A Real Data-Driven Analytical Model for Testing for the Novel Coronavirus Disease, COVID-19

Chris P. Tsokos and Lohuwa Mamudu*

University of South Florida, USA.

*Correspondence:

Lohuwa Mamudu, University of South Florida, USA.

ABSTRACT

To address the testing of the horrific pandemic disease that has terrified our global society, COVID-19, we have developed an analytical model that an individual can easily apply to determine if he or she tested positive or negative with a very high degree of accuracy. Our analytical model is real data-driven utilizing data obtained from the World Health Organization, WHO, and the United States Center for Disease Control and Prevention, CDC guidelines. Both WHO and CDC have identified several symptoms or risk factors from individuals diagnosed with the disease, COVID-19. They have identified and published nine symptoms that are associated with the disease, COVID-19. However, our structured analytical model identified only seven of the nine symptoms to statistically significantly contribute to the subject disease. They are fever, tiredness, dry cough, difficulty in breathing, sore throat, pain, and nasal congestion. Each of the symptoms shows highly likelihood of having COVID-19. Our analytical model was carefully developed, very well-validated, and statistically tested to achieve a 93% accuracy in the testing result. If a person is tested positive, we recommend that he/she seek medical evaluation and treatment. That is, once we receive the categorical data from a given individual, and we input into the proposed model, the output result will be the individual is tested positive or negative for COVID-19. The developed model identifies (estimated) the different weights of each of the seven symptoms or risk factors that play a major role in the decision process of the testing results. Our findings seek to enhance testing efficiency, treatment, control, and prevention strategy for the COVID-19 disease.

Important: The subject study is under the protection right by the University of South Florida (USF), with protection number 20B126/USF Technology Protection.

Introduction

What is COVID-19? In December 2019, the world received dreadful news about a newly identified type of infectious coronavirus disease in Wuhan, a province in the Republic of China. Since then, the disease has spread throughout the globe becoming a worldwide pandemic of respiratory illness and causing 421,856 deaths as of June 12, 2020. As of June 13, 2020, the global confirmed cases were about 7.7 million, with the United States confirming about 2 million of the total number of confirmed cases and 114,875 deaths, respectively [1]. The reported number of recovery cases as of June 13, 2020, was about 4 million people. The World Health Organization (WHO) on February 11, 2020, officially named the novel coronavirus infectious disease as coronavirus disease 2019 (COVID-19), also known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The exact cause or origin of the COVID-19 virus has not yet been discovered. However, the initial hypotheses were linked to seafood with the investigation about the origin and spread of the virus still ongoing. The COVID-19 virus spreads mainly through contact with infected persons or surfaces. A report from the U.S. Center for Disease Control and Prevention (CDC) outlined that the virus can be spread through respiratory droplets produced when an infected person coughs, sneezes, or talks. The droplets may land in the mouths or noses of nearby people or possibly be inhaled into the lungs. In some situations, the virus can be spread from animals to humans or from humans

to animals, but it is considered to be at a low rate. The disease can also be spread by an asymptomatic infected person [2]. Generally, the virus enters the body through the mouth, nose, or the eye and proceeds into the air sacs inside your lungs, known as alveoli. It causes inflammation in the lungs, which triggers an immune response after hijacking the cells to reproduce copies of the virus [3]. The COVID-19 virus is noticeable for attacking the lung. However, some reports show that it may attack the heart, causing a disruption to its critical rhythm and weakening its muscles. It may also damage the kidney very badly, crawls along the human nervous system, suppressing the ability to taste and smell, and in some cases reaching the brain. Blood clots may also be created that can kill with sudden efficiency and inflames blood vessels throughout the body. The elderly are said to be more susceptible to severe cases, with more harmful cases associated with men than women, and some cases may complicate pregnancies [4]. Figure 1 below shows the COVID-19 virus and its development in the human system [2,5].

Several symptoms have been associated with the COVID-19 disease with WHO, CDC, among others citing fever, dry cough, difficulty breathing, fatigue, sore throat, muscle pain, among others as the early signs of the disease [2]. The CDC reported that the symptoms of people after being exposed to the COVID-19 virus may appear 2-14 days. The signs and symptoms of the COVID-19 virus may vary from one infected person to another. However, the CDC reported that people with trouble breathing, persistent pain or pressure in the chest, new confusion, inability to wake or stay awake, and bluish lips or face should seek emergency medical attention [2]. WHO also pointed out that most people who contract COVID-19 may have mild flu-like symptoms, but occasionally the infection may result in a severe case of pneumonia that can be lethal, especially for older people and those with underlying medical conditions. The rate of spread and fatality of COVID-19 has prompted scientists and health researchers to draw comparisons with that of Severe Acute Respiratory Syndromes (SARS) in 2003 and the Middle East respiratory syndrome (MERS-CoV) in 2012. Although much is not known, the spread of the COVID-19 seems to be faster than SARS or MERS-CoV but may cause less severe illness and deaths [1]. How then do people get tested for COVID-19? WHO published several testing protocols with the standard testing method being the real-time reverse transcription-polymerase chain reaction (rRT-PCR) [6]. Typically, the testing is done on respiratory samples obtained by a nasopharyngeal swab, and in some instances, a nasal swab or sputum sample may also be used [7]. Most often the results after testing are available within a few hours to two days [8]. On March 21, 2020, the US Food and Drug Administration (FDA) approved the use of the first point-of-care test at the end of that month [9].

