COVID-19 Severity Prediction in SARS-CoV-2 RNA-Positive Patients by Different Ensemble Learning Strategies

Emek GÜLDOĞAN[1], Mehmet KIVRAK[2], and Cemil Colak[2]

[1]Inonu University
[2]Inonu Universitesi Tip fakultesi

March 07, 2024

Abstract

Objective: While the coronavirus persists marginally for ninety-five percent of the infected case count, the remaining five percent have been placed in a critical or vital condition. This study investigates to design an intelligent model that predicts the disease severity level by modeling the relationships between the severity of COVID-19 infection and the various demographic/clinical characteristics of individuals. Material and Methods: A public dataset of a cross-sectional study included the demographic and symptomatological characteristics of 223 COVID-19 patients. The dataset was randomly divided into training (75%) and testing (25%) datasets. During training, the class imbalance problem was solved, and the related factors with the COVID-19 severity were selected using the evolutionary method supported by a genetic algorithm. Neural Network (NN), Support Vector Machine (SVM), QUEST algorithms together with confidence weighted voting, voting, and highest confidence wins strategies (HCWS) were constructed, and the predictive power of models was evaluated by performance metrics. Results: Of the individual models, the NN model outperformed SVM and QUEST algorithms based on the performance metrics in the training and testing datasets. However, ensemble approaches gave better predictions as compared to the individual models regarding all the evaluation metrics. Conclusions: The proposed voting ensemble model outperforms other ensemble and individual machine learning approaches for the severity prediction of COVID-19 disease. The proposed ensemble learning model can be integrated into web or mobile applications in classifying the severity of COVID-19 for clinical decision support.

COVID-19 Severity Prediction in SARS-CoV-2 RNA-Positive Patients by Different Ensemble Learning Strategies

Abstract

Objective : While the coronavirus persists marginally for ninety-five percent of the infected case count, the remaining five percent have been placed in a critical or vital condition. This study investigates to design an intelligent model that predicts the disease severity level by modeling the relationships between the severity of COVID-19 infection and the various demographic/clinical characteristics of individuals.

Material and Methods: A public dataset of a cross-sectional study included the demographic and symptomatological characteristics of 223 COVID-19 patients. The dataset was randomly divided into training (75%) and testing (25%) datasets. During training, the class imbalance problem was solved, and the related factors with the COVID-19 severity were selected using the evolutionary method supported by a genetic algorithm. Neural Network (NN), Support Vector Machine (SVM), QUEST algorithms together with confidence weighted voting, voting, and highest confidence wins strategies (HCWS) were constructed, and the predictive power of models was evaluated by performance metrics.

Results : Of the individual models, the NN model outperformed SVM and QUEST algorithms based on the performance metrics in the training and testing datasets. However, ensemble approaches gave better

predictions as compared to the individual models regarding all the evaluation metrics.

Conclusions: The proposed voting ensemble model outperforms other ensemble and individual machine learning approaches for the severity prediction of COVID-19 disease. The proposed ensemble learning model can be integrated into web or mobile applications in classifying the severity of COVID-19 for clinical decision support.

Keywords: Classification, COVID-19 severity, ensemble learning, machine learning, prediction.

WHAT'S KNOWN? (what is already known about this subject?)

The severity of the COVID-19 pandemic is associated with the survival of individuals, and evaluating the level of severity, especially in individuals with chronic diseases, improves the quality of life and can reduce the mortality rates caused by this virus. The severity of the COVID-19 disease was divided into four stages, i.e., mild, moderate, extreme, and serious, in accordance with the Guidelines provided by the Ministry of Health, Labor and Welfare, Japan, on the Diagnosis and Treatment of Novel Coronavirus. There have been limited studies investigating between the severity of COVID-19 infection and the various demographic/clinical characteristics of individuals using an intelligent ensemble model. Thence, this study investigates to design an intelligent model that predicts the disease severity level by modeling the relationships between the severity of COVID-19 infection and the various demographic/clinical characteristics of individuals.

WHAT'S NEW? (what does this study contribute to the literature?)

Understanding clinical/demographic features, progression, and prognosis of the COVID-19 disease may help recognize critically ill patients, provide appropriate care, and avoid mortality. In light of this important data, this research aims to design an intelligent model that predicts the disease severity level by modeling the relationships between the severity of COVID-19 infection and the various demographic/clinical characteristics of individuals. The possible contributions of the current study are given below:

- This study investigates to design an intelligent model that predicts the disease severity level by modeling the relationships between the severity of COVID-19 infection and the various demographic/clinical characteristics of individuals.
- Neural Network (NN), Support Vector Machine (SVM), QUEST algorithms together with confidence weighted voting, voting, and highest confidence wins strategies (HCWS) were constructed.
- The proposed voting ensemble model outperforms other ensemble and individual machine learning approaches for the severity prediction of COVID-19 disease.
- The proposed ensemble learning model can be integrated into web or mobile applications in classifying the severity of COVID-19 for clinical decision support.

