Using Machine Learning Techniques to Predict COVID-19 Patient Outcomes

Noha Alnazzawi

Computer Science and Engineering Department, Yanbu University College, Royal Commission for Jubail and Yanbu, Yanbu Industrial City, Saudi Arabia

Abstract. The recent outbreak of coronavirus disease 2019 (COVID-19) has affected human life to a great extent. The pandemic has led to severe global socioeconomic disruption, causing the postponement or cancellation of major events. COVID-19 is a novel and accelerating outbreak. The virus spreads very quickly and has reached almost every part of the world with more than 100 million reported cases, and critical healthcare resources are limited. Therefore, it is very important to predict which patients are most likely to develop severe illness and face the greatest risks of complications, including death. In this study, we compare the performances of different machine learning algorithms in predicting COVID-19 patient outcomes based on combinations of patient risk factors. The best-performing algorithm is the random forest classifier, which achieves an F-score of 0.788 and an accuracy rate of 0.789. The proposed model in this study is able to predict the outcome (i.e., dead, discharged, or stable) for any patient diagnosed with COVID-19 by using the same set of risk factors, namely, gender, country, symptoms, and chronic diseases. The findings of this study can supplement clinical skills and assist doctors in predicting unexpected patterns to identify mild cases among diagnosed patients and the few cases that will progress to severe illness

Keywords: Terms-Artificial Intelligence, COVID-19 outcomes, Machine Learning, Predictive Model.

1. Introduction

In the past few decades, there have been several outbreaks of viral infections. Among them, the most significant were SARS-CoV-1 in 2003, H1N1 influenza (swine flu) in 2009, MERS (Middle East respiratory syndrome) in 2012, and Ebola in 2014. The SARS-CoV-1 pandemic in 2003 showed the organizational incompetence of many governments and healthcare systems in providing adequate and timely responses. This incompetence was particularly pronounced in China, where organizational problems within the Chinese Center for Disease Control and Prevention led to a delay in the identification of the causative agent [1]. The swine flu virus was isolated

much earlier than the outbreak, which started in April 2009 in Mexico and the USA, and it spread all over the globe in a few weeks. It was marked as the first pandemic in 40 years [2]. Since influenza viruses are known to have periodic outbreaks, this pandemic had been anticipated, but no one knew which type of influenza would appear and how widely it would spread [2]. The MERS and Ebola outbreaks were not pandemics, but these diseases have high mortality rates.

According to the World Health Organization (WHO), coronavirus disease 2019 (COVID-19) is an infectious respiratory disease caused by a newly discovered coronavirus (SARS-CoV-2) [3]. The SARS-CoV-2 outbreak

started in December 2019 in the city of Wuhan, which is located in Hubei Province, China. Chinese epidemiologists first announced a rapid increase in pneumonia cases of unknown origin [4]. Current epidemiological data suggest that the first cases were related to the Huanan Seafood Wholesale Market in Wuhan, where the virus probably underwent mutations that enabled it to infect humans who came into contact with wild animals [5]. Within a short period of time, SARS-CoV-2 became capable of human-to-human transmission via respiratory droplets in close contact when the infected person sneezes or coughs. The virus spreads very quickly and has reached almost every part of the world, with more than three million cases reported in less than 4 months, constituting a pandemic [6]. The WHO declared the outbreak to be a Public Health Emergency of International Concern on 30 January 2020 and recognized it as a pandemic on 11 March 2020 [7]. Symptoms of the disease appear in people within 14 days of exposure to the virus. The symptoms range from mild to severe, and the clinical picture varies from an asymptomatic carrier state to severe acute pneumonia. The most common symptoms include cough, fever and shortness of breath, and in some cases, COVID-19 can be severe and lead to serious problems such as severe respiratory problems, kidney failure or death.

The risk factors for COVID-19 are not yet clear, although elderly people, immunocompromised patients, and patients with chronic diseases are most likely to develop serious illness [8], [9].

Currently, there is no specific treatment for the virus. Treatment is directed at relieving symptoms and providing supportive therapy [9], [10]. It is also not possible to predict with a high level of certainty how the pandemic will settle down. This situation depends on social distancing measures, the efficacy of antiviral drugs and the duration of

acquired immunity in patients who have recovered from the infection.

It is certain that the current pandemic has already caused disturbances in all aspects of society. Educational institutions are closed, production at manufacturing facilities has decreased, and industries capable of switching to online business have already taken those steps. Additionally, resources are being redirected to healthcare. Although there are many theories, at this moment, no one can be sure what the world economy will look like after the pandemic settles down.

The abovementioned points have arisen due to the COVID-19 pandemic. The global pandemic has overwhelmed existing medical facilities and resources and forced doctors to make crucial decisions with limited time and information. Therefore, the COVID-19 crisis presents a critical situation that demands the implementation of effective machine learning (ML) based data mining techniques to guide doctors in the decision-making process by processing a given patient's data to recommend the best course of action. ML algorithms employ a variety of methods, including statistical, probabilistic and optimization methods, to learn from past experiences and detect useful patterns from large and complex datasets [11], [12]. Therefore, ML techniques have become increasingly practical to apply and explore in the field of disease prediction.

In this study, we compare the performances of different ML algorithms in predicting the potential health outcomes for patients who have been diagnosed with COVID-19, with each outcome belonging to one of the following classes: dead, discharged, or stable. The ML models are trained on combinations of risk factors (*i.e.*, age, gender, country, symptoms and chronic diseases) for COVID-19. The ML model proposed in this study can supplement clinical skills and assist

in the decision-making process in healthcare settings. For example, the proposed ML model can be used to predict the severity of illness so that low-risk patients can be separated from those at higher risk to manage the patient volumes of hospitals by reducing admission to already overwhelmed hospitals and allocating resources based on the severity of the case experienced by a given patient [13].

2. Related Work

The rapid increase and proliferation of electronic health record EHR systems, coupled with advances in data mining techniques, provides a unique opportunity to make contributions in healthcare [14].

