Turkish Journal of Computer and Mathematics Education Vol.12 No.13 (2021), 2862-2868

Research Article

A Study on Self-Diagnosis Method to Prevent the Spread of COVID-19 Based on SVM

Young-Sang Kwak[1], Seo-Won Song[2], Seong-Hee Yeo[3], Min-Soo Kang[*4]

[1,2]Medical IT Marketing, Eulji Univ., SEONGNAM, Repbulic of Korea
[3]Biomedical Laboratory Science, Eulji Univ. SEONGNAM, Republic of Korea
[*4]Medical IT, Eulji Univ. SEONGNAM, Republic of Korea
ysk1188@naver.com[1], songsrr12@naver.com[2], ysher92@eulji.ac.kr[3], mskang@eulji.ac.kr[*4]
*Corresponding Author

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract: In this paper, a study was conducted to find a self-diagnosis method to prevent the spread of COVID-19 based on machine learning. COVID-19 is an infectious disease caused by a newly discovered coronavirus. According to WHO's situation report published on May 18th, 2020, COVID-19 has already affected 4,600,000 cases and 310,000 deaths globally and still increasing. The most severe problem of COVID-19 virus is that it spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, which occurs in everyday life. And also, at this time, there are no specific vaccines or treatments for COVID-19.Because of the secure diffusion method and the absence of a vaccine, it is essential to self-diagnose or do a self-diagnosis questionnaire whenever possible. But self-diagnosing has too many questions, and ambiguous standards also take time. Therefore, in this study, using SVM(Support Vector Machine), Decision Tree and correlation analysis found two vital factors to predict the infection of the COVID-19 virus with an accuracy of 80%. Applying the result proposed in this paper, people can self-diagnose quickly to prevent COVID-19 and further prevent the spread of COVID-19.

Keywords: COVID-19,SVM(SupportVectorMachine),DecisionTree,CorrelationAnalysis

1. Introduction

Coronavirus is one of the major pathogens that primarily targets the human respiratory system. Previous outbreaks of coronaviruses (CoVs) include the severe acute respiratory syndrome (SARS)-CoV and the Middle East respiratory syndrome (MERS)-CoV, which have been previously characterized as agents that are a tremendous public health threat. In late December 2019, a cluster of patients was admitted to hospitals with an initial diagnosis of pneumonia of an unknown etiology. These patients were epidemiologically linked to the seafood and wet animal wholesale market in Wuhan, Hubei Province, China [1,2]. Early reports predicted the onset of a potential Coronavirus outbreak given the estimate of a reproduction number for the 2019 Novel (New) Coronavirus (COVID-19, named by WHO on February 11, 2020) which was deemed to be significantly larger than 1 (ranges from 2.24 to 3.58) [3]. The chronology of COVID-19 infections is as follows. The first cases were reported in December 2019 [4]. From December 18, 2019, through December 29, 2019, five patients were hospitalized with acute respiratory distress syndrome, and one of these patients died [5]. By January 2, 2020, 41 admitted hospital patients had been identified as having laboratory-confirmed COVID-19 infection, less than half of these patients had underlying diseases, including diabetes, hypertension, and cardiovascular disease [6]. These patients were presumed to be infected in that hospital, likely due to nosocomial infection. It was concluded that the COVID-19 is not a super-hot spreading virus (spread by one patient to many others), but rather likely spread due to many patients getting in fected at various locations throughout the hospital through unknown mechanisms. In addition, only patients that got clinically sick were tested; thus, there were likely many more patients that were presumably infected. As of January 22, 2020, a total of 571 cases of the 2019-new coronavirus (COVID-19) were reported in 25 provinces (districts and cities) in China [7]. The China National Health Commission reported the details of the first 17 deaths up to January 22, 2020. On January 25, 2020, a total of 1975 cases were confirmed to be infected with the COVID-19 in mainland China with a total of 56 deaths [8]. Another report on January 24, 2020, estimated the cumulative incidence in China to be 5502 cases [9]. As of January 30, 2020, 7734 cases have been confirmed in China, and 90 other cases have also been reported from a number of countries that include Taiwan, Thailand, Vietnam, Malaysia, Nepal, Sri Lanka, Cambodia, Japan, Singapore, Republic of Korea, United Arab Emirates, United States, The Philippines, India, Australia, Canada, Finland, France, and Germany. The case fatality rate was calculated to be 2.2% [10]. The first case of COVID-19 infection confirmed in the United States led to the description, identification, diagnosis, clinical course, and management of this case. This includes the patient's initial mild symptoms at presentation and progression to pneumonia on day 9 of illness [11]. Further, the first case of human-to-human transmission of COVID-19 was reported in the US on January 30, 2020. The CDC has so far screened 30,000 passengers arriving at US airports for the novel coronavirus. Following such initial screening, 443 individuals have been tested for coronavirus infection in 41 states in the USA. Only 15 (3.1%) were tested positive, 347 were negative, and results on the remaining 81 are pending. A report published in Nature revealed that Chinese health authorities concluded that as of February 7, 2019,

