-59	562	18.6
60-69	366	12.1
70-79	189	6.3
80-89	149	4.9
90 and above	45	1.5
Province		
Busan	122	4.0
Chungcheongbuk-do	44	1.5
Chungcheongnam-do	136	4.5
Daegu	63	2.1
Daejeon	39	1.3
Gangwon-do	35	1.2
Gwangju	27	0.9
Gyeonggi-do	595	19.7
Gyeongsangbuk-do	1,173	38.8
Gyeongsangnam-do	107	3.5
Incheon	80	2.7
Jeju-do	9	0.3
Jeollabuk-do	15	0.5
Jeollanam-do	15	0.5
Sejong	46	1.5
Seoul	476	15.8
Ulsan	40	1.3
Exposure		
Nursing home	45	1.5
Hospital	37	1.2
Religious gathering	158	5.2
Call center	112	3.7
Community center, shelter and apartment	50	1.7
Gym facility	34	1.1
Overseas inflow	414	13.7

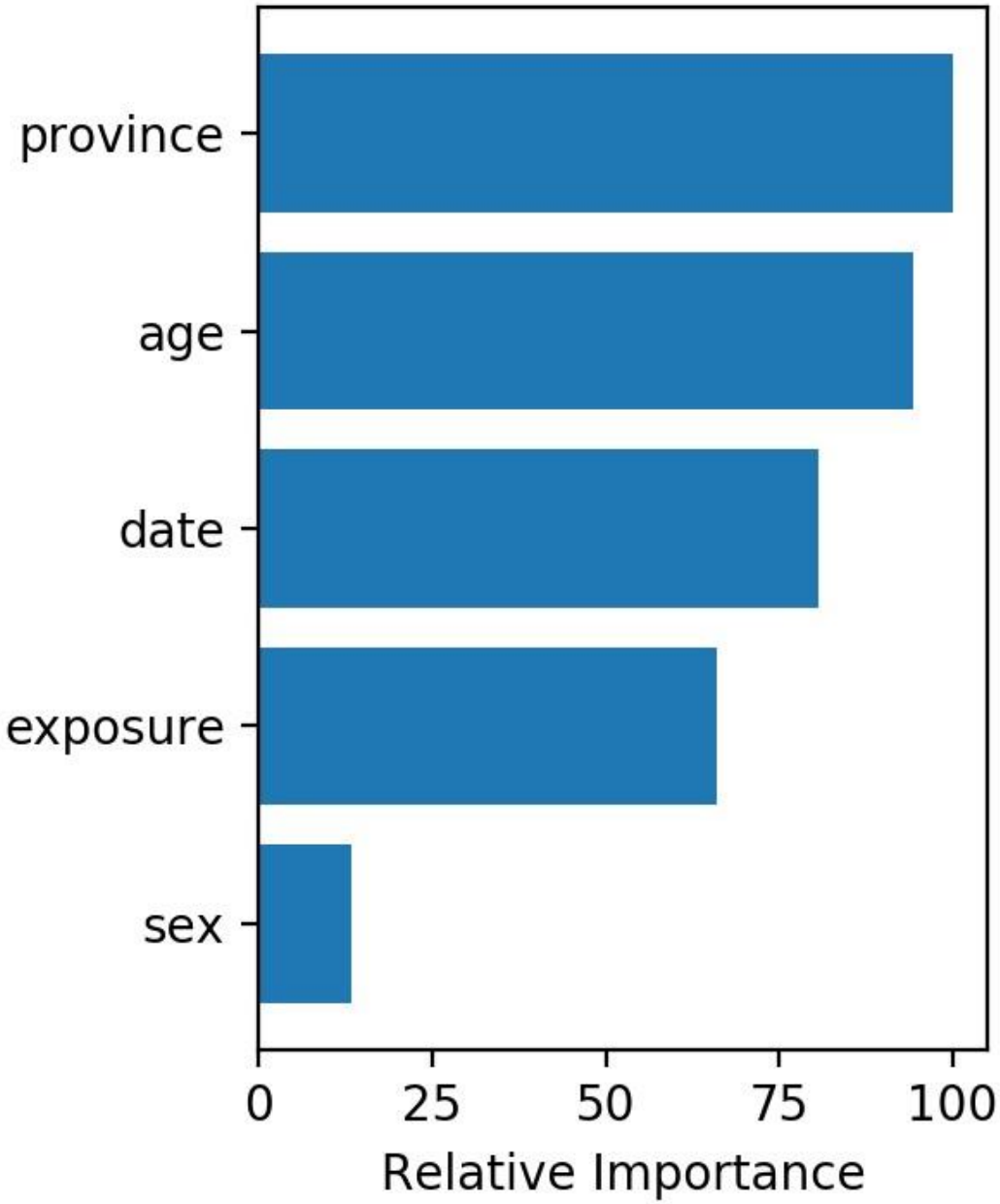
medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Contact with patients	841	27.8
Others	1,331	44.0
Date of confirmed diagnosis		
20-26 Jan 2020	2	0.1
27 Jan-02 Feb 2020	10	0.3
03-09 Feb 2020	12	0.4
10-16 Feb 2020	4	0.1
17-23 Feb 2020	186	6.2
24 Feb-01 Mar 2020	697	23.1
02-08 Mar 2020	733	24.3
09-15 Mar 2020	380	12.6
16-22 Mar 2020	354	11.7
23-29 Mar 2020	321	10.6
30 Mar-07 Apr 2020	323	10.7
Outcome		
Survived	2,961	98.0
Died	61	2.0

