# Covid-19 Stages Prediction Literature Survey :

# 1.

Title:  
A prediction model based on machine learning for diagnosing the early COVID-19 patients.

Authors:  
Nan-Nan Sun, Ya Yang, Ling-Ling Tang, Yi-Ning Dai, Hai-Nv Gao, Hong-Ying Pan, Bin Ju.

Abstract:  
With the dramatically fast spread of COVID-9, real-time reverse transcription polymerase chain reaction (RT-PCR) test has become the gold standard method for confirmation of COVID-19 infection. However, RT-PCR tests are complicated in operation andIt usually takes 5-6 hours or even longer to get the result. Additionally, due to the low virus loads in early COVID-19 patients, RT-PCR tests display false negative results in a number of cases. Analyzing complex medical datasets based on machine learning provides health care workers excellent opportunities for developing a simple and efficient COVID-19 diagnostic system. This paper aims at extracting risk factors from clinical data of early COVID-19 infected patients and utilizing four types of traditional machine learning approaches including logistic regression(LR), support vector machine(SVM), decision tree(DT), random forest(RF) and a deep learning-based method for diagnosis of early COVID-19. The results show that the LR predictive model presents a higher specificity rate of 0.95, an area under the receiver operating curve (AUC) of 0.971 and an improved sensitivity rate of 0.82, which makes it optimal for the screening of early COVID-19 infection. We also perform the verification for generality of the best model (LR predictive model) among Zhejiang population, and analyze the contribution of the factors to the predictive models. Our manuscript describes and highlights the ability of machine learning methods for improving the accuracy and timeliness of early COVID-19 infection diagnosis. The higher AUC of our LR-base predictive model makes it a more conducive method for assisting COVID-19 diagnosis. The optimal model has been encapsulated as a mobile application (APP) and implemented in some hospitals in Zhejiang Province.

# 2.

Title:  
Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID Symptom Study app.

Authors:  
Carole H. Sudre\*†, Karla A. Lee†, Mary Ni Lochlainn†, Thomas Varsavsky, Benjamin Murray, Mark S. Graham, Cristina Menni, Marc Modat, Ruth C. E. Bowyer, Long H. Nguyen, David A. Drew, Amit D. Joshi, Wenjie Ma, Chuan-Guo Guo, Chun-Han Lo, Sajaysurya Ganesh, Abubakar Buwe, Joan Capdevila Pujol, Julien Lavigne du Cadet, Alessia Visconti, Maxim B. Freidin, Julia S. El-Sayed Moustafa, Mario Falchi, Richard Davies, Maria F. Gomez, Tove Fall, M. Jorge Cardoso, Jonathan Wolf, Paul W. Franks, Andrew T. Chan, Tim D. Spector, Claire J. Steves†, Sébastien Ourselin.

Abstract:  
As no one symptom can predict disease severity or the need for dedicated medical support in coronavirus disease 2019 (COVID-19), we asked whether documenting symptom time series over the first few days informs outcome. Unsupervised time series clustering over symptom presentation was performed on data collected from a training dataset of completed cases enlisted early from the COVID Symptom Study Smartphone application, yielding six distinct symptom presentations. Clustering was validated on an independent replication dataset between 1 and 28 May 2020. Using the first 5 days of symptom logging, the ROC-AUC (receiver operating characteristic – area under the curve) of need for respiratory support was 78.8%, substantially outperforming personal characteristics alone (ROC-AUC 69.5%). Such an approach could be used to monitor at-risk patients and predict medical resource requirements days before they are required.

# 3.

Title:  
Symptom Prediction and Mortality Risk Calculation for COVID-19 Using Machine Learning.

Authors:  
Elham Jamshidi, †Amirhossein Asgary, †Nader Tavakoli, †Alireza Zali, Farzaneh Dastan, Amir Daaee, Mohammadtaghi Badakhshan, Hadi Esmaily, Seyed Hamid Jamaldini, Saeid Safari, Ehsan Bastanhagh, Ali Maher, Amirhesam Babajani, Maryam Mehrazi, Mohammad Ali Sendani Kashi, Masoud Jamshidi, Mohammad Hassan Sendani, Sahand Jamal Rahi, and Nahal Mansour.