Currently, there are many diseases that need to be identified during their early stages to provide optimal treatments and avoid severe complications. Therefore, there is a need to analyze the large amount of available and complex medical data in a short time period but with great accuracy. Data mining techniques can be applied to patient data (e.g., clinical symptoms, medical histories, risk factors, laboratory results, medication information and demographic data), and they have great potential for exploring hidden patterns and relationships that are not visible to humans [13]. Supervised ML techniques have been dominant methods in the data mining field and can facilitate the discovery of new knowledge and associations hidden in large amounts of data [13], [15]. Such approaches have been widely used in the medical domain; their implications are manifested in the areas of disease prognosis and prediction [11], [16-18], chronic disease surveillance [19] and the comparison of disease prevalence and drug outcomes [20]. For example, a variety of ML techniques, including artificial neural networks (ANNs), Bayesian networks (BNs), support vector machines (SVMs) and decision trees (DTs), have been widely used in cancer

research for prediction and prognosis purposes, resulting in effective and accurate decision making [21-24]. Hierarchical clustering has been used in the investigation of Alzheimer's disease [15]. Additionally, ANNs have been used for classifying different subtypes of psychogenic nonepileptic seizures [13], [25], [26].

Since the beginning of the COVID-19 outbreak, researchers from around the world have started to publish research summarizing their findings and observations about COVID-19. However, there is no clear picture of the disease. Based on the best available scientific evidence, the understanding of the transmission patterns, severity, clinical features and risk factors of COVID-19 is currently limited among the general population, among healthcare workers, within households and in other "closed" settings [27]. Therefore, knowing more information about the risk factors that put patients at higher risk of developing severe illness is very important for guiding clinicians to ensure that patients with the greatest or most urgent needs receive services first and to lower the potential development of severe illness.

Most of the previous works in the COVID-19 domain targeted social media posts to understand how people perceive severe outbreaks [28], [29]. For example, Li *et al.* [30] used Twitter datasets containing 1000 tweets in English to train a model to classify each tweet as one of the following emotions: anger, anticipation, disgust, fear, joy, sadness, surprise or trust. The purpose of their study was to measure the impact of COVID-19 on people's mental health [30].

Clinical information, such as laboratory results and computed tomography (CT) scans, has also been investigated by researchers to predict COVID-19 cases. For example, a study by Gifani *et al.* [31] automatically detected positive COVID-19 cases by using an

ensemble deep transfer learning system with different pretrained convolutional neural network (CNN) architectures on CT scans labeled positive or negative for COVID-19. The system can identify scans that are normal, those that contain COVID-19, and those that contain other types of lung diseases. Muhammad *et al.* [32] trained different ML methods on epidemiologically labeled datasets. The dataset contains a combination of clinical features (*i.e.*, pneumonia, diabetes, asthma, hypertension, cardiovascular diseases, obesity, chronic kidney diseases) as well as the lab reverse transcription polymerase chain reaction (RT-PCR) testing results for COVID-19 in terms of positive and negative cases in Mexico, and the proposed predictive model can be used to diagnose patients with COVID-19 [32]. Another example is a study by Kassani *et al.* [33], in which ML algorithms were used on CT images to predict abnormalities in the images of patients diagnosed with COVID-19.These predictive algorithms can aid clinicians in diagnosing COVID-19 and help provide a better understanding of the progression of the disease.

Shared tasks and competitions have also been organized to motivate researchers around the world to conduct research on COVID-19 and share their findings. Two examples are a text mining competition hosted by Kaggle [34] and an information retrieval shared task hosted by TREC [35]. Both shared tasks have no evaluation metric defined for assessing the submissions. Instead, the results regarding automated extraction and retrieval need to be judged by biomedical and medical experts [36].

All previous works used ML techniques to predict or diagnose patients as being negative or positive for COVID-19. To the best of our knowledge, our work is the first to use combinations of demographic (*i.e.*, age, gender, and country) and clinical features (*i.e.*, symptoms and chronic diseases) to predict the outcomes (dead, discharged, or stable) of patients who have already been diagnosed with COVID-19.

3. Dataset

The novel coronavirus (COVID-19) case dataset used in this research was obtained from the John Hopkins University Center for Systems Science and Engineering (JHU CSSE). The COVID-19 case dataset was extracted in CSV format from the Johns Hopkins University GitHub repository for the period from 1/1/2020 to 20/3/2020. The data were aggregated from multiple credible sources of real affected cases to track the spread of COVID-19. The CSV file consists of 44 columns of data about the affected patients, including their ages, genders, countries, symptoms and chronic diseases [37].

As the purpose of this study is to predict the outcome of a patient based on a combination of risk factors, the data were filtered to include only the columns that describe the risk factors that affect patient outcomes and remove the columns that did not directly describe risk factors (e.g., hospital_visit_date, exposure_start, exposure_end, visiting Wuhan, from Wuhan). These clinical characteristics were chosen after close consultation with medical doctor who revised the data and advised us to keep only 6 out of the original 44 columns. Those columns were chosen by the medical doctor as they present indicators of different combinations of risk factors, including demographics (*i.e.*, age, country and gender) and medical risk factors (*i.e.*, symptoms and chronic diseases).

The data were filtered further to only keep patients who had clear outcomes: dead, discharged from hospital or stable with mild symptoms. After applying this filter, the records for only 2,228 patients were retained. The patients' ages, genders, countries, symptoms and chronic diseases were used as predictive variables to predict patient outcomes.

Table 1 shows the distribution of the outcome classes. Table 2 shows a list of the symptoms and chronic diseases used in this study.

As shown in Fig. 1-3, the most common symptoms across all the patients were fever, followed by cough and sore throat. For chronic diseases, as shown in Fig. 4, the most common chronic diseases among the patients who died were: hypertension, followed by diabetes. As shown in Fig. 5 and 6, most of the discharged and stable patients did not have any chronic diseases. This suggests that patients with no chronic diseases are more likely not to develop severe illness due to their COVID-19 infection. The data were converted into dummy variables, where 1 indicated the presence of the medical condition and 0 indicated the absence of the medical condition.