1. Introduction

Since December 2019, a cohort of patients in Wuhan, Hubei Province, China, has suffered from unknown etiology acute respiratory disease. The first cases showed a connection to the Huanan wholesale market for seafood. The Chinese Centre for Disease Control and Prevention (CDC) found a new coronavirus, previously referred to by the International Committee on Taxonomy of Viruses (ICTV) as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, formerly referred to as 2019-nCoV), by analyzing the patient's throat swab sample. Human angiotensin-converting enzyme 2 (ACE2) molecules strongly interact with the SARS-CoV-2 Spike protein's receptor-binding domain (RBD). This gives human respiratory epithelial cells a high infectious proficiency of the virus. The virus induces exhaustion, fever, cough, and serious or moderate breathing impediments. SARS-CoV-2 is referred to as Coronavirus Disease 2019 (COVID-19), which showed moderate symptoms in the majority of patients (1). Early disease detection helps to provide necessary, appropriate care for clinicians. The clinical and epidemiological features of coronavirus were examined by Zheng et al. (2). Treatment, radiological, laboratory, clinical, demographic, and epidemiological data were used for 99 confirmed patients in China with COVID-19. As typical symptoms, they reported fever, tiredness, and dry cough. The median age of patients registered is 49 years, 41 percent had the underlying disorder, 49 percent had near interaction with patients affected by COVID19, and 42 percent had resided or traveled

to Wuhan. Lower counts of CD8 and CD4, lower white blood cells, lymphocytes, and neutrophils; higher levels of natriuretic peptides in the brain; higher myocardial damage levels, and higher C-reactive proteins can be used for early detection of disease in seriously ill patients. (1).

The severity of the COVID-19 pandemic is associated with the survival of individuals, and evaluating the level of severity, especially in individuals with chronic diseases, improves the quality of life and can reduce the mortality rates caused by this virus. The severity of the COVID-19 disease was divided into four stages, i.e., mild, moderate, extreme, and serious, in accordance with the Guidelines provided by the Ministry of Health, Labor and Welfare, Japan, on the Diagnosis and Treatment of Novel Coronavirus. In short, the mild disease was characterized as a lack of respiratory symptoms, no radiological pulmonary manifestations, and levels of oxygen saturation (SpO2) of about 96 percent. The moderate disease was defined as mild respiratory symptoms, 93 percent $<$ SpO2 $<$ 96 percent, and radiological evidence of pneumonia. Severe cases were described as oxygen support requiring SpO2 93 percent. Critical was described as requiring support for acute respiratory distress syndrome (ARDS) by the heart-lung machine or extracorporeal membrane oxygenation (ECMO) (3).

Medical signs, personal traits, and demographic characteristics were strongly associated with COVID-19 infection in various research studies. Other clinical characteristics, specifically obesity, cardiovascular disease, and hypertension, have also been reported in studies as important factors influencing the rate of COVID19 infection. On the other hand, studies discuss the demographic characteristics strongly associated with COVID-19 disease worldwide, such as a country's gross domestic product ratio, smoking prevalence, a country's average annual temperature, etc. (4).

Understanding clinical/demographic features, progression, and prognosis of the COVID-19 disease may help recognize critically ill patients, provide appropriate care, and avoid mortality (5). In light of this important data, this research aims to design an intelligent model that predicts the disease severity level by modeling the relationships between the severity of COVID-19 infection and the various demographic/clinical characteristics of individuals.

2. Materials and Methods

2.1. Patient data and selection

The open-accessed dataset used in the current research from a cross-sectional study (6) includes the demographic and symptomatological characteristics of 223 COVID-19 patients treated at the Sisli Hamidiye Etfal Training and Research Hospital of the University of Health Sciences between March 10 and April 21, 2020 (7). The public dataset can be achieved from the web address of*http://dx.doi.org/10.17632/w9tjhj8jy6.1.* From the beginning to the diagnosis of general and otolaryngologic symptoms of COVID-19, prevalence, seriousness, length, and time were determined. The SARS-CoV-2 RNA was detected in all patients included in this dataset in the swab specimens. The dataset encapsulates general patient characteristics (age, gender, and comorbidity), hospitalization status, dates, and swab specimen collection results, computerized tomography results, and drugs used for COVID-19. Five otorhinolaryngologists and two infectious diseases specialists performed the surveys. All the patients involved in the study gave informed consent. Detailed records of the patients (113 males, 110 females) who had SARS-CoV-2 RNA found in their nasopharyngeal or oropharyngeal swab specimens by reverse transcription-polymerase chain reaction (RT-PCR) were examined in the current study. During the study period, all patients suspected of having COVID-19 were treated in compliance with the Turkish Ministry of Health Study Board's interim guidelines. After a brief training following real interim instructions, resident or attending doctors obtained nasopharyngeal and oropharyngeal swabs. Specimens were tested in the General Directorate of Public Health Microbiology Reference Laboratory for the presence of SARS-CoV-2 RNA via Nucleic Acid Amplification Tests (RT-PCR and nucleic acid sequence analysis if necessary). Detailed information on the research design and protocol can be obtained from the related study (6). The attributes considered in the present work are summarized in Table 1.