The 2019 novel coronavirus disease was inititially incurable due to lack of vaccine. In February 2020, WHO reported that it did not expect a vaccine against COVID-19, the causative virus, to become available in less than 18 months [10]. The initial vaccines which were developed failed in early-stage clinical trials, with none progressing for licensing [11]. However, there was a

tremendous global growing effort for the COVID-19 vaccine, with the Coalition for Epidemic Preparedness Innovations (CEPI) providing a rapid investment worldwide fund of US$2 billion for the development of vaccine candidates [12]. WHO, on May 2020 received US$8.1 billion in pledges from forty countries in an organized telethon in support for rapid development of vaccines to prevent the continuous spread and fatality from COVID-19 infections [13], as well simultaneous evaluation of several vaccine candidates reaching Phase II-III clinical trials in the deployment of international solidarity trials [14]. In early December 2020, the U.S. FDA approved a vaccine developed by Pfizer Inc and BioNTech SE, which is currently been administered and said to be 95% effective at preventing the COVID-19 disease. Similarly, other candidate vaccines are closed to receiving approval in the U.S. and other countries.

Given that COVID-19 vaccine is now available, it will take sometime for every individual in the U.S. and the world to get vaccinated. Therefore whiles most people wait to get the vaccine, what are the undertaking preventive measures? WHO and CDC have proposed several preventive measures to minimize the spread and infection of the disease. They include staying home, avoiding crowded places, social distancing, washing hands with soap and water often and for at least 20 seconds, practicing good respiratory hygiene, avoiding touching the eyes, nose, or mouth with unwashed hands, and quarantine of infected individuals [15]. Other recommendations given by the CDC also include covering the mouth and nose with a tissue when coughing or sneezing and using the inside of the elbow if no tissue is available. The CDC further recommended cloth face coverings in public settings where it is difficult to maintain social distancing measures, which can reduce transmission by asymptomatic individuals [16].

Several steps are been taking by countries and health organizations to limit the spread and infection of the COVID-19 disease, often referred to as flattening the curve. This is often achieved through the practice of the recommended preventive measures, contact-tracing, and testing. Continuously testing people enables the discovery of the infected population, who can then be isolated from the susceptible or uninfected population, aiding the flattening of the curve of COVID-19 infected cases. Now, given the scarcity or limited supply of testing equipment worldwide, not everyone can be tested or must be tested. Generally, more testings are done by advance or economically robust countries than third-world or developing countries.

Tim Mackey et al. (2020) assessed the self-reporting symptoms, testing access, and COVID-19 recovery on Twitter using machine learning. They found that many Twitter users reported symptoms they thought were related to COVID-19, but they were not able to get tested to confirm their concerns. How can people self-screen for COVID-19 in the absence or inadequate testing equipment before seeking further medical assessment and treatment? What can be done to reduce the fast-spreading of COVID-19 to uninfected populations? What then can be done to prevent, or minimize the waste of the limited supply of the test tubes to ensure equitable
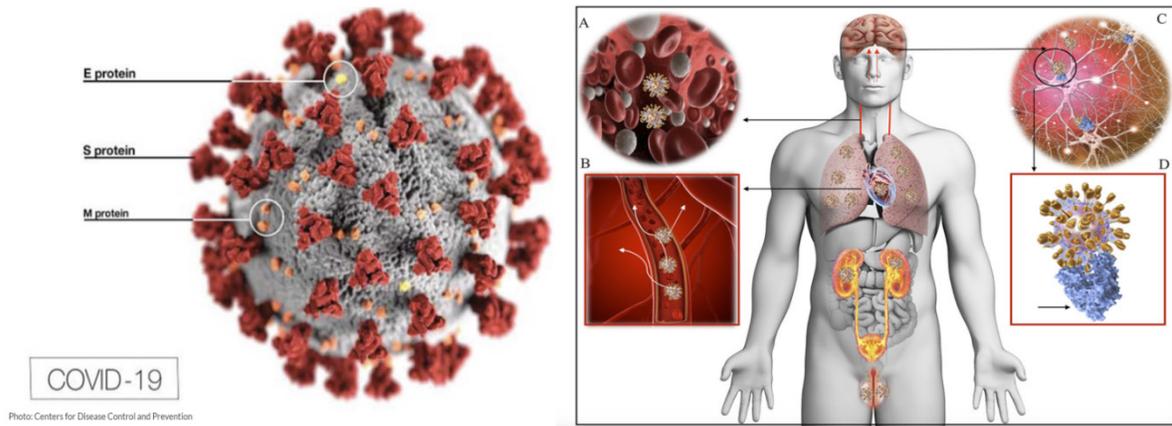
Figure 1: Development of the Novel Coronavirus Disease 2019 (COVID-19).