Abstract:  
Background: Early prediction of symptoms and mortality risks for COVID-19 patients would improve healthcare outcomes, allow for the appropriate distribution of healthcare resources, reduce healthcare costs, aid in vaccine prioritization and self-isolation strategies, and thus reduce the prevalence of the disease. Such publicly accessible prediction models are lacking, however.  
Methods: Based on a comprehensive evaluation of existing machine learning (ML) methods, we created two models based solely on the age, gender, and medical histories of 23,749 hospital-confirmed COVID-19 patients from February to September 2020: a symptom prediction model (SPM) and a mortality prediction model (MPM). The SPM predicts 12 symptom groups for each patient: respiratory distress, consciousness disorders, chest pain, paresis or paralysis, cough, fever or chill, gastrointestinal symptoms, sore throat, headache, vertigo, loss of smell or taste, and muscular pain or fatigue. The MPM predicts the death of COVID-19-positive individuals.  
Results: The SPM yielded ROC-AUCs of 0.53–0.78 for symptoms. The most accurate prediction was for consciousness disorders at a sensitivity of 74% and a specificity of 70%. 2,440 deaths were observed in the study population. MPM had a ROC-AUC of 0.79 and could predict mortality with a sensitivity of 75% and a specificity of 70%. About 90% of deaths occurred in the top 21 percentile of risk groups. To allow patients and clinicians to use these models easily, we created a freely accessible online interface at www.aicovid.net.  
Conclusion: The ML models predict COVID-19-related symptoms and mortality using information that is readily available to patients as well as clinicians. Thus, both can rapidly estimate the severity of the disease, allowing shared and better healthcare decisions with regard to hospitalization, self-isolation strategy, and COVID-19 vaccine prioritization in the coming months.

# 4.

Title:  
Covid Symptom Severity Using Decision Tree.

Authors:  
Naim Rochmawati, Hanik Badriyah Hidayati, Yuni Yamasari, Wiyli Yustanti, Lusia Rakhmawati, Hapsari P. A. Tjahyaningtijas, Yeni Anistyasari.

Abstract:  
— Corona is a very contagious virus. In a pandemic like this, people often worry whether they are infected or not. When they cough, they often worry whether it is a sign of covid19 or an ordinary cough. From the clinical symptoms can actually be known whether someone has Covid or not. In this study, a clinical symptom dataset will be used to classify the symptoms using a Decision Tree algorithm. The decision trees used in this research are J48 and Hoeffding Tree. Decision Tree is one of the most popular classification methods because it is easy to interpret by Humans. the prediction model uses a hierarchical structure. The concept is to convert data into decision trees or decision rules. the result of J48 were slightly better than the Hoeffding tree in terms of accuracy, precision, and recall. Meanwhile, from the tree view results, the Hoeffding Tree is simpler and the number of nodes is less than J48.

# 5.

Title:  
COVID-19 Early Symptom Prediction Using Blockchain and Machine Learning.

Authors:  
Sarada Kiranmayee Tadepalli, Ruppa K. Thulasiram.

Abstract:  
The COVID-19 outbreak has resulted in unprecedented and difficult times for world’s population. Social distancing and self-isolation have become very important to reduce the spread. This called upon the creation of numerous applications that have used proprietary models for symptoms-tracking and contact-tracing around the world to mitigate the spread. In most of the applications data collected is stored in a centralized database without verification and hence, the data is not reliable. In this study, a decentralized application for COVID-19 symptoms tracking using Blockchain is proposed in order to enhance reliable data collection for training Machine Learning (ML) models. The Blockchain integration in this application will help in collecting COVID-19 symptoms data from the patients with trust. In addition to this, the data would be first verified by an entity of the decentralized network (e.g. a COVID-19 testing lab). Then, with the consent of the patient, this data is provided to the centralized system for retraining the ML. In short, the main advantage of this architecture is that the data from the users is collected and checked by a laboratory first and then provided to the ML model. The process helps in identifying the incorrect ML prediction and further train the ML model with reliable data for accurate prediction. Moreover, the trust of the users is earned as the data transfer happens with their consent and, besides, all transactions are recorded on the Blockchain, which is possible with the help of the Distributed Ledger Technology (DLT).

# 6.

Title:  
Risk factors prediction, clinical outcomes, and mortality in COVID-19 patients.

Authors:  
Roohallah Alizadehsani, Zahra Alizadeh Sani,Mohaddeseh Behjati, Zahra Roshanzamir, Sadiq Hussain, Niloofar Abedini, Fereshteh Hasanzadeh, Abbas Khosravi, Afshin Shoeibi, Mohamad Roshanzamir, Pardis Moradnejad, Saeid Nahavandi, Fahime Khozeimeh, Assef Zare, Maryam Panahiazar, U. Rajendra Acharya, Sheikh Mohammed Shariful Islam.