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Using the gradient boosting algorithm, we estimated the relative importance of the predictors (figure 1). Province was the most important predictor followed by age, date, exposure and sex.

Figure 1. Relative importance of predictors



medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3.2 Performance of the algorithms

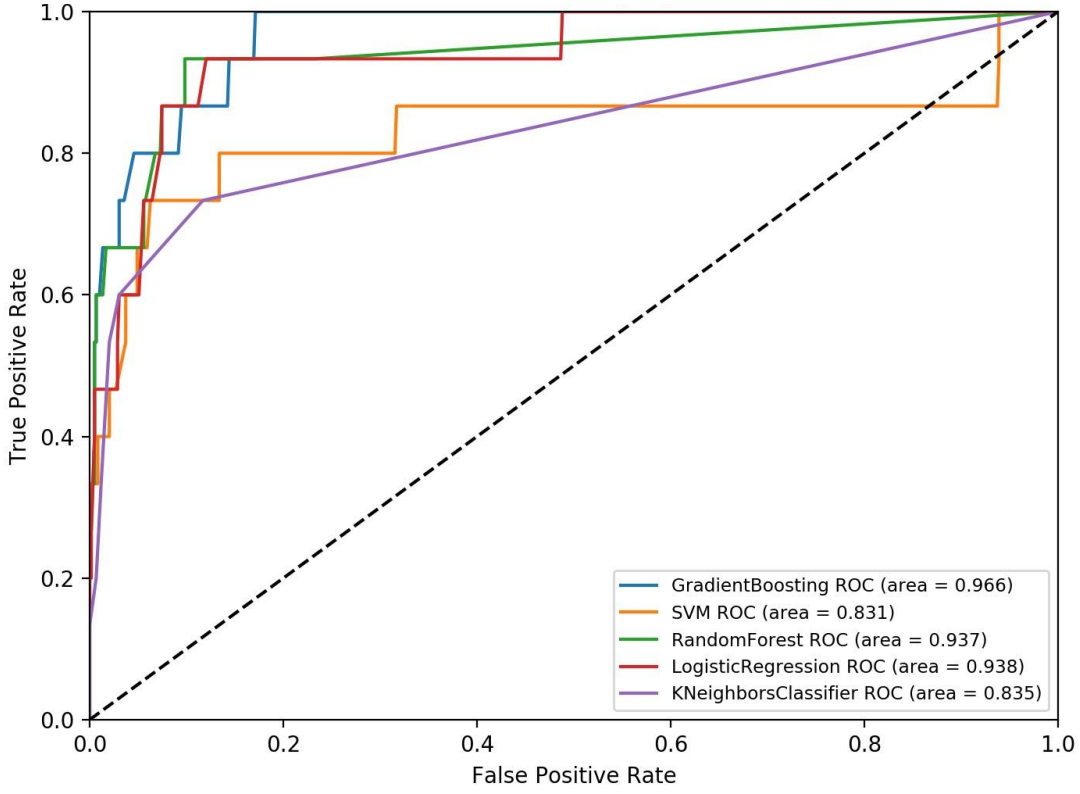
Table 2 present the performance metrics of all algorithms – logistic regression, support vector machine, K nearest neighbor, random forest and gradient boosting. The accuracy of all algorithms was very similar with the gradient boosting performing the best (0.987) and KNN with the least score (0.979). Similarly, gradient boosting performed the best on Matthews correlation coefficient (highest score) and Brier score (lowest score). Further, figure 2 shows the area under receiver operating characteristic curve (AUC) for all algorithms. The AUC ranged from 0.831 to 0.966 with the best score for the gradient boosting algorithm. Considering all the performance metrics, gradient boosting was the best performing algorithm.