Table 1. Distribution of the Outcome Classes.

Outcome class	Number of cases in the datasets
Stable	1262
Discharged	812
Died	154

Table 2. List of the Symptoms and Chronic Diseases Used in this Study.

Symptoms	Chronic diseases
abdominal pain, acute coronary syndrome, acute left heart failure, acute pharyngitis, acute respiratory viral infection, anorexia, asymptomatic, chest discomfort, chest distress, chest pain, chest tightness, chills, conjunctivitis, cough, cough with sputum, diarrhea, difficulty walking, dizziness, dry cough, dry throat, dyspnea, esophageal reflux, eye irritation, fatigue, fever, flu, headache, joint pain, lesions on chest radiographs, loss of appetite, malaise, muscle ache, muscular soreness, myalgia, nausea, pharyngalgia, pleural effusion, pleuritic chest pain, pneumonia, respiratory symptoms, rhinorrhea, sneezing, sore body, sore limbs, sore throat, sweating, toothache, vomiting, weakness, wheezing	asthma, COPD, coronary bypass surgery, coronary heart disease, diabetes mellitus, hepatitis B, HIV positive, hypertension, hypertriglyceridemia, lung cancer, pneumonia, prostate hypertrophy

4. Methods and Results

The dataset was randomly divided into an 80% training set and a 20% testing test to validate the abilities of the ML-based models to predict patient outcomes.

We compared the performance of various ML models known for their excellent performances in similar tasks [20], [34], [38]. We employed the scikit-learn implementations of the following algorithms following their default settings: Extra Trees, Random Forest, Hist Gradient Boosting, Multi-Layer Perceptron (MLP) classifier, Logistic Regression and Decision Tree [39]. The performances of the different models were compared based on their precision, recall, F-score and accuracy results.

As shown in Table 3 and Fig. 7, the best-performing model was the random forest classifier, followed by ExtraTrees. The random forest classifier achieved the best results in terms of predicting patient outcome statuses based on the selected features, as explained above.

An error analysis showed that most of the FPs were found to correspond largely to cases where patients with similar symptoms and chronic diseases experienced different outcomes. We observed that patients with the same combinations of risk factors but in different gender or age groups had different outcomes, which suggests that the reactions of patients diagnosed with COVID-19 depend on age and gender. Another phenomenon we observed was that patients with similar risk factors but from different countries sometimes experienced different outcomes, which suggests that the healthcare systems in some countries are less adequate than those in others. For the deaths, a low number of deaths made

the model less robust and unable to obtain enough information to learn from the data and predict outcomes in unseen cases.

Overall, our experiments demonstrated promising results, and we believe that training our model on larger datasets would lead to better results. Our proposed model has the potential to assist hospital doctors in the clinical decision-making process by separating low-risk patients from those at higher risk to manage the patient volumes of hospitals by reducing admission to already overwhelmed hospitals and allocating resources. Depending on the severity of the combinations of patients' risk

factors and chronic diseases, the ML model can recommend discharging the patient or admitting them to the hospital. In the future, to increase the accuracy of our model, we are planning to incorporate more features, such as laboratory and CT scan results.

To demonstrate the usability of our trained model, we developed a web-based interface where a user can enter the required information, as shown in Fig.8. The entered information is used as features to query our trained model to predict the potential outcome for the COVID-19 patient.
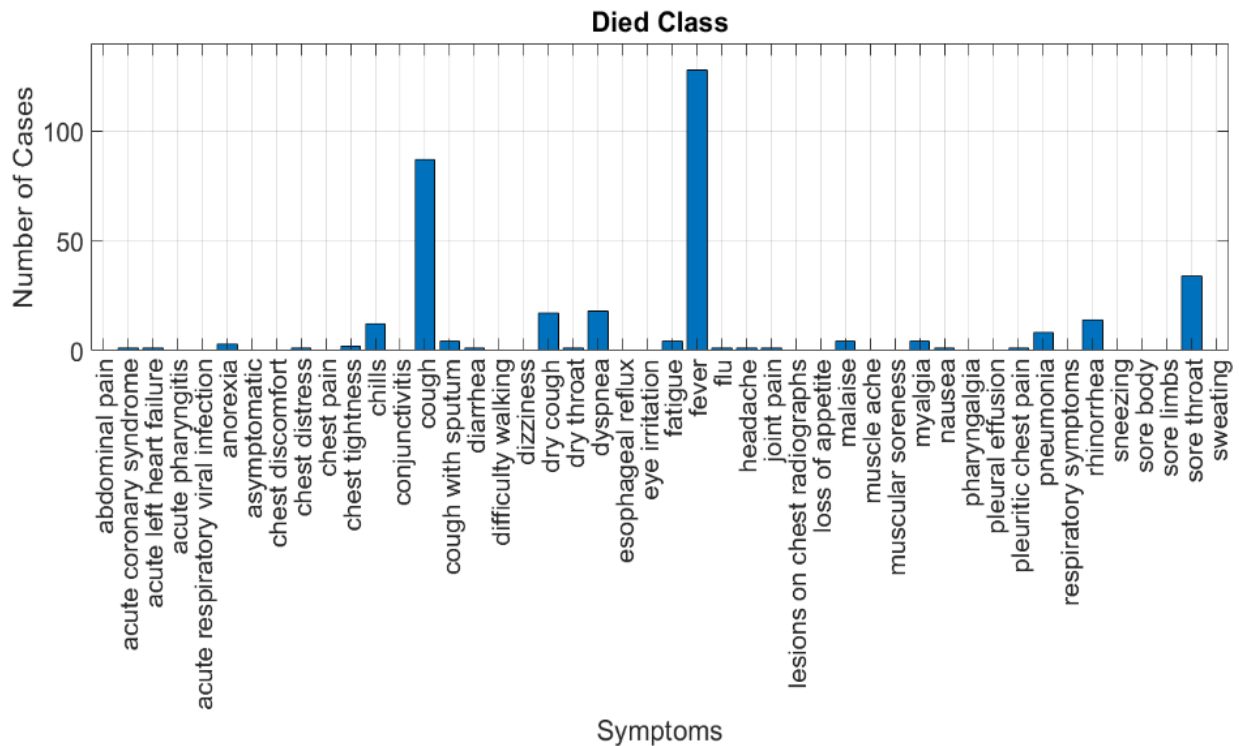