there had been 31,161 people who have contracted the infection in China, and more than 630 people have died of disease. At the time of preparing this manuscript, the World Health Organisation (WHO) reported 51,174 confirmed cases, including 15, 384 severe cases, and 1666 death cases in China. Globally, the number of confirmed cases as of this writing (February 16, 2020) has reached 51,857 in 25 countries.

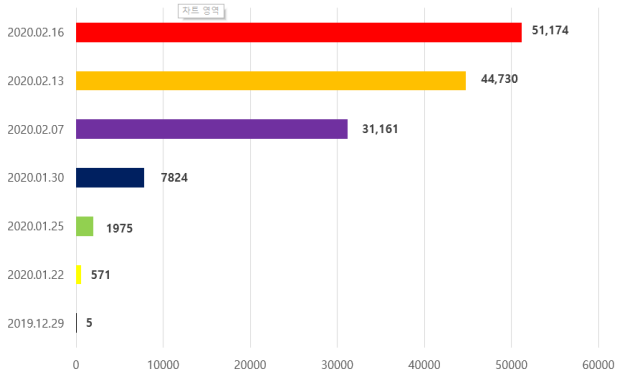

Figure1. Cases with COVID-19 Infection (2020)

Source: https://www.sciencedirect.com/science/ar
https://www.sciencedirect.com/science/article/pii/S0896841120300469)

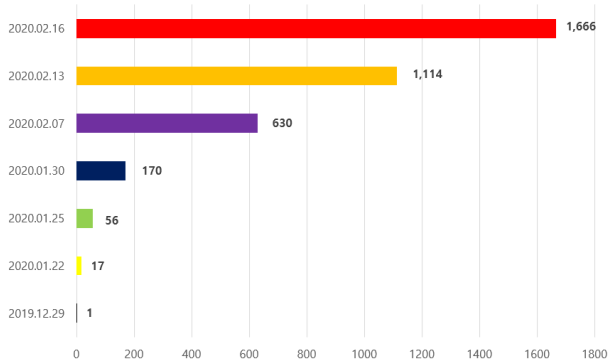

Figure*2*. Death Cases of COVID-19 Infection (2020)
Source:https://www.sciencedirect.com/science/ar
https://www.sciencedirect.com/science/article/pii/S0896841120300469)

The symptoms of COVID-19 infection appear after an incubation period of approximately 5.2 days [12]. The period from the onset of COVID-19 symptoms to death ranged from 6 to 41 days, with a median of 14 days. This period is dependent on the age of the patient and the status of the patient's immune system. It was shorter among patients >70-years old compared with those under the age of 70. The most common symptoms at the onset of COVID-19 illness are fever, cough, and fatigue, while other symptoms include sputum production, headache, hemoptysis, diarrhea, dyspnoea, and lymphopenia [13]. Clinical features revealed by a chest CT scan presented as pneumonia. However, there were abnormal features such as RNAaemia, acute respiratory distress syndrome, severe cardiac injury, and incidence of grand-glass opacities that led to death. In some cases, the multiple peripheral ground-glass opacities were observed in subpleural regions of both lungs [14] that likely induced both systemic and localized immune response that led to increased inflammation. Regrettably, treatment of some cases with interferon inhalation showed no clinical effect and instead appeared to worsen the condition by progressing pulmonary opacities

For these reasons, it is necessary to study for finding self-diagnosis methods to prevent the spread of COVID-19. There are 21 attributes that contain a demographic feature, other diseases such as diabetes, heart disease, kidney disease, and corona result. First, correlation analysis was conducted to determine which attributes profoundly affect corona results, and then, corona results were predicted using SVM(Support Vector Machine) through variables that have a potent effect.