Abstract:  
Preventing communicable diseases requires understanding the spread, epidemiology, clinical features, progression, and prognosis of the disease. Early identification of risk factors and clinical outcomes might help in identifying critically ill patients, providing appropriate treatment, and preventing mortality. We conducted a prospective study in patients with flu-like symptoms referred to the imaging department of a tertiary hospital in Iran between March 3, 2020, and April 8, 2020. Patients with COVID-19 were followed up after two months to check their health condition. The categorical data between groups were analyzed by Fisher's exact test and continuous data by Wilcoxon rank-sum test. Three hundred and nineteen patients (mean age 45.48 ± 18.50 years, 177 women) were enrolled. Fever, dyspnea, weakness, shivering, C-reactive protein, fatigue, dry cough, anorexia, anosmia, ageusia, dizziness, sweating, and age were the most important symptoms of COVID-19 infection. Traveling in the past 3 months, asthma, taking corticosteroids, liver disease, rheumatological disease, cough with sputum, eczema, conjunctivitis, tobacco use, and chest pain did not show any relationship with COVID-19. To the best of our knowledge, a number of factors associated with mortality due to COVID-19 have been investigated for the first time in this study. Our results might be helpful in early prediction and risk reduction of mortality in patients infected with COVID-19.

# 7.

Title:  
Machine learning-based prediction of COVID-19 diagnosis based on symptoms

Authors:  
Yazeed Zoabi, Shira Deri-Rozov, Noam Shomron.

Abstract:  
Effective screening of SARS-CoV-2 enables quick and efficient diagnosis of COVID-19 and can mitigate the burden on healthcare systems. Prediction models that combine several features to estimate the risk of infection have been developed. These aim to assist medical staff worldwide in triaging patients, especially in the context of limited healthcare resources. We established a machine-learning approach that trained on records from 51,831 tested individuals (of whom 4769 were confirmed to have COVID-19). The test set contained data from the subsequent week (47,401 tested individuals of whom 3624 were confirmed to have COVID-19). Our model predicted COVID-19 test results with high accuracy using only eight binary features: sex, age ≥60 years, known contact with an infected individual, and the appearance of five initial clinical symptoms. Overall, based on the nationwide data publicly reported by the Israeli Ministry of Health, we developed a model that detects COVID-19 cases by simple features accessed by asking basic questions. Our framework can be used, among other considerations, to prioritize testing for COVID-19 when testing resources are limited.

# 8.

Title:  
Using symptom-based case predictions to identify host genetic factors that contribute to COVID-19 susceptibility

Authors:  
Irene V. van Blokland, Pauline Lanting, Anil P. S. Ori, Judith M. Vonk, Robert C. A. Warmerdam, Johanna C. Herkert, Floranne Boulogne, Annique Claringbould, Esteban A. Lopera-Maya, Meike Bartels, JoukeJan Hottenga, Andrea Ganna, Juha Karjalainen, Caroline Hayward, Chloe Fawns-Ritchie, Archie Campbell, David Porteous, Elizabeth T. Cirulli, Kelly M. Schiabor Barrett, Stephen Riffle, Alexandre Bolze, Simon White, Francisco Tanudjaja, Xueqing Wang, Jimmy M. Ramirez, Yan Wei Lim, James T. Lu, Nicole L. Washington, Eco J. C. de Geus, Patrick Deelen, H. Marike Boezen, Lude H. Franke.

Abstract:  
Epidemiological and genetic studies on COVID-19 are currently hindered by inconsistent and limited testing policies to confirm SARS-CoV-2 infection. Recently, it was shown that it is possible to predict COVID-19 cases using cross-sectional self-reported disease-related symptoms. Here, we demonstrate that this COVID-19 prediction model has reasonable and consistent performance across multiple independent cohorts and that our attempt to improve upon this model did not result in improved predictions. Using the existing COVID19 prediction model, we then conducted a GWAS on the predicted phenotype using a total of 1,865 predicted cases and 29,174 controls. While we did not find any common, large effect variants that reached genome-wide significance, we do observe suggestive genetic associations at two SNPs (rs11844522, p = 1.9x10-7; rs5798227, p = 2.2x10-7). Explorative analyses furthermore suggest that genetic variants associated with other viral infectious diseases do not overlap with COVID-19 susceptibility and that severity of COVID-19 may have a different genetic architecture compared to COVID-19 susceptibility. This study represents a first effort that uses a symptom-based predicted phenotype as a proxy for COVID-19 in our pursuit of understanding the genetic susceptibility of the disease. We conclude that the inclusion of symptom-based predicted cases could be a useful strategy in a scenario of limited testing, either during the current COVID-19 pandemic or any future viral outbreak.