distribution or supply of the test tube worldwide? How can we ensure there is an effective and judicious use of scarce test kits to enhance efficiency? It is economically irrational and socially immoral to spend time and resources testing individuals with a greater likelihood of not tested positive.

Therefore, to answer the question of how efficient the testing could be, we use a machine learning technique to a developed an analytical logistic regression model that classifies individuals according to whether they are COVID-19 positive or negative. The objective of this research study is to assess or investigate the significant symptoms or risk factors of being infected with COVID-19 and then predict the probability of being tested positive or negative as a function of the symptoms the individual is experiencing.

Data and Methodology
The data used in this study consist of 316,800 observations or individuals from 10 countries combined, each consisting of 31,680 observations. The data is a public data obtained from Kaggle and available upon request. The objective of the data is to identify whether or not the individual has COVID-19 based on some pre-defined standard symptoms given under the guidelines of WHO. According to WHO, the following are the five major symptoms or signs of COVID-19 reported in the data. That is Fever, Tiredness, Difficulty in breathing, Dry cough, and sore throat. Other symptoms experienced by some individuals reported in the data include Pains, Nasal Congestion, Runny Nose, and Diarrhea. The data did not include whether or not the individual has COVID-19.

The data was collected at the early stage of the novel coronavirus outbreak, during which testing kits were not available. However, we created the column or response of whether the individual has contracted the disease or not, based on the major reported symptoms from recommendations and guidelines of WHO and the U.S. CDC. As outlined by John Hopkins Medicine, they stated that some people experience four early or first symptoms of COVID-19, including fatigue, headache, sore throat, or fever. The symptoms can be mild at first and may become intense over

5-7 days with other symptoms like cough and shortness of breath developing. Lisa Maragakis, (2020) Reviewed. Coronavirus Symptoms: Frequently Ask Questions. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/coronavirus-symptoms-frequently-asked-questions.

Therefore, individuals with four or more symptom were considered to have the coronavirus disease in our analysis. We classified individuals with at least four of the stated symptoms to be COVID-19 positive, otherwise, they are classified COVID-19 negative. We then obtained a binary response variable coded 0 and 1, representing COVID-19 negative and positive, respectively.

All nine symptoms or risk factors included in our analysis are categorical variables with each having a binary outcome of 0 and 1, not having the symptoms, and having the symptoms, respectively. In this study, we employed the classification method of binary logistic regression analytical approach for classifying the COVID-19 status of individuals (i.e. either positive or negative).

Statistical Analysis
The analytical model was developed using 70% trained data and the remaining 30% was used to test the model for prediction accuracy and precision. The stepwise backward elimination method of model selection was utilized to obtain the best model with the least Akaike information criterion (AIC) and Bayesian information criterion (BIC). The backward elimination model selection method initially places all the risk factors into the model and gradually eliminates those risk factors that are statistically insignificant at a p-value less than 5% alpha level of significance (α). It provides less biased mean square error and prevents overfitting of the model. The AIC and BIC introduce a penalty term for the number of parameters in the model to reduce overfitting. They both measure the amount of information lost in the model, hence the lowest AIC and BIC are preferred. We further applied the L1 regularization technique to assess overfitting in the model.