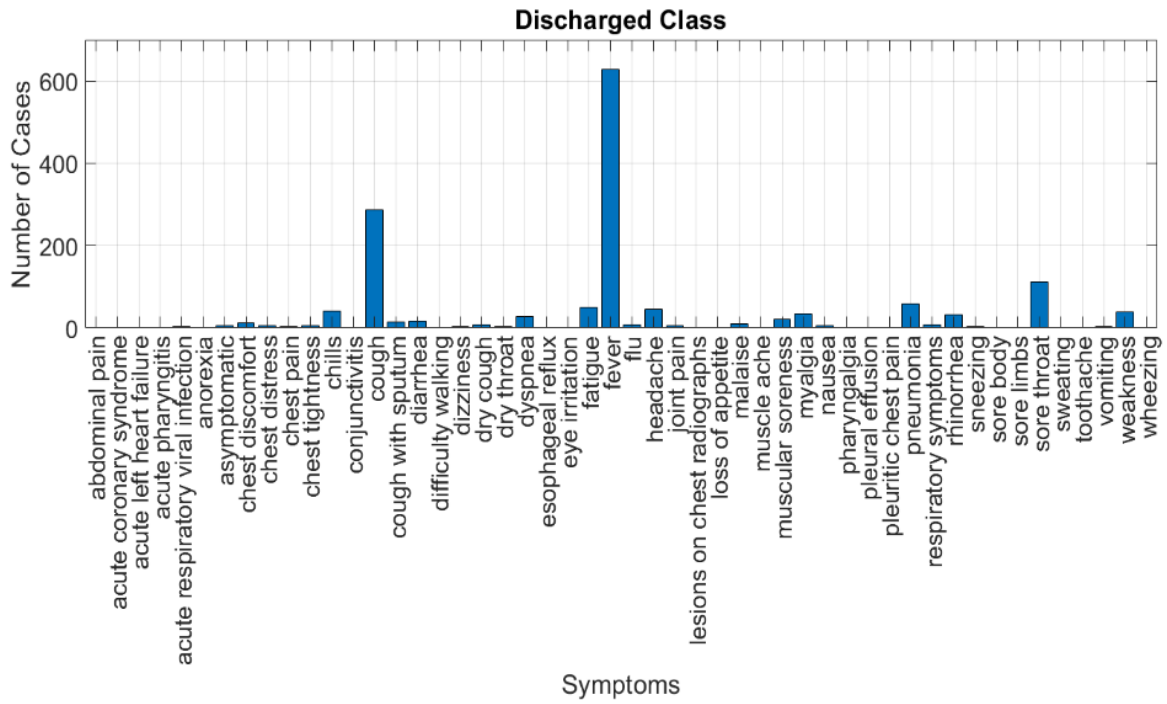Fig. 1. The most common symptoms among the patients who died.

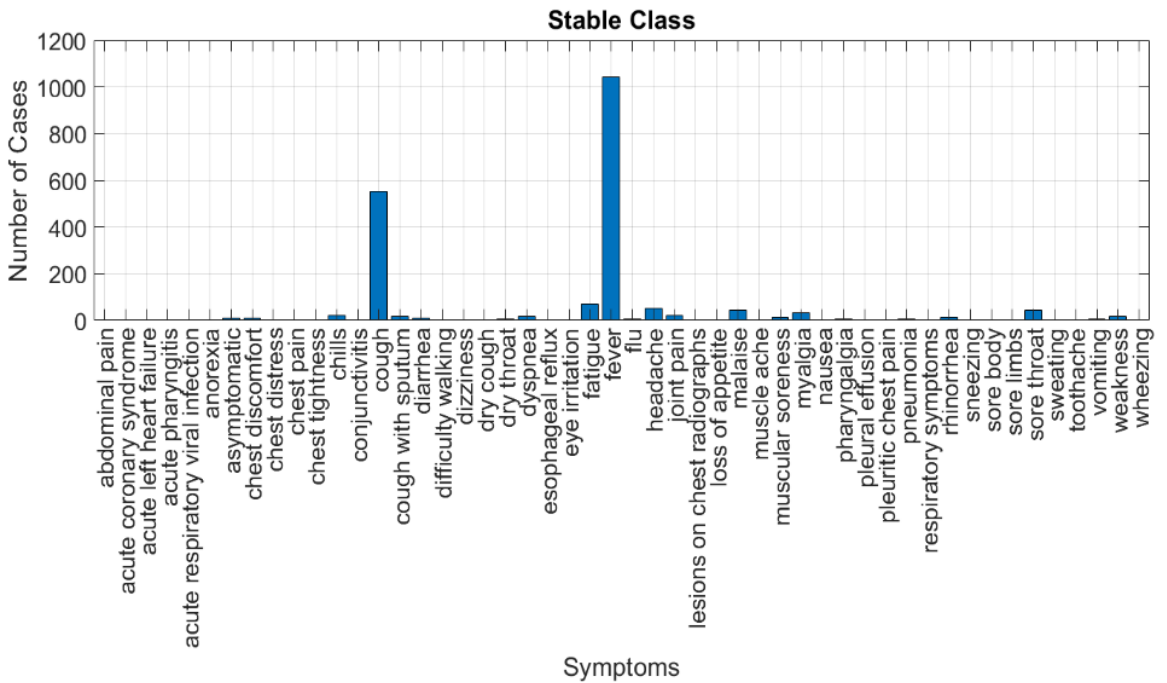Fig. 2. The most common symptoms among the discharged patients.

