

Influenza and Anosmia: important prediction factors for severity and death of COVID-19

Doo Hwan Kim, MD^{1*}, Eun Jung Lee, PhD^{2,4*}, Min Gul Kim, PhD^{3,4,6}, Seong J. Yang, PhD^{5*}, Sang Woo Yeom, MS^{2,3*}, Min Hee Lee, MD⁷, Sang Jae Noh, PhD⁸, Yeon Seok Yoo, MD², Jong Seung Kim, PhD^{2,3,4}

¹Director of Big-Data Center, National Health Insurance Service (NHIS), Wonju, Republic of Korea

²Department of Otorhinolaryngology-Head and Neck Surgery, College of Medicine, Jeonbuk National University, Jeonju, Republic of Korea

³Department of Medical Informatics, College of Medicine, Jeonbuk National University, Jeonju, Republic of Korea

⁴Research Institute of Clinical Medicine of Jeonbuk National University – Biomedical Research Institute of Jeonbuk National University Hospital, Jeonju, Republic of Korea

⁵Department of Statistics (Institute of Applied Statistics), Jeonbuk National University, Jeonju, Republic of Korea

⁶Department of Pharmacology, Jeonbuk National University, Jeonju, Republic of Korea

⁷Department of Internal Medicine, Division of Allergy and Pulmonology, Presbyterian Medical Center, Jeonju, Republic of Korea

⁸Department of Forensic Medicine, College of Medicine, Jeonbuk National University, Jeonju, Republic of Korea

*These authors contributed equally to this study.

Correspondence: Jong Seung Kim, MD, PhD

Department of Medical Informatics, Department of Otorhinolaryngology, Jeonbuk National University, Jeonju, Republic of Korea; 54907, Korea, Tel: +82-63-250-2792; Fax: +82-63-250-1986; E-mail: kjsjdk@gmail.com

Running title: Machine learning model prediction of severity and death of COVID-19 patients in South Korea

Text word count: 3432 words

Number of tables and/or figures: 1/1

Number of videos: 0

Funding Statement: None

Conflict of Interests: None

Abbreviations: NHIS, National Health Insurance Service; ML, Machine learning; IRB, Institutional Review Boards; ICU, Intensive care unit; ECMO, Extracorporeal membrane oxygenation; HTN, Hypertension; ICD, International Classification of Disease; DM, Diabetes mellitus; ARB, Angiotensin Receptor Blocker; GERD, Gastroesophageal reflux disease; A_sinusitis, Acute sinusitis; C_sinusitis, Chronic sinusitis; BMD, bone mineral density; CVD, Cardiovascular disease; PVD, Peripheral vascular disease; CHF, Congestive heart failure; RA, rheumatologic disease; MI, Myocardial infarction; IBD, Inflammatory bowel disease; NTM, Non-tuberculosis mycobacterium; EN, elastic net; SVM, support vector machine; AUC, area under the ROC curve; RSS, residual sum of squares; BD, binomial deviance; SD, standard deviation; ACE, angiotensin converting enzyme; FVC, forced vital capacity; FEV1, forced

exhalation volume in 1 second

Author contributions

EJL, JSK, DHK, MHL contributed to the study design, protocol and study materials. DHK, SWY, JSK collected study data. MGK, SWY, SJN, SJY designed the statistical plan and data analysis. JSK, SWY, YSY performed the statistical analysis. EJL, MGK, JSK, SJY, YSY wrote the first draft of the manuscript. All authors contributed to interpretation of the data and revision of the manuscript.

Financial support: None

Conflicts of Interest: None

Funding: None

Level of Evidence: level 3

Keywords: Machine Learning; Modeling; COVID 19; SARS-CoV-2; Mortality; Severity; Comorbidities