Table 1: Descriptive Statistics of Symptoms for Sex and Age

	Fever	Tiredness	Dry Cough	Difficulty Breathing	Sore Throat	Pains	Nasal Congestion	Runny Nose	Diarrhea
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Overall	99000 (31.25)	158400 (50)	178200 (56.25)	158400 (50)	99000 (31.25)	115200 (36.36)	19800 (6.25)	172800 (54.54)	172800 (54.54)
Sex									
Male	33000 (33.33)	52800 (33.33)	59400 (33.33)	52800 (33.33)	33000 (33.33)	38400 (33.33)	6600 (33.33)	57600 (33.33)	57600 (33.33)
Female	33000 (33.33)	52800 (33.33)	59400 (33.33)	52800 (33.33)	33000 (33.33)	3840 0(33.33)	6600 (33.33)	57600 (33.33)	57600 (33.33)
Transgender	33000 (33.33)	52800 (33.33)	59400 (33.33)	52800 (33.33)	33000 (33.33)	38400 (33.33)	6600 (33.33)	57600 (33.33)	57600 (33.33)
Age									
0-9	19800 (20)	31680 (20)	35640 (20)	31680 (20)	19800 (20)	23040 (20)	3960 (20)	34560 (20)	34560 (20)
10-19	19800 (20)	31680 (20)	35640 (20)	31680 (20)	19800 (20)	23040 (20)	3960 (20)	34560 (20)	34560 (20)
20-24	19800 (20)	31680 (20)	35640 (20)	31680 (20)	19800 (20)	23040 (20)	3960 (20)	34560 (20)	34560 (20)
25-59	19800 (20)	31680 (20)	35640 (20)	31680 (20)	19800 (20)	23040 (20)	3960 (20)	34560 (20)	34560 (20)
60+	19800 (20)	31680 (20)	35640 (20)	31680 (20)	19800 (20)	23040 (20)	3960 (20)	34560 (20)	34560 (20)
COVID-19	**n (%)**								
Positive	185400 (58.5)								
Negative	131400 (41.5)								

Table 2: Statistical Significance of COVID-19 Symptoms or Risk Factors

Variable	Description	Coef (β)	Std Err	p-value	OR (95% CI)
Constant	Intercept	-8.7479	0.0387	<0.0001	0.51
X_1	Fever	2.4325	0.0205	<0.0001	11.42 (11.04, 11.81)
X_2	Tiredness	2.4100	0.0195	<0.0001	11.14 (10.69, 11.45)
X_3	Dry Cough	2.1879	0.0193	<0.0001	8.94 (8.66, 9.23)
X_4	Difficulty Breathing	2.3893	0.0195	<0.0001	11.04 (10.69, 11.40)
X_5	Sore Throat	2.4654	0.0206	<0.0001	11.45 (11.24, 11.81)
X_6	Pains	1.7129	0.0166	<0.0001	5.53 (5.38, 5.68)
X_7	Nasal Congestion	3.2513	0.0206	<0.0001	25.93 (25.06, 26.82)

Results

COVID-19 Logistic Analytic Regression Model

In Table 1, we displayed the descriptive statistics of individuals with COVID-19 symptoms according to sex and age. It is shown that 185,400 (58.5%) were found positive. Out of the total number of N=316800, fever [n=99000 (31.25%)], tiredness [n=158400 (50%)], It is dry cough [n=179200 (56.25%)], difficulty breathing [n=158400 (50%)], sore throat [n=99000 (31.25%)], pains [n=115200 (36.36%)], nasal congestion [n=19800 (6.25%)], runny nose [n=172800 (54.54%)], and diarrhea [n=172800 (54.54%)].

After careful and rigorous data processing and analysis, employing some sophisticated statistical methods and machine learning, we significantly identify seven symptoms or risk factors that effectively and efficiently predict an individual COVID-19 status with 93% accuracy from the test data. Table 2 below displays the statistical significance of the seven identified symptoms of COVID19, along with their odd ratios. It is shown that each of the identified symptoms has a very high odd ratio greater than 1 and a positive coefficient, implying that each of the identified symptoms is positively associated with COVID-19.

The optimization process of the identified model successfully occurred after 8 iterations based on Newton's optimization method. The analytical COVID-19 logistic regression model that predicts the status of an individual COVID-19 based on the statistically significantly identified seven symptoms or risk factors is given by

$$\log\left(\frac{\pi}{1-\pi}\right) = -8.7478 + 2.4325X_1 + 2.4100X_2 + 2.1879X_3 + 2.3893X_4 + 2.4654X_5 + 1.7129X_6 + 3.2513X_7$$

where

$$\pi_i = \left(\frac{\exp(-8.7478 + 2.4325X_1 + 2.4100X_2 + 2.1879X_3 + 2.3893X_4 + 2.4654X_5 + 1.7129X_6 + 3.2513X_7)}{1 + \exp(-8.7478 + 2.4325X_1 + 2.4100X_2 + 2.1879X_3 + 2.3893X_4 + 2.4654X_5 + 1.7129X_6 + 3.2513X_7)}\right)$$

and

$$COVID-19 = \begin{cases} 1(test-positive), & if\ \pi_i \geq 0.5 \\ 0(test-negative), & otherwise. \end{cases}$$

Note that the weights that have been calculated for each of the seven symptoms (risk factors) that are incorporated in the model, statistically identified the test results as either positive or negative with 93% accuracy.