2. Supervised Learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value(also called the supervisory signal). A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances.

2.1 SVM(Support Vector Machine)

In machine learning, Support-Vector Machine is supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, and an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the separate categories that are divided by a clear fap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall [18]. The reason for using a two-class support vector machine is that technically it can be used in both classification and forecasting problems, and the second is less likely to be overfitted than neural network techniques, and thirdly, it is more accurate to predict, and lastly for its simplicity. For its experimental usage, the performance of two-class logistic regression and two-class neural networks were lower than the two-class support vector machine.

3. Related Research

In order to achieve sustainable development with its holistic concept and approach, there must be political will of states and a willingness of societies and individuals to achieve them.

A study of the epidemiological characteristics of an Outbreak of 2019 Novel Coronavirus Diseases in China was conducted to find how and why the COVID-19 spread so quickly. The objective was about an outbreak of 2019 novel COVID-19 in Wuhan, China has spread quickly nationwide. All COVID-19 cases reported through February 11, 2020, were extracted from China's Infectious Disease Information System. The analysis included a summary of patient characteristics, examination of age distributions and sex ratio, calculation of case fatality and mortality rates, geo-temporal analysis of viral spread, epidemiological curve construction, and subgroup analysis. The results were that the COVID-19 spread outward from Hubei sometime after December 2019, and by February 11, 2020, 1 386 counties across all 31 provinces were affected. The epidemic curve of onset of symptoms peaked on January 23-26, then began to decline leading up to February 11. A total of 1 716 health workers have become infected, and five have died [19]. And Pathological findings of COVID-19 associated with acute respiratory distress syndrome was conducted in China. The pathological features of COVID-19 greatly resemble those seen in SARS and Middle Eastern respiratory syndrome (MERS) coronavirus infection. In addition, the liver biopsy specimens of the patient with COVID-19 showed moderate microvascular steatosis and mild lobular and portal activity, indicating the injury could have been caused by either SARS-CoV-2 infection or drug-induced liver injury. There were a few interstitial mononuclear inflammatory infiltrates, but no other substantial damage in the heart tissue. The clinical and pathological findings in this severe case of COVID-19 can not only help to identify a cause of death, but also provide new insights into the pathogenesis of SARS-CoV-2-related pneumonia, which might help physicians to formulate a timely therapeutic strategy for similar severe patients and reduce mortality [20]. Next, a study was conducted about COVID-19 and the cardiovascular system also in China. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infects host cells through ACE2 receptors, leading to coronavirus disease (COVID-19)-related pneumonia, while also causing acute myocardial injury and chronic damage to the cardiovascular system. Therefore, attention should be given to cardiovascular protection during treatment for COVID-19. Conclusively, SARS-CoV-2 is thought to infect host cells through ACE2 to cause COVID-19, while also causing damage to the myocardium, although the specific mechanisms are uncertain. Patients with underlying CVD and SARS-CoV-2 infection have an adverse prognosis. Therefore, attention should be given to cardiovascular protection during treatment for COVID-19 [21]. And a paper called "COVID-19 and Italy: what next?" was conducted to figure out the status after the COVID-19 break. The spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has already taken on pandemic proportions, affecting over 100 countries in a matter of weeks. A global response to prepare health systems worldwide is imperative. Although containment measures in China have reduced new cases by more than 90%, this reduction is not the case elsewhere, and Italy has been particularly affected. There is now grave concern regarding the Italian national health system's capacity to respond to the needs of patients who are infected and require intensive care for SARS-CoV-2 pneumonia effectively. The percentage of patients in intensive care reported daily in Italy between March 1 and March 11, 2020, has consistently been between 9% and 11% of patients who are actively infected. The number of