Abstract

Many people are being hospitalized and are dying from COVID-19 as the right treatment has not yet been identified. We investigate the factors related to the severity and mortality of COVID-19 using big data-machine learning techniques. This retrospective study included 8070 SARS-CoV-2 confirmed patients of the 129,120 SARS-CoV-2 RNA tested patients in South Korea between January and July 2020, and whose data were available from the National Health Insurance Service. Primary endpoint was comorbidity, severity and mortality rate in COVID-19. Machine learning algorithms were performed to evaluate the effects of comorbidities on severity and mortality rate of COVID-19. The most common comorbidities of COVID-19 were pulmonary inflammation followed by anosmia. The model that best predicted severity was a neural network (AUC: 85.06%). The most important variable for predicting severity in the neural network model was a history of influenza (relative importance: 0.129). The model that best predicted mortality was the logistic regression elastic net (EN) model (AUC: 93.86%). The most important variables for mortality in the EN model were age (coefficient: 2.136) and anosmia (coefficient: -1.438). Through the state-of-the-art machine learning algorithm and 8070 patients of COVID-19 patients in South Korea, influenza was found to be a major adverse factor in addition to old age and male sex. In addition, anosmia was found to be a major factor associated with lower severity and mortality rates. The patient's history of influenza and anosmia will be an important indicator for predicting the severity and mortality of COVID-19 patients.

INTRODUCTION

At November 27 2020, 60 million patients had been infected with COVID-19 worldwide, and about 1.4 million had died from the disease.¹ This virus belongs to the same coronavirus family as the MERS virus that circulated in 2015, but it is much more infectious, and the world is currently experiencing a pandemic.² However, the factors affecting disease severity and mortality have not yet been clearly identified.

In Korea, which has had widespread medical insurance coverage since 1990 (National Health Insurance Service; NHIS), almost all citizens including all patients diagnosed with COVID-19 have benefiting from this coverage. The NHIS has the advantage of being able to provide information on underlying diseases as well as the demographic factors of all the COVID-19 patients.

Machine learning (ML) was developed back in the 1970s, but its widespread application has only been in the spotlight since 2010 once appropriate computer technology had evolved. Developed by statisticians, but unlike statistics, it is an algorithm where prediction is the most important outcome. Therefore, it is appropriate to use the ML algorithm to try to predict the mortality and severity of COVID-19.

In this paper, we have analyzed the factors affecting the severity and mortality of COVID-19 patients registered in the NHIS of South Korea using ML algorithms.

METHODS

Ethical considerations

This study passed the institutional review board (IRB) of Jeonbuk National University Hospital (no. 2020-04-067). All methods were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained from all subjects.

Data sources

The NHIS data for this study included 8070 patients diagnosed with COVID-19 between 1 January and 30 July 2020 in South Korea. The NHIS data provided the patients' history of use of medical services (hospital visits, billing data) from 2015 to 2020. (Figure 1A) This data is in the form of combining data from the Korea Centers for Disease Control and the National Health Insurance Service, and included 129120 patients who underwent SARS-CoV-2 RNA test from hospital admission to death from January 1, 2020 to May 1, 2020. Of the 129,120 SARS-CoV-2 RNA tested patients, we enrolled 8070 SARS-CoV-2 confirmed patients.

Definition of severity and mortality of COVID-19

The severity of COVID-19 was defined as the end result with one of following conditions. 1) Intensive care unit (ICU) care; 2) Extracorporeal membrane oxygenation (ECMO) treatment; 3) Mechanical ventilator care; 4) high flow oxygen supply.

The mortality of COVID-19 was also checked because the NHIS data was connected to the Korea Disease Control and Prevention Agency and Statistics Korea, which has the mortality data.

Comorbidity of COVID-19

A total of 21 diseases were chosen as the underlying diseases in the 8070 COVID-19 patients. NHIS-customized data for the past 5 years were selected for the patients confirmed with COVID-19, and hospital use records for the past 5 years were used to identify the following inclusion criteria.

Hypertension (HTN): a patient with 1) high blood pressure ICD-10 code (I10, I101, I109) and 2) taking antihypertensive drugs at least twice every 3 months.

Diabetes mellitus (DM): a patient with 1) diabetes ICD-10 code (E10, E11, E12, E13, E14, O24) and 2) taking antidiabetic drugs at least twice every 3 months.

Influenza: a patient who 1) has had the influenza ICD-10 code (J10, J11) and 2) took the influenza medication 3) a positive result of a real-time RT-PCR assay of nasal or pharyngeal swabs

Cancer: patients with an ICD-10 code for any type of cancer.

Pulmonary disease: patients with an ICD-10 code for lower respiratory tract disease (asthma, COPD, bronchitis, bronchiolitis).

Angiotensin Receptor Blocker (ARB) among hypertensive patients: 1) patients taking angiotensin converting enzyme inhibitor or ARB; 2) patients who took these drugs at least twice every 3 months.