2.2. Data Preprocessing

There were no missing values in the available dataset regarding COVID-19 severity and other attributes of the patients. However, there was a problem of class imbalance between categories of COVID-19 severity. The class imbalance problem affects data when class distributions are strongly imbalanced. Many classification learning algorithms in this context have poor predictive accuracy for the rare class. The proportion of cases in a data set that belongs to each class plays an important role in machine learning. The real-world data, however, often suffer from class imbalances. It is harder to deal with multi-class tasks that require different class misclassification costs than with two-class tasks (8). The Sample (Balance) operator in the RapidMiner software was used to solve the class imbalance problem that emerged in this study. An automated balancing of an example set with labels is achieved by the Sample (Balance) operator, which allows up and downsampling (9). For feature/attribute selection, the most important attributes of the given data set were selected by the Optimize Selection (Evolutionary operator in the RapidMiner software, which uses a Genetic Algorithm. A genetic algorithm (GA) is a search heuristic that imitates the natural evolution process. This heuristic is regularly used to construct helpful solutions to problems of optimization. Genetic algorithms belong to the broader class of evolutionary algorithms (EA), which use techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover, to produce solutions to optimization problems (10).

2.3. Data Mining Approaches

2.3.1. Neural Network (NN) Algorithm

Neural networks predict a continuous or categorical goal based on one or more predictors by identifying trends in the data which are unknown and probably complex. A feed-forward, supervised learning network of up to two hidden layers is the multilayer perceptron (MLP). The MLP network is a feature of one or more predictors that minimize one or more target prediction errors. A combination of categorical and continuous fields can be predictors and targets (11). In the current study, the activation function was a hyperbolic tangent, the error function was cross-entropy, the number of hidden layers was two, and the number of component models for boosting was ten. The boosting is an algorithm used to enhance model stability/precision, can be used in all models, and can minimize prediction variances and biases. Scaled Conjugate Gradient technique was used for tuning model hyperparameters (12).

2.3.2. Support Vector Machine (SVM) Algorithm

A supervised learning technique that generates input-output mapping functions from a collection of labeled training data is the Support Vector Machine (SVM). The mapping function may be a function of classification or a function of regression. Nonlinear kernel functions are also used for classification to translate input data to a high-dimensional feature space where the input data becomes more separable than the original input space. Hyperplanes of maximum-margin are then formed. Only a subset of the training data near the class boundaries depends on the generated model. In the present study, kernel type was radial basis polynomial, regularization parameter (C) was ten, gamma was 0.1, and stopping criteria was 1.0E (-3). The hyperparameters of the SVM algorithm was tune using a modified sequential minimal optimization (SMO) method (11, 13).

2.3.3. QUEST Algorithm

QUEST stands for a quick, unbiased, efficient statistical tree and is a relatively new algorithm for binary tree increasing. This deals with split field selection and different split-point selection. In Search, the univariate split conducts roughly unbiased field selection. If all the predictor fields are equally informative concerning the target field, QUEST selects any of the predictor fields with equal likelihood. QUEST offers many of Classification and Regression Trees' (C&RT) benefits. Still, the trees can become unusable, like C&RT. Automatic cost-complexity pruning can be applied to the QUEST tree to reduce its scale. Concerning this study, the boosting method was used to build the QUEST model to enhance accuracy (11). The hyperparameters of the SVM algorithm were maximum tree depth of 5, maximum surrogates of 5, minimum records in parent branch of 2%, minimum records in child branch of 1%, and the number of component models for boosting was ten. The model hyperparameters were optimized by SMO approximation (13).

2.4. Ensemble Learning

2.4.1. Confidence weighted voting (CWV) strategy

More specific predictions can be given by assembling scores from individual models. Limitations in individual models can be avoided by integrating scores from many models, resulting in higher overall accuracy. Models blended in this way usually perform at least as well as the best and sometimes better individual models. Another weighted majority voting form is Confidence Weighted Voting (CWV). However, rather than giving high weights to the nearest sensors, CWV gives higher weights to those sensors that are more likely to be right (i.e., with higher confidence of correctness). A distributed comparison of sensing results and neighbors that share overlapping coverage can be made for the confidence value of each sensor. For this study, a confidence-weighted voting strategy was used to combine the scores from individual NN, SVM, and QUEST algorithms (11, 14).

$$\sum_{m=1}^{M} p_{m,j} d_{m,j} = \max_{k=1}^{K} p_{m,j} \left(\sum_{m=1}^{M} d_{m,k} \right)$$

$p_{m,j}$ is the posterior likelihood for a given vector of predictor values calculated for the kth target group by the mth base model. M is the number of base models, and K is the number of target categories (15).