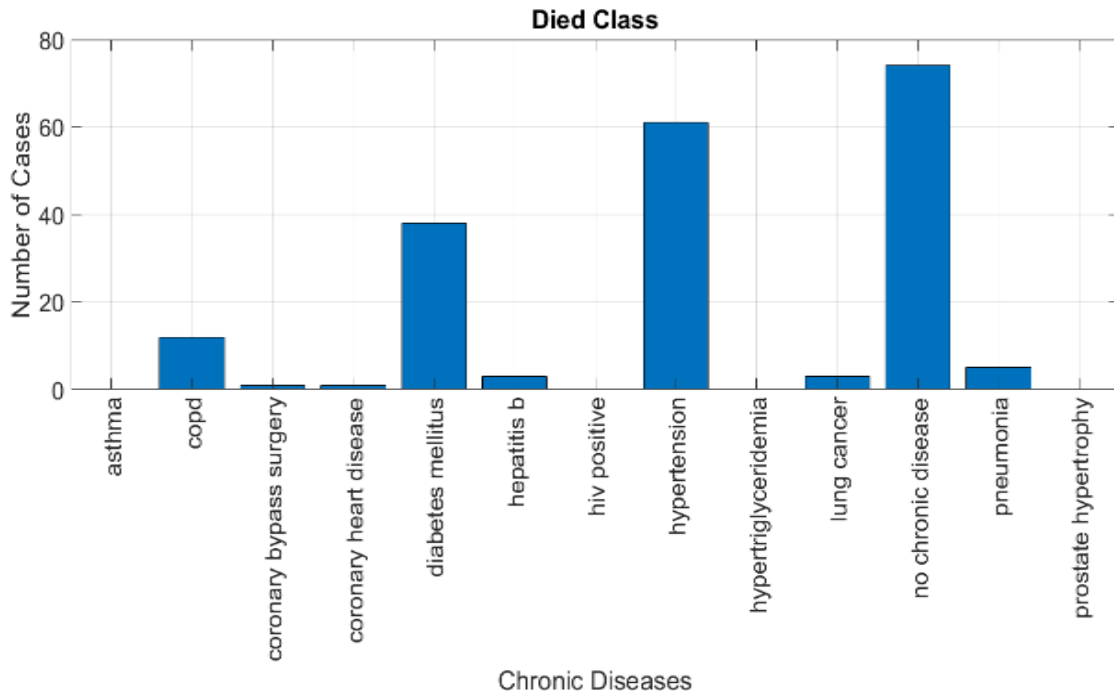Fig. 3. The most common symptoms among the stable patients.

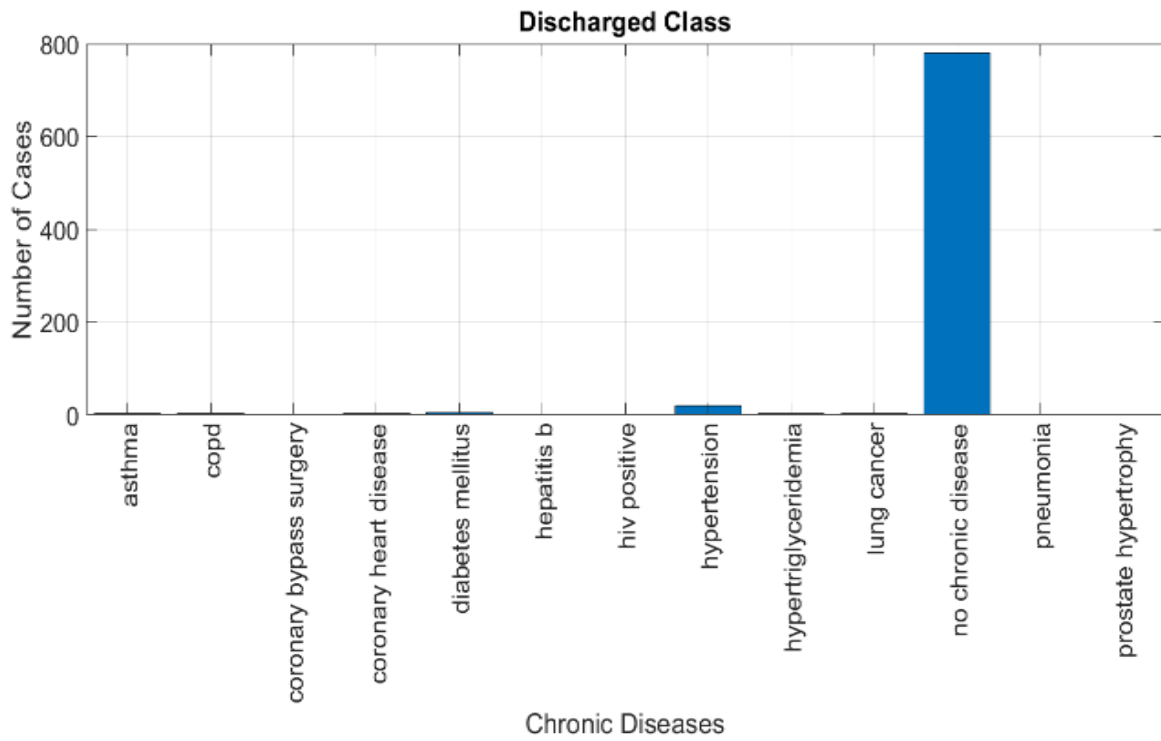**Fig. 4. The most common chronic diseases among the patients who died.**