Gastroesophageal reflux disease (GERD): 1) patients who have an ICD-10 code (K21) for GERD and 2) who had undergone laryngoscopy or gastroscopy in their local practice.

Acute sinusitis (A_sinusitis): a patient whose ICD-10 code (J01) for acute sinusitis was present more than twice every 3 months.

Chronic sinusitis (C_sinusitis): a patient whose ICD-10 code (J32) for chronic sinusitis was present more than twice every 3 months.

Osteoporosis: 1) patients with osteoporosis ICD-10 code (M80, M81) and 2) bone mineral density (BMD) in their treatment.

Cardiovascular disease (CVD): a patient whose ICD-10 code for CVD (G45, G46, I60 to I63) was present more than twice every 3 months.

Angina: 1) a patient with an ICD-10 code for angina (I20) and 2) who had received an exercise (tolerance) test.

Peripheral vascular disease (PVD): patients with PVD ICD-10 code (I70, I71).

Congestive heart failure (CHF): 1) Patients with ICD-10 codes for CHF (I43, I50) and 2) who had received an echocardiogram for diagnosis or treatment.

Depression: a patient with ICD-10 code (F32 to F34) for depression.

Rheumatologic disease (RA): patients with ICD-10 code (M05, M06, M32 to M34) for rheumatologic disease.

Hepatitis: 1) Patients who have ICD-10 code (K73) for liver disease and 2) who underwent abdominal ultrasonography.

Myocardial infarction (MI): 1) Patients who have ICD-10 codes (I21, I22) for myocardial infarction and 2) who underwent an echocardiogram for diagnosis or treatment.

Inflammatory bowel disease (IBD): a patient with ICD-10 code (K50, K51) for IBD.

Non-tuberculosis mycobacterium (NTM): a patient with ICD-10 code (A31) for NTM.

Anosmia (olfactory loss): patients diagnosed with ICD-10 code (R430) for anosmia; this only

targeted patients who were diagnosed after their COVID-19 confirmation diagnosis; anosmia was only included in patients who had a 5-year wash-out period.

Machine Learning (ML) modeling

To apply the ML model, we selected 24 independent variables: HTN, DM, Influenza, Cancer, Pulmonary, ARB, GERD, A_Sinusitis, C_Sinusitis, Osteoporosis, CVD, Angina, PVD, CHF, Depression, RA, Hepatitis, MI, IBD, NTM, Anosmia, Age, Sex, Residential area.

Our interest was in predicting and identifying underlying diseases and demographic factors related to the severity and mortality of COVID-19 patients diagnosed from January 1, 2020 to July 30, 2020 in Korea.

We used several ML algorithms. First, a logistic regression model was used. The logistic regression model is an analytical method commonly used in medicine to identify independent variables that affect the result when the dependent variable has a binary classification such as 0 or 1. In this study, lasso, ridge and elastic net (EN) methods were used.

$$l(\beta) - \lambda \left[\frac{(1 - \alpha)}{2} \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j| \right],$$

Here, if $\alpha=0$, it is ridge, if $\alpha=1$, it is lasso, and if $0<\alpha<1$, it is elastic net. To do this, we selected λ using fivefold cross-validation for each α value, and used R's "glmnet" package (www.r-project.org).

Second, a tree-based model was selected. One decision tree model is easy to explain to a clinician, but it has the disadvantage of low prediction power. In this study, we applied the bagging and random forest (RF) models using multiple trees. The multiple tree model

eventually improved the classification and prediction performance. This was applied using the “randomForest” package in R, and the number of trees was calculated by increasing bagging and RF to 500.

We also used other ML models such as support vector machine (SVM) and neural networks. SVM is an algorithm that efficiently handles non-linear classification by converting it to a plane of a higher dimension than the original dimension, and it can be performed by introducing a kernel trick.

The neural network model is composed of hidden layers and nodes, and the model was constructed using the “nnet” package in R.

Measure of performance

Since we have fitted several ML models, model comparison was an essential procedure. First, the AUC value (the area under the ROC curve) was calculated for several models. In addition, we calculated the binomial deviance (BD) as follows:

$$2 \sum_{i=1}^n [y_i \log(y_i/\hat{\mu}_i) + (1 - y_i) \log((1 - y_i)/(1 - \hat{\mu}_i))],$$

where y_i and $\hat{\mu}_i$ are, respectively, the i -th observed response and the probability of success estimated from the model.