COVID-19 Logistic Analytic Regression Model Validation

As we stated earlier, the final model has the lowest AIC and BIC. The model prediction using the test data resulted in 93% prediction accuracy. To further validate the identified model, we performed sensitivity and specificity analysis of the proposed analytical model prediction using the test data. Figure 2 shows the heatmap of the confusion matrix for sensitivity and specificity analysis of the proposed model prediction accuracy and precision. The upper left position of the heatmap represents the true negatives (specificity), the lower left position presents the false negatives,
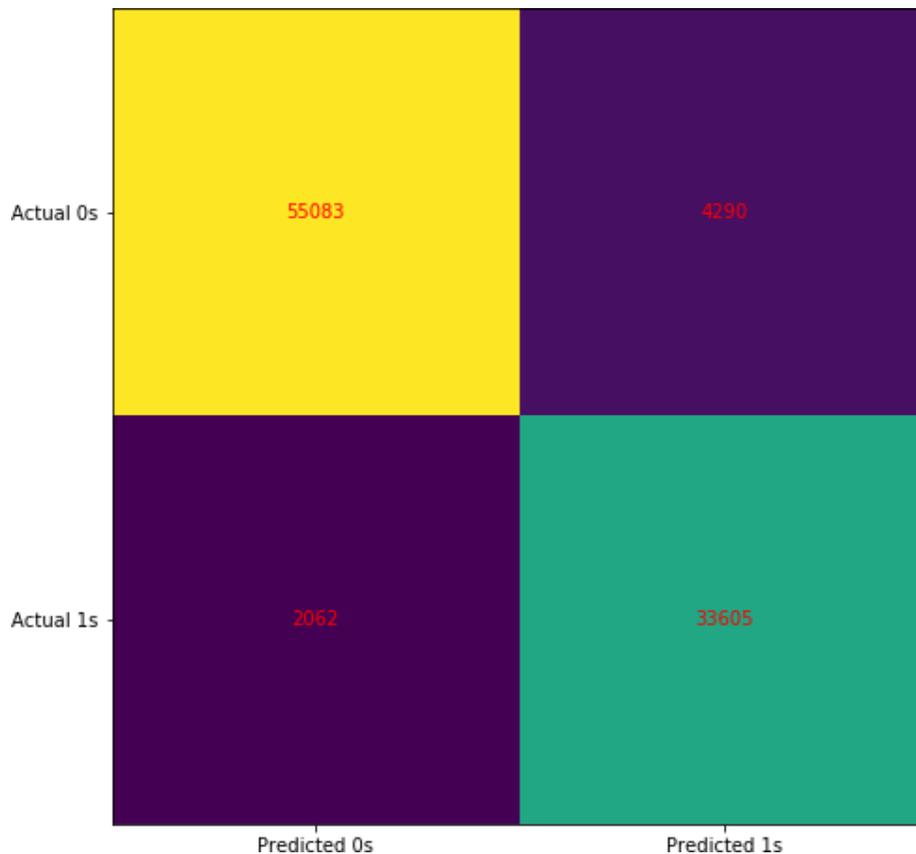
Figure 2: Heatmap of Confusion Matrix.

the top-right position denotes the false positives, and the lower right position denotes the true positives (sensitivity). Of the 95,040 test observations, the true COVID-19 positives and negatives are 33,605 and 55,083, respectively. The number of false positives and negatives is 4,290 and 2,062 observations, respectively. The sum of the false positive and negative individual equals a total number of 6,352 misclassified observations. Whereas, the sum of the true positive and negative observations equals a total of 88,688 well-classified individuals. Therefore, dividing 6,352 and 88,688 by the total individuals of the test data results in about 7% misclassification and 93% accurate classification by the proposed model, respectively. See Table 3 for more information on the classification report of the COVID-19 logistic regression model.

Table 3: Mode Prediction Classification Report.

Classification Report		
	Precision	**Support**
0 (COVID-19 Negative)	0.96	59373
1 (COVID-19 Positive)	0.89	35667
Accuracy	0.93	95040
Macro Avg	0.93	95040
Weighted Avg	0.93	95040

Another model validation method employed was the k−Fold cross-validation. We applied cross-validation of k = 10 folds. Thus, the data was divided into 10 folds with an equal number of observations. Within each fold, the model was trained with k−1 folds of the data, and the resulting model is validated on the remaining test data. The average of the prediction score or accuracy of the 10 folds was obtained and compared with the general prediction accuracy. The 10 − folds cross-validation resulted in an average classification prediction accuracy of 93%, which equals the original prediction score, attesting to the high quality of the proposed model. Table 4, below shows the prediction scores of the 10 − folds cross-validation. Figure 3 displays a receiver operating characteristic curve (ROC) commonly used with binary classification. The dotted line represents the ROC curve of a purely random classifier. For a good classification, the blue line stays as far away from that dotted line as possible (toward the top-left corner), as shown by our case.