Table 2. Performance of the algorithms with test data

Metrics	**Logistic regression**	**Support vector machine**	**K nearest neighbor**	**Random forest**	**Gradient boosting**
Area under ROC curve	0.938	0.831	0.835	0.937	0.966
Accuracy	0.983	0.983	0.979	0.985	0.987
Matthews correlation coefficient	0.530	0.584	0.264	0.593	0.656
Brier score	0.017	0.017	0.021	0.015	0.013

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2. ROC and AUC of machine learning algorithms

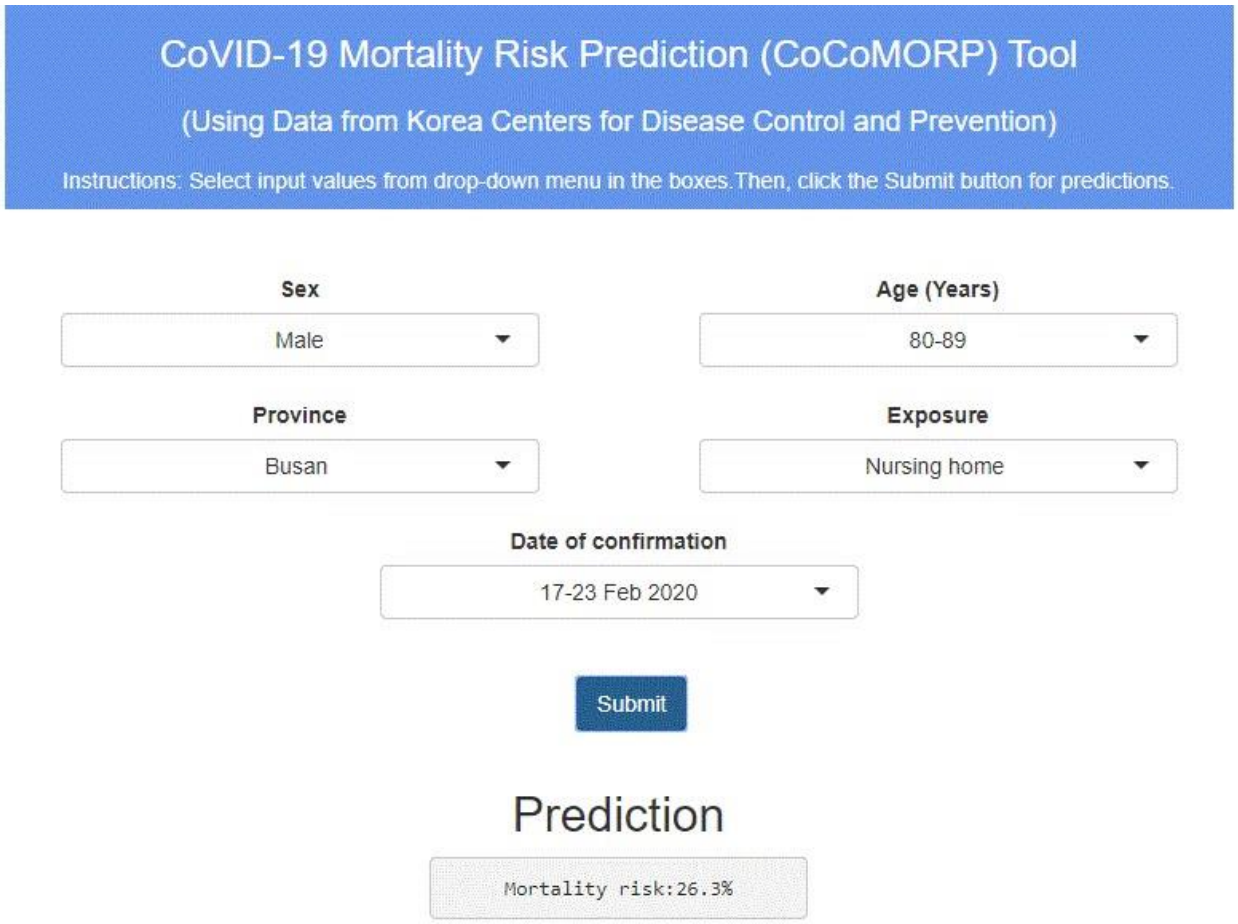


medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3.3 Online CoVID-19 mortality risk prediction tool – CoCoMoRP

The best performing model – gradient boosting was deployed as the online mortality risk prediction tool named as "**Co**VID-19 **Co**mmunity **Mo**rtality **R**isk **P**rediction" – CoCoMoRP" (https://ashis-das.shinyapps.io/CoCoMoRP/). Figure 3 presents the user interface of the prediction tool. The web application is optimized to be conveniently used on multiple devices such as desktops, tablets, and smartphones.

Figure 3. CoCoMORP online **Co**VID-19 **Co**mmunity **Mo**rtality **R**isk **P**rediction tool



medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The user interface has five boxes to select input features as drop-down menus. The features are sex (two options – male and female), age (ten options – below 10 years, 10-19 years, 20-29 years, 30-39 years, 40-49 years, 50-59 years, 60-69 years, 70-79 years, 80-89 years, 90 years and above), province (all 17 provinces – Busan, Chungcheongbuk-do, Chungcheongnam-do, Daegu, Daejeon, Gangwon-do, Gwangju, Gyeonggi-do, Gyeongsangbuk-do, Gyeongsangnam-do, Incheon, Jeju-do, Jeollabuk-do, Jeollanam-do, Sejong, Seoul, Ulsan), exposure (nine options – nursing home; hospital; religious gathering; call center; community center, shelter and apartment; gym facility; overseas inflow; contact with patients; and others) and date of confirmation in weeks (eleven options – 20-26 Jan 2020, 27 Jan-02 Feb 2020, 03-09 Feb 2020, 10-16 Feb 2020, 17-23 Feb 2020, 24 Feb-01 Mar 2020, 02-08 Mar 2020, 09-15 Mar 2020, 16-22 Mar 2020, 23-29 Mar 2020, 30 Mar-07 Apr 2020).

The user has to select one option each from the input feature boxes and click the submit button to estimate the CoVID-19 mortality risk probability in percentages. For instance, the tool gives a CoVID-19 mortality risk prediction of 26.3% for a male patient aged between 80 and 89 years from Busan province with exposure in a nursing home who got confirmation of diagnosis during the week of 17-23 February 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4. Discussion

The CoVID-19 pandemic is a threat to global health and economic security. Recent evidence for this new disease is still evolving on various clinical and socio-demographic dimensions.[18–20] Simultaneously, health systems across the world are constrained with resources to efficiently deal with this pandemic. We describe the rapid development and deployment of an open-source artificial intelligence tool to predict mortality risk among CoVID-19 confirmed patients using publicly available surveillance data. This tool can be utilized by potential stakeholders such as health providers and policy makers to triage patients at the community level in addition to other approaches.