This measure can be thought of as a generalization of RSS (residual sum of squares), and as the calculated probability value approaches the true value, BD tends to decrease. A cross-validated approach was used to evaluate these two parameters, AUC and BD.

RESULTS

Demography of COVID-19 patients in Korea

This retrospective study included 8070 SARS-CoV-2 confirmed patients of the 129,120 SARS-CoV-2 RNA tested patients in South Korea between January and July 2020, and whose data were available from the National Health Insurance Service. Their average age was 39.9 years (SD: 19.7 years), 3236 (40.1%) males and 4834 (59.9%) females. Of the 785 patients classified as severe, 374 were men and 411 were women ($p < 0.001$). The mean age of severely ill patients was 61.6 years (SD 16.0 years). There were a total of 248 patients who died. Among the patients who died, 136 were male and 112 were female ($p = 0.0008$). The average age of the patients who died was 72.1 years (10.2 years) (Figure 1B).

Regarding the underlying diseases in COVID-19 patients, 2259 patients had a history of pulmonary disease, 806 patients with a history of hypertension (HTN), 1371 patients with anosmia, 659 patients with cancer, 725 patients with diabetes mellitus (DM), 943 patients with chronic sinusitis, 931 patients with osteoporosis, 796 patients with rheumatic disease (RA), 193 patients with cardiovascular disease (CVD), 231 patients with ACE inhibitor or Angiotensin Receptor Blocker (ARB), 303 patients with peripheral vascular disease (PVD), 1389 patients with depression, 1471 patients with gastroesophageal reflux disease (GERD), 629 patients with angina, 556 patients with congestive heart failure (CHF), 77 patients with acute sinusitis, 33 patients with inflammatory bowel disease (IBD), 286 patients with acute myocardial infarction (MI), 674 patients with influenza, and 20 patients with non-tuberculosis mycobacterium (NTM) (Figure 1C).

Factors that can predict the severity of COVID-19

Model selection was made by comparing the AUC and BD values for each model. Among the various models, the model with the best prediction of severity was the neural network with an AUC value of 85.06%, followed by logistic regression EN (84.91%), logistic regression ridge (84.89%), logistic regression lasso (84.88%), random forest (83.57%), and the bagging model (82.61%) (Figure 2A).

The lowest value for BD was with the neural network model with a BD value of 25867.68, followed by logistic regression lasso (27821) and logistic regression EN (27863.47). The random forest based on the tree model had a BD value of 28130.22, followed by bagging with 28192.47 (Figure 2B).

In the neural network model, variable importance was in the order of influenza> NTM> IBD> ARB> anosmia (Table 1, Figure 2C).

In the logistic regression EN model, variable importance was in the order of age> DM> anosmia>male>HTN> CVD>ARB, and in lasso, it was in the order of age>DM>male>anosmia>HTN>ARB>CVD. In ridge, it was age>anosmia>DM>HTN>male>influenza>ARB (Table 1).

In the tree model, variable importance in bagging was age>HTN>DM>male>pulmonary disease>cancer history, and in RF, the order of importance was Age>HTN>DM>CVD>osteoporosis (Table 1).

Factors that can predict the mortality of COVID-19

Model selection was again made by comparing the AUC and BD values for each model. Among the various models, the model with the best prediction of death was the logistic regression EN

model with an AUC value of 93.86%, followed by the logistic regression lasso model (93.74%), the neural network model (93.73%), logistic regression ridge model (93.67%), random forest (91.98%), and finally the bagging model (91%) (Figure 3A).

The lowest BD value was with the neural network model with a BD value of 5932.504, followed by logistic regression EN (6978.869), lasso (6999.053), and ridge (7067.44). The random forest based on tree model had a BD value of 7179.297, followed by bagging with 7333.569 (Figure 3B).

In the logistic regression EN model, variable selection importance was in the order of age>anosmia>metropolitan> NTM>influenza>male> DM>HTN, and in lasso, it was in the order age> metropolitan> influenza>anosmia>male> NTM>DM>HTN. In ridge, the order was age>anosmia>NTM>metropolitan>influenza> HTN>male>DM.

In the neural network model, variable selection was important in the order of CVD Hx> age> male> RA> C sinusitis> influenza Hx> IBD (Table 1, Figure 3C).

In the tree model, variable importance in bagging was age> male> HTN> DM> CVD> lung disease> cancer history, and in RF, the order of importance was age> HTN> male> DM> CVD> PVD (Table 1).