Table 4: 10 − Folds Cross-Validation Prediction Accuracy.

10-Fold Cross-Validation										
Fold	1	2	3	4	5	6	7	8	9	10
Accuracy	0.935	0.937	0.933	0.931	0.936	0.929	0.930	0.935	0.932	0.931

We also applied L1 regularization to the coefficients to assess overfitting in the model, which may result from noise in the dataset. Regularization normally tries to reduce or penalize the complexity of the model. In logistic regression, regularization mostly tends to penalize the coefficient parameters. The L1 regularization penalizes the coefficients with the scaled sum of the absolute values of the weights; |b0|, |b1|, ..., |bk|. The application of the L1 regularization strength C = 10 resulted in the same coefficient values of the risk factors and the same prediction accuracy, a further indication of

the very high quality of the proposed analytical model. Note that a higher regularization value means weaker penalization, which relates to high values of β. If increasing the regularization value does not affect or change the coefficient values, implies minimum or no overfitting of the model. We applied different regularization strengths to assess the behavior of the coefficients. The values of the coefficients did not significantly change, indicating that there is no overfitting, and further attesting to the strength and quality of the COVID-19 proposed model.

Figure 3: Test for Classification Accuracy.
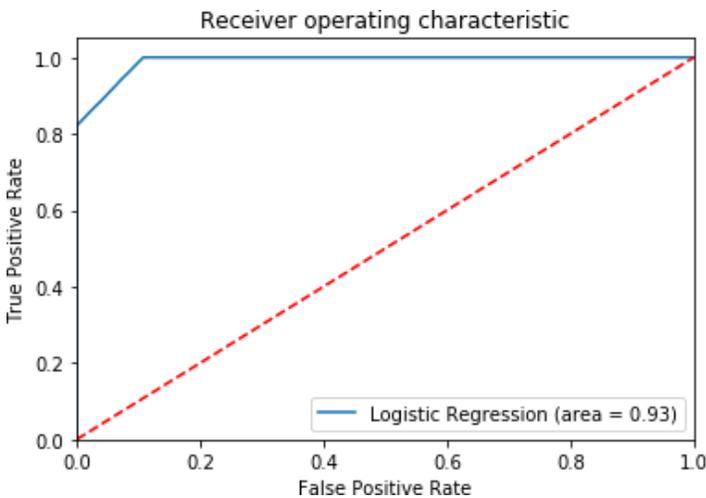
Discussion

In the present study, we developed an analytical model to predict the COVID-19 status of people based on the symptoms or risk factors identified with them. Our goal is to: (1) Identify the statically significant symptoms or risk factors of the COVID19 among all the risk factors published by WHO and the US CDC. (2) To ensure that scarce test kits are effectively and rightly used for people who are more likely to test positive, thereby enhancing efficiency, and allowing for testing to reach a vast number of highly infected or COVID-19 hot-spot communities. We developed a binary logistic analytic regression to predict the COVID-19 status of individuals (i.e. whether testing positive or negative) based on nine symptoms presumed by WHO and US CDC. We utilized a combined data of 316,800 from ten countries. Due to the absence of a response variable in the data, we created an artificial COVID-19 status response for individuals to be classified positively based on having at least four of the published symptoms, otherwise, they are classified as negative. We proceeded to develop the COVID-19 logistic regression model in which we statistically identified seven symptoms or risk factors out of nine to be significantly associated with COVID-19. The identified symptoms were found after a careful and rigorous data analysis utilizing the stepwise backward elimination model selection method to select the best model with the least AIC and BIC. The proposed COVID-19 logistic model has 93% classification prediction accuracy. The significant symptoms identified include fever, tiredness, dry cough, difficulty in breathing, sore throat, pains, and nasal congestion. The developed model identifies (estimated) the different weights of each of the seven symptoms or risk factors that play a major role in the decision process of the testing results. Also, each of the symptoms identified showed very high increased odds of tested positive for COVID-19, given by the odds ratios greater than 1. Thus, individuals identified with any of the symptoms are highly likely or at risk to be infected by COVID-19. Although the odds ratio of each symptom in Table 2 showed an increased likelihood of an individual having COVID-19, the rate of increase is not the same for all the symptoms. The higher the odds ratio, the more likelihood of getting tested positive. For instance, Nasal Congestion was found to have the highest odds ratio of 25.95, followed by Sore Throat with an odds ratio of 11.45, and Pains has the least odds ratio of 5.53. That is, if a person is identified with Nasal Congestion, he/she is more at risk of been infected with COVID-19 than another person with Pains. Also, the positive coefficient of identified symptoms implies that all the symptoms are positively associated with COVID-19. These symptoms we identified are consistent with those published by the US CDC [2]. However, runny nose and diarrhea were not identified to have a statistically significant association with COVID-19 in our analysis, hence were excluded in the proposed COVID-19 logistic analytic regression model. Our findings contradict two of the three additional symptoms recently published by the CDC [17]. This may be because the COVID-19 is still spreading to uninfected population, with new cases identified on a daily basis. The medical professions are continuously learning more about the disease, and new information is continuously being discovered about the disease.