patients infected since February 21 in Italy closely follows an exponential trend. If this trend continues for one more week, there will be 30 000 infected patients. Intensive care units will then be at maximum capacity; up to 4000 hospital beds will be needed by mid-April 2020. Our analysis might help political leaders and health authorities to allocate enough resources, including personnel, beds, and intensive care facilities, to manage the situation in the next few days and weeks. If the Italian outbreak follows a similar trend as in Hubei province, China, the number of newly infected patients could start to decrease within 3–4 days, departing from the exponential trend. However, this cannot currently be predicted because of differences between social distancing measures and the capacity to quickly build dedicated facilities in China [22]. Lastly, a study called "Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?" was conducted. This paper suggests that patients with cardiac diseases, hypertension, or diabetes, who are treated with ACE2-increasing drugs, are at higher risk for severe COVID-19 infection and, therefore, should be monitored for ACE2-modulating medications, such as ACE inhibitors or ARBs. Based on a PubMed search on February 28, 2020, we did not find any evidence to suggest that antihypertensive calcium channel blockers increased ACE2 expression or activity. Therefore, these could be a suitable alternative treatment in these patients [23].

4. Experiment

4.1 Experimental Environment

In this paper, Microsoft Azure Machine Learning Studio was used to assess the COVID-19 disease feature and to find the most influential feature when diagnosing COVID-19 disease patients. Azure is a cloud computing platform of Microsoft that began servicing since 2010, along with the commencement of PaaS service in 2011, followed by IaaS service in 2013. Azure platform provides more than 600 services, and we are using Azure Machine Learning Studio, which is one of such services [24,28]. Azure Machine Learning Studio is used because of the provision of modules in various formats that assist with the development of and distribution of one's machine learning model in the actual environment. Moreover, it is void of the lack of flexibility in the computing resources necessary and inconvenience of the setting works for GPU-based learning such as Tensor Flow library, which are the chronic problem of the existing Machine Learning library and tool, thereby making it very easy to install and set tools and environment necessary for learning [25,27]. In addition, it solves the difficulties of the recording of experimental processes and version management and enables easy direct installation and setting of a solution system for collaboration with Jupyter Notebook, etc. Utilization of Azure Machine Learning Studio enables quick execution of repetitive works by producing and testing several models within several minutes. Generally, designing of the prediction model, and editing and experimenting of parameter or model are executed repetitively when producing experiment. Series of such works can be handled easily through the utilization of Azure Machine Learning Studio.

4.2 Data Processing

COVID-19 data utilized in this study were collected from the open source site, Kaggle specifically 128 rows and 21 columns, including demographic data, various symptoms, and diseases. Table 1 illustrates the data collected.

Table 2. COVID – 19 Data

Data Group	Example
Demographic data	Age, Gender, Body Temperature
Various symptoms	Dry Cough, Sour Throat, Weakness, Breathing Problem, Drowsiness, Pain in chest, Travel History to infected countries, Change in Appetite, Loss of sense of smell
Various Disease	Diabetes, Heart Disease, Lung Disease, Stroke or Reduced Immunity, Symptoms progressed, High Blood Pressure, Kidney Disease,
Result	Corona Result

Pre-processing of data collected was executed using correlation analysis. Figure 2 illustrates the pre-processing results of the data.