DISCUSSION

The reproduction number (R_0) of COVID-19 is 2.56 (95%CI: 2.49–2.63), which is considerably higher than that of MERS ($R_0 < 1$).^{3,4} So the infectivity of COVID-19 is high, but mortality (3.4%) is still relatively low compared to other MERS (case fatality rate (CFR): 34.4%) and SARS-CoV (9.6%); however, the COVID-19 pandemic is still in progress, so we need to continue to monitor the trends.^{5,6} Nevertheless, since the death toll with COVID-19 is so high (about 1,400,000 worldwide as of November 26, 2020), it is of considerable significance to investigate the level of severity and mortality.

We have predicted the severity and mortality rate using the ML technique. Based on AUC value, the model that best predicted severity was the neural network followed by logistic regression EN, logistic regression ridge, logistic regression lasso, random forest and bagging model (Figure 2A). Based on BD, the model that best predicted severity was the neural network followed by logistic regression lasso, logistic regression EN, random forest then bagging.

So the neural network was the model with the best performance among the models for predicting COVID-19 severity. Influenza, NTM, IBD, ARB, and anosmia were the important variables selected in the neural network model.

It has been reported that the status of the immune system of patients who have recovered from COVID-19 is similar to that in individuals when they recover from influenza with activation of immune cells.⁷ Infection caused by the influenza virus can cause pulmonary fibrosis and emphysema in severe cases, and as a result, the cross-sectional area of the pulmonary alveoli is reduced.⁸ This aspect seems to be a major factor in increasing COVID-19 severity and can lead to death. These results are in line with the current policy recommending influenza virus vaccination, mainly considering the current COVID-19 epidemic and the prevalence of

influenza during the period from autumn to spring. Non-tuberculosis mycobacterium (NTM) can also cause a decline in forced exhalation volume in 1 second (FEV1) and forced vital capacity (FVC). It is a causative disease, and appears to contribute to increasing the severity of COVID-19 when patients are infected with this disease.⁹

In our study, influenza history was a very important variable in terms of COVID-19 severity (neural network 1st, ridge 6th) and mortality (EN 4th, lasso 2nd, ridge 5th). It has been reported that oseltamivir cannot prevent worsening of symptoms and disease in patients with COVID-19 as different molecular docking sites have been found in in vitro and retrospective studies in COVID-19.¹⁰ Among recent papers, it has been reported that influenza vaccination can alleviate the risk of death in a pandemic situation caused by COVID-19.¹¹ Since the symptoms of influenza and COVID-19 are similar, it can be confusing which disease is present, so vaccination can be important in preventing the twindemic of COVID-19 and influenza co-infection. In this paper, we studied the history of influenza and the severity of COVID-19. A history of influenza can sometimes cause pulmonary fibrosis, a common sequelae of virus-induced pneumonia, and this complication is estimated to cause increased severity and mortality of COVID-19 infection.

Among all variables, age was the top factor affecting disease severity and mortality in COVID-19 patients in Korea. The average age of patients with severe COVID-19 was 21.7 years higher than the average age of all COVID-19 confirmed cases, and the average age of those who died from COVID-19 was 10.5 years higher than the age of those with severe COVID-19.

Among the demographic factors, sex (male) was significant. Among patients with severe COVID-19, 631 were male (46.7%), and among patients who died from COVID-19, 151 were male (51.5%). As the severity and mortality increased, the proportion of males increased statistically significantly ($p < 0.05$, respectively). This trend supports previous studies in China,

Italy, and the United States which reported that men and old age are two important risk factors.¹²⁻¹⁴ In our paper, the proportion of underlying diseases is higher in men than in women, and the proportion of underlying diseases is particularly high in the elderly, which is the main reason for the severity and mortality of COVID-19. In our study, patients living in metropolitan areas did not have significantly increased severity of COVID-19 compared to other patients, and there were significantly lower mortality rates in metropolitan areas. In Korea, the mortality rate is thought to be significantly lower in large cities, thanks to the well developed healthcare system there.