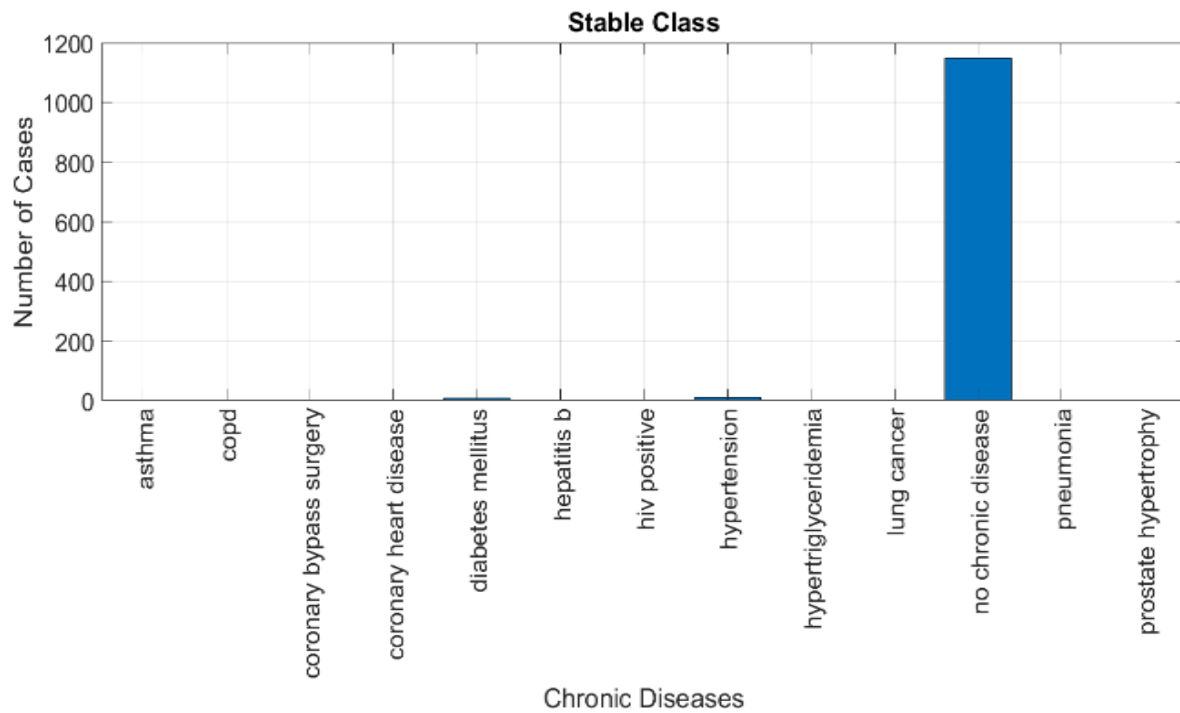**Fig. 5. The most common chronic diseases among the discharged patients.**

Fig. 6. The most common chronic diseases among the stable patients.

Table 3. Comparison of the Different Methods Developed and Evaluated on the COVID-19 Dataset.

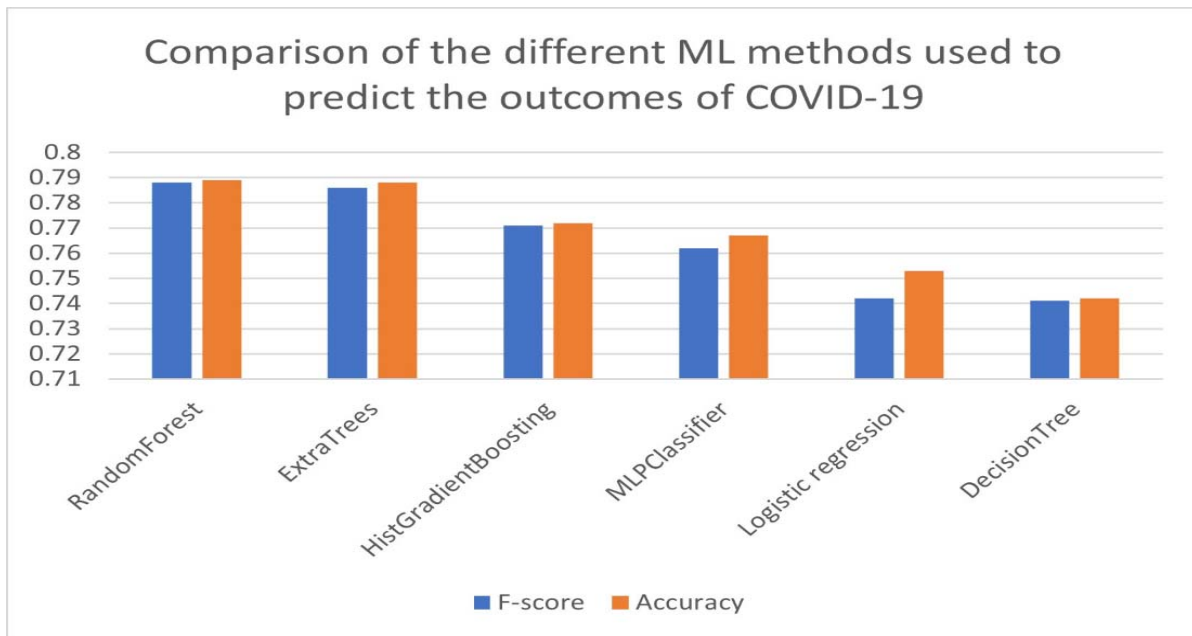| Model	Precision	Recall	F-score	Accuracy
Random Forest Classifier	0.789	0.788	**0.788**	**0.789**
Extra Trees Classifier	0.787	0.788	0.786	0.788
Hist Gradient Boosting Classifier	0.771	0.772	0.771	0.772
MLP Classifier	0.767	0.767	0.762	0.767
Logistic regression	0.762	0.753	0.742	0.753
Decision Tree Classifier	0.741	0.742	0.741	0.742

Fig. 7. Comparison of different ML classifiers used to predict the outcomes of COVID-19 in terms of F-score and accuracy.

Fig. 8. An interface to query our trained model to predict the potential outcome for a COVID-19 patient.