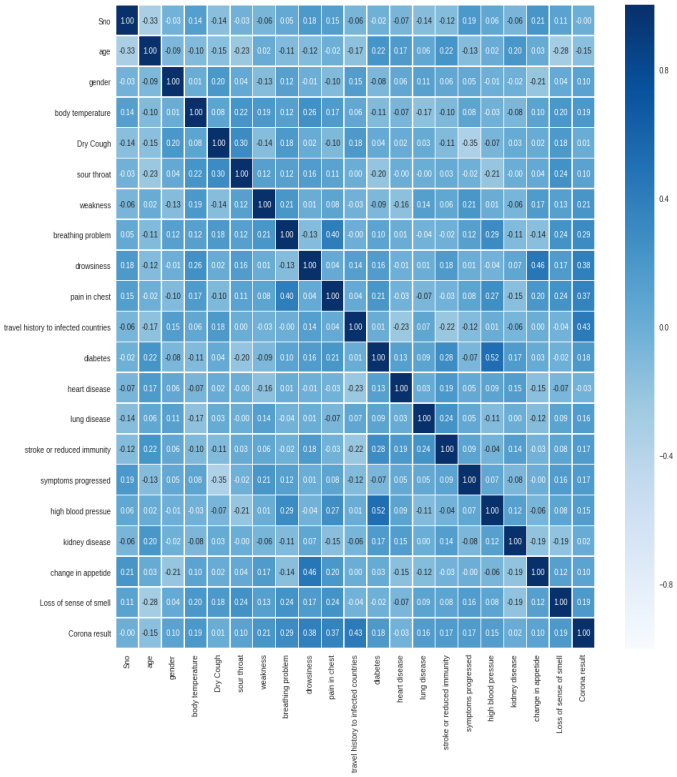

Figure**3**. Correlation Analysis Result using COVID – 19Data

Through the correlation result analysis result Drowsiness, Pain in chest, Travel history to infected countries attributes have 0.38, 0.37, 0.43 relation to the COVID-19 results. So, 1 experiment was executed using these three attributes applying SVM and Decision tree to predict COVID-19 results.

5. Results

The experiment with Drowsiness, Pain in Chest, Travel history to infected countries feature using SVM resulting in 0.806 accuracy, 0.737 precision, 0.933 recall, and finally 0.824 F1 score. Figure 3 visualizes how evaluate model came out. And with Decision Tree only three steps took to diagnose COVID-19 patients whether patients get COVID-19 or not.

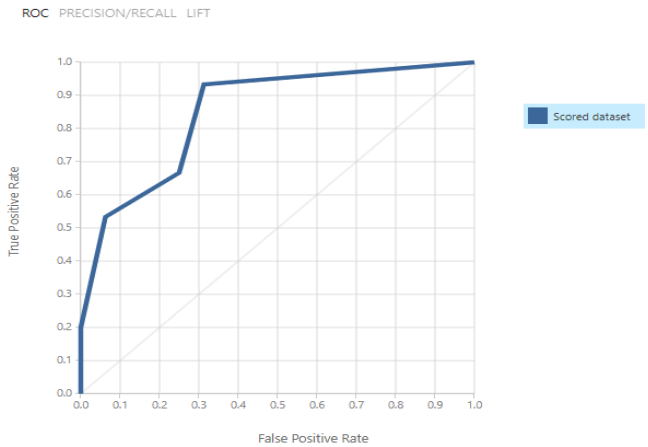

Figure**4**. Evaluate Model (ROCGraph)

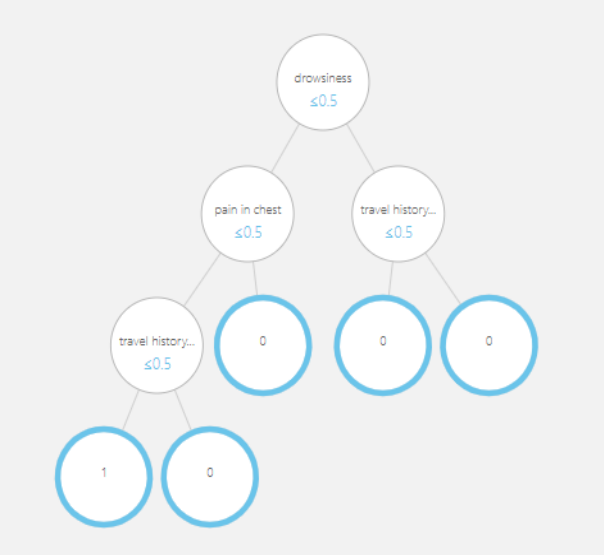

Figure**5**. Evaluate Model (Decision Tree)

6. Conclusion

In this paper, 1 experiment was made to predict COVID-19 patients using two – class support vector machine and two-class boosted decision tree. The result showed means that when classifying COVID-19 patients, only checking the potential patient's drowsiness, pain in chest, and their travel history to infected countries is more efficient than doing self-diagnosis questionnaire, which takes time and has ambiguous standards. So, applying the result proposed in this paper, doctors could be able to reduce time when diagnosing COVID-19 patients and diagnose other potential COVID-19 patients more. Furthermore, doing this process could prevent the spread of COVID-19.

Acknowledgment

'This research was supported by 2020 eulji university University Innovation Support Project grant funded'.