In the models for predicting severity, diabetes was the second to fourth most important variable in the EN, ridge, lasso, and RF models. In the models for predicting mortality, it was the 8th most important variable in the EN model and 5th in the lasso model. When diabetes is poorly controlled, lymphocyte proliferation is inhibited,¹⁵ as are neutrophil and monocyte functions.¹⁶ Although the mechanism for the role of DM in COVID-19 has not yet been fully elucidated, its relevance is recognized epidemiologically, and this study supports that.¹⁷

Hypertension is an important factor in COVID-19. It was the 4th to 5th most important variable in the EN, ridge, and lasso models with regard to severity of COVID-19 and the 8th most important variable in the EN and lasso models with regard to death from COVID-19. Some studies have shown that over 60% of the elderly have a prevalence for hypertension.¹⁸ With logistic regression analysis of the variables age and hypertension in our data, statistically significant results were found with a very high prevalence of hypertension in old age ($p = 2e-16 < 0.001$).

Angiotensin converting enzyme II (ACE2) is a controversial factor. It is one of the enzymes in the renin-angiotensin system (RAS) and is a predominant cellular receptor of COVID-19 and SARS-CoV. ACE2 is present in the lungs, heart, kidneys, endothelium, liver, intestine, testis

and oral mucosa.¹⁹ In addition, ACE2 is known to be related to IL-6, which is closely related to cytokine storm syndrome, and is known to affect the severity of COVID-19.²⁰ In theory, it can be assumed that the ACE inhibitor blocks ACE receptors in cells of various organs in the human body through which COVID-19 can penetrate, and thus the severity of COVID-19 can be reduced. However, there is also a study reporting that ACE inhibitors have no effect on plasma ACE2 levels, so it is difficult to confidently predict the effect on COVID-19.²¹ In recent research results, there are reports that RAS blockers raise the risk of COVID 2 virus infection,²² and there are papers which report that RAS enzymes are not related to COVID-19 infection or severity.²³ In our study, the use of angiotensin receptor blocker (ARB) was the 4th to 6th most important variable for severity of COVID-19, and was considered to be a variable that had no effect on mortality.

Anosmia was also identified as an important variable in predicting the severity of COVID-19. The best predictive models for severity were neural networks followed by logistic regression EN, ridge, and lasso. In the neural network model, olfactory loss was the fifth most important variable, whereas in logistic regression EN and ridge, it was the second most important variable, and in lasso, the fourth most important variable to lower severity. The best predictive models for mortality were the EN and lasso models, and the second most important variable in both these models was olfactory loss. This means that the mortality rate was low in patients with olfactory loss after the COVID-19 diagnosis. There are papers which indicate that recent olfactory loss in mild to moderate COVID-19 patients is an important factor that differentiates COVID-19 from other infectious disease, and in most cases, the sense of smell recovers well.^{24,25} The novel finding in our study is that anosmia is an important indicator for COVID-19, and most patients with such anosmia will not be severely affected by the disease or die. It is assumed that the anosmia in COVID-19 is a result of transneuronal penetration.²⁴ As recent

studies have reported that loss of sense of smell is a major symptom which indicates that patients are infected with COVID-19, the presence of this indicator should also be carefully examined, as it is a major indicator of lower severity and mortality. Taken together with the results of this paper, olfactory loss will continue to be an indicator that should be carefully examined in COVID-19 infection.²⁶

CONCLUSION

Influenza was found to be a major adverse factor in COVID-19 in addition to the factors of old age and male sex, and which are already known to be related to disease severity and mortality. In addition, anosmia was found to be a major factor associated with lower severity and mortality rates. Therefore, in the current situation where there is no adequate COVID-19 treatment at present, examining the history of influenza and anosmia in addition to age and sex will be important indicators for predicting the severity and mortality of COVID-19 patients.

References

1. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
2. Goh GK, Dunker AK, Foster JA, Uversky VN. Rigidity of the outer shell predicted by a protein intrinsic disorder model sheds light on the COVID-19 (Wuhan-2019-nCoV) infectivity. *Biomolecules* 2020;10.
3. Zhao S, Musa SS, Lin Q, et al. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: A data-driven modelling analysis of the early outbreak. *J Clin Med* 9:2020.
4. Chen J. Pathogenicity and transmissibility of 2019-nCoV - A quick overview and comparison with other emerging viruses. *Microbes Infect* 22:69–71, 2020.
5. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14:523–34, 2016.
6. Ahmadzadeh J, Mobaraki K, Mousavi SJ, Aghazadeh-Attari J, Mirza-Aghazadeh-Attari M, Mohebbi I. The risk factors associated with MERS-CoV patient fatality: A global survey. *Diagn Microbiol Infect Dis* 96:114876, 2020.
7. Thevarajan I, Nguyen THO, Koutsakos M, et al. Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. *Nat Med* 26:453–5, 2020.
8. Keeler SP, Agapov EV, Hinojosa ME, Letvin AN, Wu K, Holtzman MJ. Influenza A virus infection causes chronic lung disease linked to sites of active viral RNA remnants. *J Immunol* 201:2354–68, 2018.
9. Park HY, Jeong BH, Chon HR, Jeon K, Daley CL, Koh WJ. Lung function decline according to clinical course in nontuberculous mycobacterial lung disease. *Chest* 150:1222–32, 2016.