5. Conclusion

COVID-19 is a novel and accelerating outbreak. Knowing detailed information about the disease risk factors and how these risk factors can contribute to predicting potential outcomes is very important. This study compared the performances of different ML algorithms on a dataset obtained from JHU CSSE to predict patient outcomes based on risk factor combinations. The best-performing algorithm was the random forest classifier, which achieved an F-score of 0.788 and an accuracy rate of 0.789. The ML model generated by this study is promising and will benefit clinicians, decision-makers, and researchers by providing an improved understanding of patient symptoms to prioritize certain patients in clinical settings. Gaining access to larger datasets and incorporating more clinical features, such as laboratory and CT scan results, will be very important for increasing the accuracy of the proposed model.

References

[1] **Enserink, M.** (2003). "SARS in China. China's missed chance," *Science*, **301**(5631): 294–296, Jul. 2003, doi: 10.1126/science.301.5631.294.

[2] **Al-Muharrmi, Z.** (2010). "Understanding the influenza A H1N1 2009 pandemic," *Sultan Qaboos Univ. Med. J.*, 10(2): 187–195.

[3] **World Heath Orgnization (WHO)**, "*CoronaVirus overview*." https://www.who.int/health-topics/coronavirus#tab=tab_1.

[4] **Ciotti, M.** *et al.* (2020). "COVID-19 outbreak: An overview," *Chemotherapy*, **64** (5-6): 215–223, Apr., doi: 10.1159/000507423.

[5] **Lu, H., Stratton, C.** W. and **Tang, Y. W.** (2020). "Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle," *J. Med. Virol.*, **92** (4): 401–402, Apr., doi: 10.1002/jmv.25678.

[6] **Mackenzie, J. S.** and **Smith, D. W.** (2020). "COVID-19: A novel zoonotic disease caused by a coronavirus from China: What we know and what we don't," *Microbiology Aust.*, p. MA20013, Mar., doi: 10.1071/MA20013.

[7] **World Health Orgnization (WHO)**, "*WHO Director-General's opening remarks at the media briefing on COVID-19.*" https://www.who.int/dg/speeches/detail/who-director-

general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

[8] **Wang, D.** *et al.* (2020). "Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China," *JAMA*, **323**(11): 1061–1069, Feb., doi: 10.1001/jama.2020.1585.

[9] **Centers for Disease Control and Prevention**, "*Global COVID-19 pandemic notice.*" https://wwwnc.cdc.gov/travel/notices/warning/coronavirus-global.

[10] **Mayo Clinic**, "*Coronavirus disease: What is it and how can I protect myself?*" https://www.mayoclinic.org/diseases-conditions/coronavirus/expert-answers/novel-coronavirus/faq-20478727.

[11] **Uddin, S., Khan, A., Hossain, M. E.** and **Moni, M. A.** (2019). "Comparing different supervised machine learning algorithms for disease prediction," *BMC Med. Inform. Decis. Making*, **19**(1): 281, Dec., doi: 10.1186/s12911-019-1004-8.

[12] **Mitchell, T. M.** (1997). *Machine Learning WCB*. Boston, MA: McGraw-Hill Boston,

[13] **Jayatilake, S. M. D. A. C.** and **Ganegoda, G. U.** (2021). "Involvement of machine learning tools in healthcare decision making," *J. Healthcare Eng.*, **2021**: 6679512, Jan., doi: 10.1155/2021/6679512.

[14] **Avati, A., Jung, K., Harman, S., Downing, L., Ng, A.** and **Shah, N. H.** (2018) "Improving palliative care with deep learning," *BMC Med. Inform. Decis. Making*, **18**(4): 122, Dec., doi: 10.1186/s12911-018-0677-8.

[15] **Alashwal, H., El Halaby, M., Crouse, J. J., Abdalla, A.** and **Moustafa, A. A.** (2019). "The application of unsupervised clustering methods to Alzheimer's disease," *Frontiers Comput. Neuroscience*, **13**: 31, May, doi: 10.3389/fncom.2019.00031.

[16] **Davis, D. A., Chawla, N. V., Christakis, N. A.** and **Barabási, A. L.** (2010), "Time to CARE: A collaborative engine for practical disease prediction," *Data Mining Knowl. Discovery*, **20** (3): 388–415, Nov., doi: 10.1007/s10618-009-0156-z.

[17] **Soni, J., Ansari, U., Sharma, D.** and **Soni, S.** (2011). "Predictive data mining for medical diagnosis: An overview of heart disease prediction," *Int. J. Comput. Appl.*, **17** (8): 43–48, Mar., doi: 10.5120/2237-2860.

[18] **McCormick, T., Rudin, C.** and **Madigan, D.** (2011) *A Hierarchical Model for Association Rule Mining of Sequential Events: An Approach to Automated Medical Symptom Prediction*. Columbia, SC: Columbia University.

[19] **Yiannakoulias, N., Schopflocher, D.** and **Svenson, L.** (2009). "Using administrative data to understand the geography of case ascertainment," *Chronic Diseases in*

Canada, **30**, (1): 20–28, Dec., doi: 10.24095/hpcdp.30.1.05.

[20] **Fisher, E. S., Malenka, D. J., Wennberg, J. E.** and **Roos, N. P.** (1990). "Technology assessment using insurance claims. Example of prostatectomy," *Int. J. Technol. Assess Health Care*, **6**(2): 194–202, Jan., doi: 10.1017/s0266462300000714.

[21] **Cruz, J. A.** and **Wishart, D. S.** (2006). "Applications of machine learning in cancer prediction and prognosis," *Cancer Inform.*, **2**: 59–77, Feb., doi: 10.1177/117693510600200030.

[22] **Li, J.** *et al.* (2018). "Support Vector Machines (SVM) classification of prostate cancer Gleason score in central gland using multiparametric magnetic resonance images: A cross-validated study," *Eur. J. Radiol.*, **98**: 61–67, Jan., doi: 10.1016/j.ejrad.2017.11.001.

[23] B**ouazza, S. H., Hamdi, N., Zeroual, A.** and **Auhmani, K.** (2015). "Gene-expression-based cancer classification through feature selection with KNN and SVM classifiers," in *2015 Intell. Syst. Comput. Vision (ISCV)*, Fez, Morocco, Mar. 25–26, 2015, pp. 1–6.