We validated and assessed the quality of the proposed model employing sensitivity and specificity analysis, 10 − folds cross-validation, and L1 regularization methods. Apart from the model having the least AIC and BIC, the sensitivity and specificity analysis shows that only 6,352 out of 95,040 individuals were misclassified for their COVID-19 status, indicating that 93% of individuals in the test data were accurately or correctly classified. The 10 − folds cross-validation also resulted in an average prediction score or accuracy of 93%, which is consistent with the previous results. The L1 regularization was employed to penalized any possible overfitting by the model parameters. The penalization term did not change the initial values of the coefficients, a good indication of goodness-of-fit of the proposed model. Also, it is fascinating to recognize that the results of all the ten countries combined were consistent with that of individual countries. Our proposed model can be updated as more data and possibly new proven symptoms become available to increase the very high accuracy of the test results of the proposed model. It is important to note that the model does not apply to COVID-19 asymptomatic individuals. Our findings will provide further improvement in the therapeutic/treatment and preventive process of the COVID-19 disease. Future research needs to focus on understanding possible differences in the results we found by socio-demographic characteristics such as age, race, and gender. These factors have been known to affect survival from other health outcomes including both Infectious and noninfectious diseases [18-20], which is important given symptoms could impact survival. Also, survival analysis has been

conducted for other disease outcomes such as cancer [21-23]. Therefore, future research will focus on performing the survival analysis of individuals diagnosed with COVID-19.

Conclusion

We developed a high quality and well-validated COVID-19 analytical logistic regression model that statistically significantly identified seven symptoms or risk factors that predict the COVID-19 status of all individuals tested positive or negative. The identified symptoms include fever, tiredness, dry cough, difficulty in breathing, sore throat, pains, and nasal congestion. The proposed model can be used to predict the COVID-19 status of an individual with 93% accuracy based on the identified symptoms without been tested with the test kit. The developed model identifies (estimated) the different weights of each of the seven symptoms or risk factors that play a major role in the decision process of the testing results. The proposed model provides a supplementary mechanism for COVID-19 screening. It will speed up and enhance the efficiency of the testing process by classifying people who are tested for positive or negative COVID-19. It will further ensure the judicious distribution of the already scarce COVID-19 test kits around the globe, enhancing efficacy and efficiency of resource distribution. Once an individual is diagnosed with the identified symptoms, without testing with the test kit but rather using the proposed model, the appropriate treatment or preventive measures can be taken to stop the possibility of spreading the disease to others. Also, the subject testing will be useful or important for individuals in areas or settings that do not have immediate access to the medical test instruments for checking their COVID-19 status upon experiencing some of the outlined identified symptoms. This model can serve public health importance for only individuals with suspected symptoms of COVID-19. However, the model cannot be used for asymptomatic individuals. Finally, once an individual test positive, he/she can decide to self-quarantine or immediately seek medical treatment. Attached in the Appendix is the questionnaire to be administered for the COVID-19 test based on the identified risk factors.

Acknowledgment

References

1. Johns Hopkins Medicine. COVID-19 Update. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus

2. World Economic Forum COVID Action Platform. https://www.weforum.org/agenda/2020/04/covid-19-does-to-your-body/

3. Lenny Bernstein and Ariana Eunjung Cha (MAY 10, 2020). The Washington Post. Doctors keep discovering new ways the coronavirus attacks the body. Damage to the kidneys, heart, brain - even covid toes-prompts reassessment of the disease and how to treat it. https://www.washingtonpost.com/health/2020/05/10/coronavirusattacksbodysymptoms/?arc404=true

4. Merk Serono, Novartis, Roche and Celgene. BartMS Blog, COVIDMS GUEST POST. https://multiple-sclerosis-research.org/2020/03/ mscovid19-guest-post/

5. Center for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19), Symptoms. https://www.cdc.gov/coronavirus/2019-ncov/ symptoms-testing/symptoms.html

6. Grenfell R, Drew T (17 February, 2020). "Here's Why It's Taking So Long to Develop a Vaccine for the New Coronavirus". ScienceAlert. Archived from the original on 28 February 2020. Retrieved 26 February 2020. https://www.sciencealert.com/who-says-a-coronavirus-vaccine-is-18-months-away

7. 2019 Novel Coronavirus 2019-nCoV Situation Summary. Centers for Disease Control and Prevention. 30 January 2020. Archived from the original on 26 January 2020.