10. Tan Q, Duan L, Ma Y, et al. Is oseltamivir suitable for fighting against COVID-19: In silico assessment, in vitro and retrospective study. *Bioorg Chem* 104:104257, 2020.
11. Grohskopf LA, Liburd LC, Redfield RR. Addressing influenza vaccination disparities during the COVID-19 pandemic. *JAMA* 324:1029–30, 2020.
12. Shen N, Zhu Y, Wang X, et al. Characteristics and diagnosis rate of 5630 subjects receiving SARS-CoV-2 nucleic acid tests from Wuhan, China. *JCI Insight* 5: 2020.
13. Grasselli G, Zangrillo A, Zanella A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA* 323:1574–81, 2020.
14. Miller J, Fadel RA, Tang A, et al. The impact of sociodemographic factors, comorbidities and physiologic response on 30-day mortality in COVID-19 patients in metropolitan Detroit. *Clin Infect Dis* 2020.
15. Geerlings SE, Hoepelman AI. Immune dysfunction in patients with diabetes mellitus (DM). *FEMS Immunol Med Microbiol* 26:259–65, 1999.
16. Knapp S. Diabetes and infection: is there a link? - A mini-review. *Gerontology* 59:99–104, 2013.
17. Hussain A, Bhowmik B, do Vale Moreira NC. COVID-19 and diabetes: Knowledge in progress. *Diabetes Res Clin Pract* 162:108142, 2020.
18. Zhang Y, Moran AE. Trends in the prevalence, awareness, treatment, and control of hypertension among young adults in the United States, 1999 to 2014. *Hypertension* 70:736–42, 2017.
19. Venkatakrishnan AJ, Puranik A, Anand A, et al. Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. *Elife* 9, 2020.

20. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395:497–506, 2020.
21. Sriram K, Insel PA. Risks of ACE inhibitor and ARB usage in COVID-19: Evaluating the evidence. *Clin Pharmacol Ther* 108:236–41, 2020.
22. Jarcho JA, Ingelfinger JR, Hamel MB, D'Agostino RB, Sr., Harrington DP. Inhibitors of the Renin-Angiotensin-Aldosterone System and Covid-19. *N Engl J Med* 382:2462–4, 2020.
23. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Cardiovascular disease, drug therapy, and mortality in Covid-19. *N Engl J Med* 382:e102, 2020.
24. Lee Y, Min P, Lee S, Kim SW. Prevalence and duration of acute loss of smell or taste in COVID-19 patients. *J Korean Med Sci* 35:e174, 2020.
25. Baron-Sanchez J, Santiago C, Goizueta-San Martin G, Arca R, Fernandez R. Smell and taste disorders in Spanish patients with mild COVID-19. *Neurologia* 2020.
26. Calica Utku A, Budak G, Karabay O, Guclu E, Okan HD, Vatan A. Main symptoms in patients presenting in the COVID-19 period. *Scott Med J* 2020, DOI: 10.1177/36933020949253.

Acknowledgements

This paper was supported by a fund of the Biomedical Research Institute at Jeonbuk National University Hospital.

Figure legends

Figure 1. A. Flowchart of the entire study design. B. Age distribution of 8070 COVID-19 confirmed patients in Korea in this study C. Comorbidity of 8070 COVID-19 confirmed patients in Korea in this study

Figure 2. A. ROC curves and AUC values in the prediction of severity of COVID-19. B. BD in the prediction of severity of COVID-19. C. Variable importance of the neural network model in the prediction of severity of COVID-19.

Figure 3. A. ROC curves and AUC values in the prediction of mortality of COVID-19. B. BD in the prediction of mortality of COVID-19. C. Coefficient heatmap of the three logistic model in the prediction of mortality of COVID-19.