2.4.2. Voting strategy (VS)

Assume that for a given vector of predictor values, the mark performance of the mth base model represents. If the kth goal group is the mark given by the mth base model and 0, otherwise. Complete M base models and K target categories are given. If the plurality of base models assigns it, the majority vote method selects the jth group. The following equation fulfills it:

$$\sum_{m=1}^{M} d_{m,j} = \max_{k=1}^{K} \left(\sum_{m=1}^{M} d_{m,k} \right)$$

$d_{m,k}$ is the label performance of the mth base model for a given predictor values vector. M is the number of base models, and K is the number of target categories (15).

2.4.3. Highest confidence wins strategy (HCWS)

The highest confidence wins strategy is calculated as follows:

$$\max_{m=1}^{M} (p_{m,j}) = \max_{k=1}^{K} \left(\max_{m=1}^{M} (p_{m,j}) \right)$$

$p_{m,j}$ is the posterior likelihood for a given vector of predictor values calculated for the kth target group by the mth base model. M is the number of base models, and K is the number of target categories (16).

2.5. Model Validation and Performance Metrics

A nested operator used for validating the models in the current study is the split validation operator. The operator consists of two subprocesses: a training subsection and a test subsection. The training subprocess is used to learn or construct a model. The learned model is then employed in the subprocess testing. The model's efficiency is also evaluated during the test process. The input dataset is divided into two sub-sets. One subset is used as a training set, the other as a test set. The model is learned from the training set and adapted to the test set. The learning process normally optimizes the model parameters so that the model fits as well as possible into the training data. If we then take an independent sample of test data, the model normally does not match the test data and the training data. Split validation is a means of predicting how a model fits into a hypothesis set if an explicit test set is not available. The Split Validation operator also

provides instruction for one set of data and testing for another explicit test set (17). The whole dataset was portioned into two subsets: %75 for training and 25% for testing samples. All the models were trained on the first dataset and tested on the second dataset. Afterward, the performance metrics were calculated for each model by DTROC software (18). The evaluation metrics were accuracy, precision, sensitivity, specificity, F1-score, Matthews correlation coefficient (MCC), and G-mean. A detailed explanation of the metrics is available from the study (19).

2.6. Data Analysis

Quantitative data were summarized as mean (standard deviation), and qualitative data were given as numbers (percentages). Pearson's Chi-square test was utilized to examine the differences among the categories of categorical variables. When significant differences were detected, the Bonferroni corrected Pearson's Chi-square test was also used for pairwise comparisons. A p-value of 0.05 was considered statistically significant. IBM SPSS Statistics 26.0, IBM SPSS Modeler 18.0, and RapidMiner Studio Educational Edition 9.8 softwares were used for the analyses (17, 20, 21).

3. Results

3.1. Baseline characteristics of the patients

Baseline characteristics of the patients by disease severity are summarized in Table 2. Of the data set consisting of 223 patients, 113 were male (50.7%), and 110 were female (49.3%). There were 41 patients (18.4%) with mild disease, 137 (61.4%) with moderate, 32 (14.3%) with severe, and 13 (5.8%) with critical. The average age of all patients was 50.09 $\pm$ 16.46 years; the mild ones were 36.17 $\pm$ 14.28 years, the moderate ones were 50.28 $\pm$ 14.88 years, the severe ones were 62.75 $\pm$ 13.63 years, and the critical ones were 60.85 $\pm$ 11.19 years, respectively.

Approximately 98.7% of the patients used Hydroxychloroquine, 63.7% Oseltamivir, 19.7% Azithromycin, 0.9% Lopinavir/Ritonavir, and 13.5% Favipiravir. About 50.7% of the patients had a fever, 54.3% had a cough, 71.3% had fatigue, 37.7% had dyspnea, and 50.7% had Myalgia/Arthralgia. The proportion of patients with headache is 26.5, the proportion of those with frontal type headache is 10.8, the proportion of those with nasal obstruction is 16.6, the proportion of those with Rhiorrhea is 11.7. Sore throat in approximately 26.0% of patients, dry throat in 16.1%, loss of smell in 31.8%, loss of taste in 34.5%, ear pain in 2.7%, dizziness in 2.2%, and hearing loss in 0.9% were the common complaints of the patients.

There are significant differences between the Patient Clinical Status variable in terms of Comorbidity Status, General Medication, Lopinavir/Ritonavir, Favipiravir, Fever Presence, Cough Presence, Cough Presence, Fatigue Presence, Dyspnea Presence, Smell Loss Presence, Dizziness Presence, and Hearing Loss Presence variables (Pearson's Chi-Squared Test; $p<0.05$).