One major limitation of this tool is unavailability of crucial clinical information on symptoms, risk factors and clinical parameters. Recent research has identified certain symptoms, preexisting illnesses and clinical parameters as strong predictors of prognosis and severity of progression for CoVID-19.[20–22] These crucial pieces of information are not publicly available so far in the surveillance data, so the tool could not be tested to include these features. Inclusion of these additional features may improve the reliability and relevance of the tool. Therefore, we urge the users to balance the predictions from this tool against their own and/or health provider's clinical expertise and other relevant clinical information.

To the best of our knowledge, our CoVID-19 community mortality risk prediction tool is the first of its kind. Our tool offers an additional approach to informing decision making for CoVID-19 patients. We believe our experience of rapidly developing a mortality risk prediction tool during

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

a crisis using limited data will guide future development of similar approaches using locally available data during epidemics and other disasters.

Authors' contributions

Conceived and designed this study: Ashis Kumar Das, Shiba Mishra, Saji Saraswathy Gopalan

Analyzed and explained the data: Ashis Kumar Das, Shiba Mishra, Saji Saraswathy Gopalan

All authors contributed to the writing and approved the final manuscript.

Acknowledgements

We are grateful to Korea Center for Disease Control and Prevention for making this data publicly available.

Declaration of competing interest

The authors declare that there is no conflict of interest. The views expressed in the paper are that of the authors and do not reflect that of their affiliations. This particular work was conducted outside of the authors' organizational affiliations.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

[1] WHO. WHO Coronavirus disease (COVID-2019) situation reports 2020 n.d.

[2] COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) n.d. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6. (Accessed on April 18, 2020)

[3] KCDC. Korea Centers for Disease Control and Prevention; Seoul, Korea: 2020. The updates on COVID-19 in Korea as of 18 April.

[4] Chen JH, Asch SM. Machine learning and prediction in medicine-beyond the peak of inflated expectations. N. Engl. J. Med. 2017 Jun 29;376(26):2507-2509.

[5] Qu Y, Yue G, Shang C, Yang L, Zwiggelaar R, Shen Q. Multi-criterion mammographic risk analysis supported with multi-label fuzzy-rough feature selection. Artif Intell Med 2019;100:101722.

[6] Benke K, Benke G. Artificial intelligence and big data in public health. Int J Environ Res Public Health 2018;15(12).

[7] Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328.

[8] Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920-30.

[9] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: Past, present and future. Stroke Vasc Neurol 2017;2(4):230-243.

[10] Raeisi Shahraki H, Pourahmad S, Zare N. K Important Neighbors: A Novel Approach to Binary Classification in High Dimensional Data. Biomed Res Int 2017;7560807.

[11] Rigatti SJ. Random Forest. J Insur Med 2017; 47(1):31-39.

[12] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21.

[13] Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002;6:321–357.

[14] Nnamoko N, Korkontzelos I. Efficient treatment of outliers and class imbalance for diabetes prediction. Artif Intell Med 2020;104:101815.

[15] Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc 2020;27(4):621-633.

[16] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020; 21(1):6.

[17] Xie J, Coggeshall S. Prediction of transfers to tertiary care and hospital mortality: A gradient boosting decision tree approach. Stat Anal Data Min 2010;3(4).

[18] Sun P, Lu X, Xu C, Sun W, Pan B. Understanding of COVID-19 based on current evidence. J Med Virol 2020. https://doi.org/10.1002/jmv.25722.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.27.20081794; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[19] Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet 2020; 395(10226):809-815.

[20] Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol 2020. https://doi.org/10.1007/s00392-020-01626-9.

[21] Li L quan, Huang T, Wang Y qing, Wang Z ping, Liang Y, Huang T bi, et al. 2019 novel coronavirus patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. J Med Virol 2020. https://doi.org/10.1002/jmv.25757.

[22] Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020. https://doi.org/10.1056/nejmoa2002032.