# 9.

Title:  
Potential Factors for Prediction of Disease Severity of COVID-19 Patients

Authors:  
Huizheng Zhang, Xiaoying Wang, Zongqiang Fu, Ming Luo, Zhen Zhang, Ke Zhang, Ying He, Dongyong Wan, Liwen Zhang, Jing Wang, Xiaofeng Yan, Mei Han, Yaokai Chen.

Abstract:  
Objective Coronavirus disease 2019 (COVID-19) is an escalating global epidemic caused by SARS-CoV-2, with a high mortality in critical patients. Effective indicators for predicting disease severity in SARS-CoV-2 infected patients are urgently needed.  
Methods: In this study, 43 COVID-19 patients admitted in Chongqing Public Health Medical Center were involved. Demographic data, clinical features, and laboratory examinations were obtained through electronic medical records. Peripheral blood specimens were collected from COVID-19 patients and examined for lymphocyte subsets and cytokine profiles by flow cytometry. Potential contributing factors for prediction of disease severity were further analyzed.  
Results: A total of 43 COVID-19 patients were included in this study, including 29 mild patients and 14 sever patients. Severe patients were significantly older (61.9±9.4 vs 44.4±15.9) and had higher incidence in co-infection with bacteria compared to mild group (85.7%vs27.6%). Significantly more severe patients had the clinical symptoms of anhelation (78.6%) and asthma (71.4%). For laboratory examination, 57.1% severe cases showed significant reduction in lymphocyte count. The levels of Interluekin-6 (IL6), IL10, erythrocyte sedimentation rate (ESR) and D-Dimer (D-D) were significantly higher in severe patients than mild patients, while the level of albumin (ALB) was remarkably lower in severe patients. Further analysis demonstrated that ESR, D-D, age, ALB and IL6 were the major contributing factors for distinguishing severe patients from mild patients. Moreover, ESR was identified as the most powerful factor to predict disease progression of COVID-19 patients.  
Conclusion: Age and the levels of ESR, D-D, ALB and IL6 are closely related to the disease severity of COVID-19 patients. ESR can be used as a valuable indicator for distinguishing severe COVID-19 patients in early stage, so as to increase the survival of severe patients.

# 10.

Title:  
A Novel Scoring System for Prediction of Disease Severity in COVID-19

Authors:  
Chi Zhang†, Ling Qin†, Kang Li†, Qi Wang, Yan Zhao, Bin Xu, Lianchun Liang, Yanchao Dai, Yingmei Feng, Jianping Sun, Xuemei Li, Zhongjie Hu, Haiping Xiang, Tao Dong, Ronghua Jin, Yonghong Zhang.

Abstract:  
Background: A novel enveloped RNA beta coronavirus, Corona Virus Disease 2019 (COVID-19) caused severe and even fetal pneumonia in China and other countries from December 2019. Early detection of severe patients with COVID-19 is of great significance to shorten the disease course and reduce mortality. Methods: We assembled a retrospective cohort of 80 patients (including 56 mild and 24 severe) with COVID-19 infection treated at Beijing You’an Hospital. We used univariable and multivariable logistic regression analyses to select the risk factors of severe and even fetal pneumonia and build scoring system for prediction, which was validated later on in a group of 22 COVID-19 patients.  
Results: Age, white blood cell count, neutrophil, glomerular filtration rate, and myoglobin were selected by multivariate analysis as candidates of scoring system for prediction of disease severity in COVID-19. The scoring system was applied to calculate the predictive value and found that the percentage of ICU admission (20%, 6/30) and ventilation (16.7%, 5/30) in patients with high risk was much higher than those (2%, 1/50; 2%, 1/50) in patients with low risk (p = 0.009; p = 0.026). The AUC of scoring system was 0.906, sensitivity of prediction is 70.8%, and the specificity is 89.3%. According to scoring system, the probability of patients in high risk group developing severe disease was 20.24 times than that in low risk group.  
Conclusions: The possibility of severity in COVID-19 infection predicted by scoring system could help patients to receiving different therapy strategies at a very early stage.