3.2. Modelling and performance evaluation

Table 3 presents the performance metrics of the individual and ensemble models for each COVID-19 severity category. Of the individual models, the NN model outperformed SVM and QUEST algorithms based on the performance metrics in the training and testing datasets. However, ensemble approaches (i.e., HCWS, VS, CWVS) gave better predictions as compared to the individual models regarding all the evaluation metrics. As the ensemble models' estimates were compared, VS achieved slightly better prediction performance than the HCWS and CWVS algorithms.

Predictor importance values for each separate model are summarized in Table 4. Based on the estimates of the best-performing individual model (i.e., NN), the three most important predictors were age, Favipiravir use, and the presence of dyspnea, respectively. The predictor significance values of other individual models are also shown in Table 4.

4. Discussion

The COVID-19 pandemic posed a big threat to global health, as well as a massive burden on healthcare

systems. Besides, the COVID-19 pandemic has impacted the lives of many people worldwide in recent days and needs a massive number of screening tests to identify the presence of the coronavirus. At the same time, the rise of concepts of deep learning (DL) helps to build a COVID-19 diagnosis model effectively to achieve maximum detection rate with minimum computation time (22). A precise forecast of the magnitude of COVID-19 could provide realistic insights into directing critical hospitalization and treatment decisions to relieve the burden on the healthcare system. Clinical/demographic knowledge of COVID-19 disease progression and prognostics will help diagnose critically ill patients, provide adequate treatment, and prevent mortality (23). Based on these important data, the purpose of the research is to build an intelligent model predicting the severity of the disease by modeling the associations between the severity of COVID-19 infection and the demographic/clinical properties of individuals.

Predictive models acquire knowledge about a software project and predict whether the instances added in the future will be faulty or not by studying historical software information. Nevertheless, in most programs, there are many more non-defective (i.e., the majority class) cases than faulty (i.e., the minority class), which is referred to as the problem of class imbalance. Conversely, traditional algorithms in the field of machine learning presume that the numbers of minority and majority groups are essentially the same. Predictive models built from such highly imbalanced datasets tend to disregard faulty instances and predict non-default performance. As a consequence, the models yield highly skewed results and are not technically applicable. Data resampling techniques are commonly used to resolve the class imbalance issue. Two forms of general resampling strategies are oversampling and subsampling: the former creating new cases and introducing them to the minority class and the latter eliminating existing cases from the mainstream. Both techniques strive to balance the distribution of data sets to enhance prediction models' efficiency (24). In the current study, the class imbalance problem arose in the data set in the process of forming individual models. In order to solve this problem, the imbalance among the disease categories was resolved by balancing the classes in the preprocessing stage of the data set. Individual models of NN, SVM, and QUEST algorithms were constructed to predict the COVID-19 severity categories on the balanced medical records of the patients. Based on the experimental results of each singular model, the NN model produced better predictions as compared to the SVM and QUEST algorithms. Additionally, the most important factors estimated from the NN algorithm in the classification of COVID-19 severity were age, Favipiravir use, the presences of dyspnea, cough and smell loss, Lopinavir/Ritonavir use, the presences of fatigue, fever, frontal type headache, and gender, respectively.

Ensemble learning approaches use many machine-learning algorithms to generate poor predictive results based on features derived from a wide range of data forecasts and merge results to achieve higher performance than any single constituent algorithm, with different voting or other mechanisms. Therefore, ensemble learning is extremely expandable, combined with various machine learning models for different tasks such as general classification tasks, clustering tasks, etc. In general, current methods of ensemble learning can be divided into four categories: supervised classification of the ensemble, semi-supervised classification of the ensemble, the clustering ensemble, and a semi-supervised clustering ensemble (25). In the current study, different ensemble learning algorithms were implemented to combine individual predictions to classify the severity of the COVID-19 pandemic. Voting strategy, one of the ensemble learning methods, gave slightly better predictive results than other ensemble techniques (i.e., HCWS and CWVS) in predicting the severity of Covid-19 disease.

Recent studies on COVID-19 severity prediction have been reported on the applications of machine learning algorithms and artificial intelligence models. A novel study aims to develop a COVID-19 severity prediction model and explain dynamic changes in key clinical characteristics over seven weeks. In accordance with this purpose, a support vector machine model was constructed with a genetic algorithm for feature selection and achieved an accuracy of over 94% for COVID-19 severity prediction. The authors report that the proposed model includes 11 routine clinical features commonly available during COVID-19 management, which may predict the severity and guide the treatment of COVID-19 patients (26). In another recently published study, RNA-Seq and high-resolution mass spectrometry on 128 blood samples from COVID-19 positive and negative patients with diverse disease severities were performed on 219 molecular features with high significance to COVID-19 status and severity. The researchers present an interactive web-based tool (covid-omics.app)