8. Interim Guidelines for Collecting Handling and Testing Clinical Specimens from Persons for Coronavirus Disease 2019 COVID-19. Centers for Disease Control and Prevention CDC. 11 February 2020. Archived from the original on 4 March 2020. Retrieved 26 March 2020.

9. Curetis Group Company Ares Genetics and BGI Group Collaborate to Offer Next-Generation Sequencing and PCR-based Coronavirus 2019-nCoV Testing in Europe. GlobeNewswire News Room. 30 January 2020. Archived from the original on 31 January 2020. Retrieved 1 February 2020.

10. Coronavirus COVID-19 Update FDA Issues first Emergency Use Authorization for Point of Care Diagnostic Press release. US Food and Drug Administration FDA. 21 March 2020. Archived from the original on 21 March 2020. Retrieved 22 March 2020.

11. Diamond, Michael S, Pierson, et al. The challenges of vaccine development against a new virus during a pandemic. Cell Host and Microbe. 2020.

12. CEPI welcomes UK Government's funding and highlights need for $2 billion to develop a vaccine against COVID-19. Coalition for Epidemic Preparedness Innovations Oslo Norway. 6 March 2020.

13. Damon Wake. World leaders urge cooperation in vaccine hunt raise $8 billion. Yahoo Finance. 2020.

14. Update on WHO Solidarity Trial Accelerating a safe and effective COVID-19 vaccine.

15. World Health Organization. Advice for Public. Archived from the original on 26 January 2020. Retrieved 10 February 2020.

16. Recommendations for Cloth Face Covers. Centers for Disease Control and Prevention. 3 April 2020.

17. Jak Connor, Tweak Town. June 26, 20202. 3 new coronavirus symptoms officially added to the list, here they are; The Centers for Disease Control and Prevention has added three new coronavirus symptoms to the list. Read more: https://www.

tweaktown.com/news/73435/ 3-new-coronavirus-symptoms-officially-added-to-the-list-here-they-are/ index.html

18. Utuama O, Mukhtar F, Pham YTH, et al. Racial/ethnic age and sex disparities in leukemia survival among adults in the United States during 1973-2014 period. 2019; 14: e0220864.

19. Mukhtar F, Boffetta P, Dabo B, et al. Disparities by race age and sex in the improvement of survival for lymphoma Findings from a population-based study. PloS one. 2019; 13: e0199745.

20. Tim Mackey, Vidya Purushothaman, Jiawei Li, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access and Recovery Associated With COVID-19 on Twitter Retrospective Big Data Infoveillance Study. JMIR public health and surveillance 2020; 6: e19509.

21. Lohuwa Mamudu, Chris P. Tsokos. Parametric and Non-Parametric Analysis of the Survival Times of Patients with Multiple Myeloma Cancer. Open Journal of Applied Sciences. 2020; 10: 118-134.

22. Lohuwa Mamudu, Chris P Tsokos, Otunuga Oluwaseun E. Survival Analysis of Multiple Myeloma Cancer Using the Cox-PH Model. Medical Clinical Research Journal. 2020.

23. Lohuwa M, Tsokos C. Data-Driven Statistical Modeling and Analysis of the Survival Times of Multiple Myeloma. Health Science Journal. 2020; 14: 1.

APPENDIX A

TESTING FOR COVID-19

DATE: _____

APPLICANT ID: _____

APPLICANT NAME:_____

Please answer the following questions regarding your present health symptoms. Answer **Yes** or **No** by checking the appropriate box that apply to you. Thank you.

1. Do you presently have fever? ☐ Yes ☐ No

2. Do you feel tiredness? ☐ Yes ☐ No

3. Do you experience dry cough? ☐ Yes ☐ No

4. Do you experience a sore throat? ☐ Yes ☐ No

5. Do you have difficulty breathing? ☐ Yes ☐ No

6. Do you experience any pain, anywhere in your body? ☐ Yes ☐ No

7. Do you have nasal congestion? ☐ Yes ☐ No

Officials use only:

COVID-19 Results: ☐ Test Positive ☐ Test Negative