References

1. Bogoch, A.; Watts, A.; Thomas-Bachli; C. Huber; M.U.G. Kraemer; K. Khan.Pneumonia of unknown etiology in wuhan, China: potential for international spread via commercial air travel
2. H. Lu; C.W. Stratton; Y.W. TangOutbreak of pneumonia of unknown etiology in wuhan China: the mystery and the miracle
3. S. Zhao; Q. Lin; J. Ran; S.S. Musa; G. Yang; W. Wang; *et al.*Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreakInt. J. Infect.
4. A. Du ToitOutbreak of a novel coronavirus
5. Nat. Rev. Microbiol., 18 (123) (2020), 10.1038/s41579-020-0332-0
6. L.L. Ren; Y.M. Wang; Z.Q. Wu; Z.C. Xiang;L. Guo; T. Xu; *et al.*Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study
7. Chinese
8. C. Huang; Y. Wang; X. Li; L. Ren; J. Zhao; Y. Hu. *et al.*Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China
9. H. LuDrug treatment options for the 2019-new coronavirus (2019-nCoV)
10. Biosci. Trends (2020), 10.5582/bst.2020.01020
11. W. Wang; J. Tang; F. WeiUpdated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China
12. H. Nishiura; S.M. Jung; N.M. Linton; R. Kinoshita; Y. Yang; K. Hayashi; et al.The extent of transmission of novel coronavirus in wuhan, China, 2020
13. M. Bassetti; A. Vena; D. Roberto.GiacobbeThe Novel Chinese Coronavirus (2019-nCoV) Infections: challenges for fighting the storm

14. M.L. Holshue; C. DeBolt; S. Lindquist; K.H. Lofy; J. Wiesman; H. Bruce. *et al.*First case of 2019 novel coronavirus in the United States
15. Q. Li; X. Guan; P. Wu; X. Wang; L. Zhou; Y. Tong. *et al.*Early transmission dynamics in wuhan, China, of novel coronavirus-infected pneumonia
16. W.G. Carlos; C.S. DelaCruz; B. Cao; S. Pasnick; S. JamilNovelwuhan (2019-nCoV) coronavirus
17. J. Lei; J. Li; X. Li; X. QiCT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia
18. Radiology (2020), p. 200236, 10.1148/radiol.2020200236
19. Samuel & Arthur L (1959). Some studies in machine learning using the game of checkers, IBM Journal of research and development 3.3, pp 210-229.
20. EuiJoong Kim. (2016). Artificial Intelligence, Machine Learning, Deep Learning introduction, wikibooks,
21. James, Gareth & et al. (2013). An introduction to statistical learning. Vol. 112. New York: springer, 2013.
22. Cortes, Corinna, *Vapnik& Vladimir N. (1995).* "Support-vector networks" (PDF). Machine Learning. *20 (3): 273–297.* CiteSeerX 10.1.1.15.9362. doi*:*10.1007/BF00994018.
23. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. Zhonghua Liu Xing Bing Xue Za Zhi. 2020;41(2):145-151. doi:10.3760/cma.j.issn.0254-6450.2020.02.003
24. Pathological findings of COVID-19 associated with acute respiratory distress syndrome, The Lancet Respiratory Medicine.
25. Zheng, Y.; Ma, Y.; Zhang, J. et al. COVID-19 and the cardiovascular system. Nat Rev Cardiol 17, 259–260 (2020). https://doi.org/10.1038/s41569-020-0360-5
26. Andrea Remuzzi, Giuseppe Remuzzi, COVID-19 and Italy: what next?, The Lancet, Volume 395, Issue 10231, 2020, Pages 1225-1228, ISSN 0140-6736
27. Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?. The Lancet. Respiratory medicine, 8(4), e21. https://doi.org/10.1016/S2213-2600(20)30116-8
28. Microsoft Azure Machine Learning, Website :https://docs.microsoft.com/ko-kr/azure/machine-learning/studio/what-is-ml-studio
29. Min Soo Kang, Eun Soo Choi, Getting started Machine Learning with MicroSoft AZURE ML, Hanti Media, 2018
30. Kaggle. COVID-19 Dataset From https://www.kaggle.com/bitsofishan/covid19-patient-symptoms
31. Aydın, İ. S. (2019). Improvement of preservice Turkish teachers' perceived writing self-efficacy beliefs. Educational Sciences: Theory & Practice, 19(1).
32. Aydin, S., Öztürk, A., Büyükköse, G. T., Er, F., & Sönmez, H. (2019). An Investigation of Drop-Out in Open and Distance Education. Educational Sciences: Theory and Practice, 19(2), 40-57.