to illustrate its utility by comparing the data published and a machine learning approach to COVID-19 severity prediction (27). Another research assesses the predictive accuracy of the severity classification of WHO COVID-19 and compares its predictive power based on the Bayesian network analysis with the new prediction model, COVID-19 EPI-SCORE. The selected variables from the machine learning model were the classification of WHO severity, acute kidney injury, age, Lactate dehydrogenase levels (LDH), lymphocytes, and activated prothrombin time (aPTT). The findings of the study demonstrate that the severity classification of the WHO is accurate for predicting serious results in patients with COVID-19 (28). Other newly published work performs a comparative analysis using machine learning algorithms [i.e., the support vector machine (SVM), decision tree (DT), k-nearest neighbor (kNN), and convolution neural network (CNN)] to classify the COVID-19 confirmed patients' pneumonia level (mild, progressive, and severe stage). Extensive experiments have been performed, and the findings show the accuracy values for kNN, SVM, DT, and CNN of 91.304%, 91.4%, 87.5%, and 95.622%, respectively (29). Some factors in this study are consistent with other reported researches. Besides, the calculated performance metrics in the current study are higher as compared to similar works (29). The results of the current and above-mentioned research studies demonstrate that machine learning and statistical learning models can predict the severity of the COVID-19 pandemic.

In conclusion, the proposed voting ensemble model outperforms other ensemble and individual machine learning approaches for the severity prediction of COVID-19 disease. The proposed ensemble learning model can be integrated into web or mobile applications for classifying the severity of COVID-19 for clinical decision support.

Ethical approval, conflict of interest and funding

This study does not require ethical approval and informed consent because the open-source data set is used. Also, there is no conflict of interest among the authors. Any institution or organization did not financially support this research.

References

1. Alizadehsani R, Behjati M, Roshanzamir Z, Hussain S, Abedini N, Hasanzadeh F, et al. Risk Factors Prediction, Clinical Outcomes, and Mortality of COVID-19 Patients. medRxiv. 2020.

2. Zheng Y-Y, Ma Y-T, Zhang J-Y, Xie X. COVID-19 and the cardiovascular system. Nature Reviews Cardiology. 2020;17(5):259-60.

3. Sugiyama M, Kinoshita N, Ide S, Nomoto H, Nakamoto T, Saito S, et al. Serum CCL17 level becomes a predictive marker to distinguish between mild/moderate and severe/critical disease in patients with COVID-19. Gene. 2020;766:145145.

4. Khan W, Hussain A, Khan SA, Al-Jumailey M, Nawaz RJapa. Association Learning Between the COVID-19 Infections and Global Demographic Characteristics Using the Class Rule Mining and Pattern Matching. 2020.

5. Alizadehsani R, Behjati M, Roshanzamir Z, Hussain S, Abedini N, Hasanzadeh F, et al. Risk Factors Prediction, Clinical Outcomes, and Mortality of COVID-19 Patients. 2020.

6. Salepci E, Turk B, Ozcan SN, Bektas ME, Aybal A, Dokmetas I, et al. Symptomatology of COVID-19 from the otorhinolaryngology perspective: a survey of 223 SARS-CoV-2 RNA-positive patients. 2020:1-11.

7. Salepci E, Turk B, Ozcan SN, Bektas ME, Aybal A, Dokmetas I, et al. Otorhinolaryngologic and General Symptoms Survey of 223 COVID-19 Patients. V1 ed. Mendeley Data2020.

8. Feng W, Huang W, Ren JJAS. Class imbalance ensemble learning based on the margin theory. 2018;8(5):815.

9. Hofmann M, Klinkenberg R. RapidMiner: Data mining use cases and business analytics applications: CRC Press; 2016.

10. Kotu V, Deshpande B. Predictive analytics and data mining: concepts and practice with rapidminer: Morgan Kaufmann; 2014.

11. Modeler IS. Algorithms Guide.[(accessed on 25 November 2020)].

12. Andrei NJCO, Applications. Scaled conjugate gradient algorithms for unconstrained optimization. 2007;38(3):401-16.

13. Feng L, Li Z, Wang Y, Zheng C, Guan Y, editors. VLSI design of modified sequential minimal optimization algorithm for fast SVM training. 2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT); 2016: IEEE.

14. Tang J, Ning J, Liu X, Wu B, Hu RJCC-ADD. A novel amino acid sequence-based computational approach to predicting cell-penetrating peptides. 2019;15(3):206-11.

15. McCormick K, Salcedo J. IBM SPSS Modeler essentials: Effective techniques for building powerful data mining and predictive analytics solutions: Packt Publishing Ltd; 2017.

16. Wendler T, Gröttrup S. Data mining with SPSS modeler: theory, exercises and solutions: Springer; 2016.

17. Mierswa I, Klinkenberg R. RapidMiner Studio (9.8)[Data science, machine learning, predictive analytics]. 2020.

18. S. Y, Arslan AK, Yologlu S, Colak C. DTROC: Tanı Testleri ve ROC Analizi Yazılımı 2019 [Available from:*http://biostatapps.inonu.edu.tr/DTROC/*.