[24] **Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V.** and **Fotiadis, D. I.** (2015). "Machine learning applications in cancer prognosis and prediction," *Comput. Structural Biotechnology J.*, **13**: 8–17, Nov., doi: 10.1016/j.csbj.2014.11.005.

[25] **Magaudda, A.** *et al.* (2016). "Validation of a novel classification model of psychogenic nonepileptic seizures by video-EEG analysis and a machine learning approach," *Epilepsy Behav.*, 60: 197–201, Jul., doi: 10.1016/j.yebeh.2016.03.031.

[26] **Calamuneri, A., Donato, L., Scimone, C., Costa, A., D'Angelo, R.** and **Sidoti, A.** (2017). "On machine learning in biomedicine," *Life Saf. Secur.*, vol. 5, no. 12, pp. 96–99, Dec., doi: 10.12882/2283-7604.2017.5.12.

[27] **World Health Orgniazation (WHO)**, "*Coronavirus disease 2019 (COVID-19) situation report – 72.*" https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn=3dd8971b_2.

[28] **Karisani, N.** and **Karisani, P.** (2020). "Mining coronavirus (COVID-19) posts in social media,", *arXiv:2004.06778*.

[29] **Tran, T.** and **Lee, K.** (2016). "Understanding citizen reactions and Ebola-related information propagation on social media," in *2016 IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, San Francisco, CA, Aug. 18–21, 2016, pp. 106–111.

[30] **Li, I., Li, Y., Li, T., Alvarez-Napagao, S.** and **Garcia, D. (**2020**).** "What are we depressed about when we talk about COVID19: Mental health analysis on tweets using natural language processing, 2020", *arXiv:2004.10899*.

[31] **Gifani, P., Shalbaf, A.** and **Vafaeezadeh, M.** (2020). "Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans," *Int. J. Comput. Assisted Radiol. Surgery*, **16**(1): 115–123, Jan., doi: 10.1007/s11548-020-02286-w.

[32] **Muhammad, L., Algehyne, E. A., Usman, S. S., Ahmad, A., Chakraborty, C.** and **Mohammed, I.** (2021). "Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset," *SN Comput. Sci.*, **2**(1): 11, Nov., doi: 10.1007/s42979-020-00394-7.

[33] K**assani, S. H., Kassasni, P. H., Wesolowski, M. J., Schneider, K. A.** and **Deters, R.** (2020). "Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: A machine learning-based approach," 2020, *arXiv:2004.10641*.

[34] **Mariani, M. C., Tweneboah, O. K.** and **Bhuiyan, M. A. M.** (2019). "Supervised machine learning models applied to disease diagnosis and prognosis," *AIMS Public Health*, **6**(4): 405–423, Oct., doi: 10.3934/publichealth.2019.4.405.

[35] **Roberts, K.** *et al.* (2020). "TREC-COVID: Rationale and structure of an information retrieval shared task for COVID-19," *J. Am. Med. Inform. Assoc.*, **27**(9): 1431–1436, May, doi: 10.1093/jamia/ocaa091.

[36] **Wang, L. L.** *et al.* (2020) "*CORD-19: The Covid-19 open research dataset,*", *arXiv:2004.10706*.

[37] **Systems Science and Engineering (CSSE) at Johns Hopkins University**, *"COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University."* https://github.com/CSSEGISandData/COVID-19.

[38] **Asgari, S., Scalzo, F.** and **Kasprowicz, M.** (2019). "Pattern recognition in medical decision support," *BioMed Res. Int.*, vol. 2019, p. 6048748, Jun., doi: 10.1155/2019/6048748.

[39] **Pedregosa, F.** *et al.* (2011). "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, **12** (85): 2825–2830, Nov.

استخدام تقنيات التعلم الآلي للتنبؤ بنتائج مرضى كوفيد –19

نهى النزاوي

قسم علوم وهندسة الحاسب الآلي، كلية ينبع الجامعية، الهيئة الملكية للجبيل وينبع، مدينة ينبع الصناعية، المملكة العربية السعودية

المستخلص. أثر التفشي الأخير لمرض فيروس كورونا 2019 (COVID-19) على حياة الإنسان إلى حد كبير. وأدى الوباء إلى اضطراب اجتماعي واقتصادي عالمي حاد، مما تسبب في تأجيل أو إلغاء الأحداث الكبرى، COVID-19 هو تفشي جديد ومتسارع. وينتشر الفيروس بسرعة كبيرة وقد وصل تقريبًا إلى كل جزء من العالم مع أكثر من 100 مليون حالة تم الإبلاغ عنها، وموارد الرعاية الصحية الحرجة محدودة. لذلك، من المهم للغاية التنبؤ بالمرضى الأكثر عرضة للإصابة بهذا المرض الخطير، ومواجهة مخاطر المضاعفات، بما في ذلك الوفاة. في هذه الدراسة، قمنا بمقارنة أداء خوارزميات التعلم الآلي المختلفة في التنبؤ بنتائج مرضى COVID-19 بناءً على مجموعات لعوامل الخطر الخاصة بالمريض. وأفضل الخوارزمية أداءً هي مصنف الغابة العشوائية، والذي يحقق درجة F تبلغ 0,788 ومعدل دقة 0,789. والنموذج المقترح في هذه الدراسة قادر على التنبؤ بالنتيجة (أي ميت أو خرج أو مستقرة) لأي مريض تم تشخيصه بـ COVID-19، باستخدام نفس مجموعة عوامل الخطر، وهي: الجنس، والبلد، والأعراض، والأمراض المزمنة. ويمكن لنتائج هذه الدراسة أن تكمل المهارات السريرية، وتساعد الأطباء في التنبؤ بالأنماط غير المتوقعة لتحديد الحالات الخفيفة بين المرضى الذين تم تشخيصهم، والحالات القليلة التي ستتطور إلى مرض شديد.

الكلمات المفتاحية: مصطلحات – الذكاء الاصطناعي، نتائج COVID-19، التعلم الآلي، النموذج التنبئي.