19. Chicco D, Jurman GJBg. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. 2020;21(1):6.

20. Released IC. IBM SPSS Statistics for Windows, Version 26.0. IBM Corp Armonk, NY; 2019.

21. IBM_Corp R. IBM SPSS Modeler for Windows, Version 18.0. IBM Corp, Armonk, NY. 2016.

22. Pustokhin DA, Pustokhina IV, Dinh PN, Phan SV, Nguyen GN, Joshi GP. An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19. Journal of Applied Statistics. 2020:1-18.

23. Yip SS, Klanecek Z, Naganawa S, Kim J, Studen A, Rivetti L, et al. Performance and Robustness of Machine Learning-based Radiomic COVID-19 Severity Prediction. 2020.

24. Feng S, Keung J, Yu X, Xiao Y, Bennin KE, Kabir MA, et al. COSTE: Complexity-based OverSampling TEchnique to alleviate the class imbalance problem in software defect prediction. 2020;129:106432.

25. Dong X, Yu Z, Cao W, Shi Y, Ma QJFoCS. A survey on ensemble learning. 2020:1-18.

26. Zhou K, Sun Y, Li L, Zang Z, Wang J, Li J, et al. Eleven Routine Clinical Features Predict COVID-19 Severity Uncovered by Machine Learning of Longitudinal Measurements. 2020.

27. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et al. Large-scale multi-omic analysis of COVID-19 severity. Cell systems. 2020.

28. de Terwangne C, Laouni J, Jouffe L, Lechien JR, Bouillon V, Place S, et al. Predictive Accuracy of COVID-19 World Health Organization (WHO) Severity Classification and Comparison with a Bayesian-Method-Based Severity Score (EPI-SCORE). Pathogens. 2020;9(11):880.

29. Ali AM, Ghafoor KZ, Maghdid HS, Mulahuwaish A. Diagnosing COVID-19 Lung Inflammation Using Machine Learning Algorithms: A Comparative Study. Internet of Medical Things for Smart Healthcare: Springer; 2020. p. 91-105.

Table 1: The detailed explanation of the variables/attributes in the dataset

Abbreviation	Explanation
Age	Birth year (year)
Gender	Gender (0=female, 1=male)
ComorbidityStatus	Comorbidity Status (0=no comorbidities, 1=with comorbidities)
GeneralMedication	General Medication Use (0=absence, 1=presence)
Hydroxychloroquine	Hydroxychloroquine use (0=not used, 1=used)
Oseltamivir	Oseltamivir use (0=not used, 1=used)
Azithromycin	Azithromycin use (0=not used, 1=used)
LopinavirRitonavir	Lopinavir/Ritonavir use (0=not used, 1=used)
Favipravir	Favipravir use (0=not used, 1=used)
Fever	Fever Presence (0=absence, 1=presence)
Cough	Cough Presence (0=absence, 1=presence)
Fatigue	Fatigue Presence (0=absence, 1=presence)
Dyspnea	Dyspnea Presence (0=absence, 1=presence)
MyalgiaArthralgia	Myalgia/Arthralgia Presence (0=absence, 1=presence)
Headache	Headache Presence (0=absence, 1=presence)
FrontalTypeHeadache	Frontal Type Headache Presence (0=absence, 1=presence)
NasalCongestion	Nasal Congestion Presence (0=absence, 1=presence)
Rhinorrhea	Rhinorrhea Presence (0=absence, 1=presence)
SoreThroat	Sore Throat Presence (0=absence, 1=presence)
DryThroat	Dry Throat Presence (0=absence, 1=presence)
SmellLoss	Smell Loss Presence (0=absence, 1=presence)
TasteLoss	Taste Loss Presence (0=absence, 1=presence)
Earache	Earache Presence (0=absence, 1=presence)
Dizziness	Dizziness Presence (0=absence, 1=presence)
HearingLoss	Hearing Loss Presence (0=absence, 1=presence)
PatientClinicalStatus	Patient Clinical Status (1=Mild Disease, 2=Moderate Disease, 3=Severe Disease, 4=Critical Diseas

Table 2. Baseline characteristics of the patients by disease severity

Variable	**Class**	**Patient Clinical Status**	**Patient Clinical Status**	**Patient Clinical**
		Mild Disease	**Mild Disease**	**Moderate Disea**
		n	**%**	**n**
Gender	Female	20	48.8	70
	Male	21	51.2	67
Comorbidity Status	Absence	28[a]	68.3	67[a]
	Presence	13[a]	31.7	70[a]
General Medication	Absence	34[a]	82.9	83[b]
	Presence	7[a]	17.1	54[b]
Hydroxychloroquine	Not Used	1	2.4	1
	Used	40	97.6	136
Oseltamivir	Not Used	18	43.9	47
	Used	23	56.1	90
Azithromycin	Not Used	33	80.5	115
	Used	8	19.5	22
Lopinavir/Ritonavir	Not Used	41[a,b]	100	137[b]
	Used	0[a,b]	0	0[b]
Favipiravir	Not Used	41[a]	100	137[a]
	Used	0[a]	0	0[a]
Fever Presence	Absence	29[a]	70.7	66[a.b]

Variable	Class	Patient Clinical Status	Patient Clinical Status	Patient Clinical
	Presence	12[a]	29.3	71[a.b]
Cough Presence	Absence	28[a]	68.3	59[b]
	Presence	13[a]	31.7	78[b]
Fatigue Presence	Absence	19[a]	46.3	38[a.b]
	Presence	22[a]	53.7	99[a.b]
Dyspnea Presence	Absence	33[a]	80.5	93[a]
	Presence	8[a]	19.5	44[a]
Myalgia/Arthralgia Presence	Absence	25	61	62
	Presence	16	39	75
Headache Presence	Absence	33	80.5	97
	Presence	8	19.5	40
Frontal Type Headache Presence	Absence	37	90.2	123
	Presence	4	9.8	14
Nasal Congestion Presence	Absence	33	80.5	115
	Presence	8	19.5	22
Rhiorrhea Presence	Absence	35	85.4	123
	Presence	6	14.6	14
Sore Throat Presence	Absence	30	73.2	101
	Presence	11	26.8	36
Dry Throat Presence	Absence	38	92.7	113
	Presence	3	7.3	24
Smell Loss Presence	Absence	23[a]	56.1	90[a.b]
	Presence	18[a]	43.9	47[a.b]
Taste Loss Presence	Absence	23	56.1	90
	Presence	18	43.9	47
Earache Presence	Absence	38	92.7	136
	Presence	3	7.3	1
Dizziness Presence	Absence	41[a.b]	100	136[b]
	Presence	0[a.b]	0	1[b]
Hearing Loss Presence	Absence	41[a.b]	100	137[b]
	Presence	0[a.b]	0	0[b]

*: Pearson's Chi-Squared Test. In each row, different superscript letters indicate statistical significance (Bonferroni-corrected Pearson chi-square test; $p<0.05$)

Table 3: The performance metrics of the individual and ensemble models for each COVID-19 severity category

Algorithm	Sample	Category	Accuracy	Precision	Sensitivity	Specificity	F1-Score	MCC	G-me
NN	**Training**	**Mild**	0.9976	0.9905	1	0.9968	0.9952	0.9936	0.9
		Moderate	0.9976	1	0.9899	1	0.9949	0.9934	0.9
		Severe	1	1	1	1	1	1	1
		Critical	1	1	1	1	1	1	1
	Testing	**Mild**	0.9389	0.8205	0.9697	0.9286	0.8889	0.8528	0.9
		Moderate	0.9179	0.9655	0.7368	0.9896	0.8358	0.7951	0.8
		Severe	0.984	0.9429	1	0.9783	0.9706	0.9604	0.9
		Critical	0.9919	0.9677	1	0.9894	0.9836	0.9785	0.9
SVM	**Training**	**Mild**	0.9879	0.9541	1	0.9838	0.9765	0.9688	0.9

Algorithm	Sample	Category	Accuracy	Precision	Sensitivity	Specificity	F1-Score	MCC	G-mc
		Moderate	0.9951	1	0.9798	1	0.9898	0.9867	0.9
		Severe	0.9879	1	0.9519	1	0.9754	0.9678	0.9
		Critical	0.9951	0.9817	1	0.9934	0.9907	0.9875	0.9
	Testing	**Mild**	0.9219	0.7805	0.9697	0.9053	0.8649	0.8203	0.9
		Moderate	0.9077	0.9643	0.7105	0.9891	0.8182	0.7741	0.8
		Severe	0.9672	1	0.8788	1	0.9355	0.9171	0.9
		Critical	0.9516	0.8333	1	0.9362	0.9091	0.8833	0.9
GUEST	**Training**	**Mild**	0.8958	0.8764	0.75	0.9562	0.8083	0.7415	0.8
		Moderate	0.848	0.6721	0.8283	0.8551	0.7421	0.643	0.8
		Severe	0.8665	0.6923	0.9519	0.8327	0.8016	0.725	0.8
		Critical	0.8665	0.9833	0.5514	0.9962	0.7066	0.6729	0.7
	Testing	**Mild**	0.8718	0.75	0.8182	0.8929	0.7826	0.6933	0.8
		Moderate	0.8293	0.7297	0.7105	0.8824	0.72	0.5973	0.7
		Severe	0.8644	0.6977	0.9091	0.8471	0.7895	0.7052	0.8
		Critical	0.8947	1	0.6	1	0.75	0.7246	0.7
Ensemble VS	**Training**	**Mild**	0.9976	0.9905	1	0.9968	0.9952	0.9936	0.9
		Moderate	0.9976	1	0.9899	1	0.9949	0.9934	0.9
		Severe	1	1	1	1	1	1	1
		Critical	1	1	1	1	1	1	1
	Testing	**Mild**	0.9398	0.8205	0.9697	0.93	0.8889	0.8536	0.9
		Moderate	0.9328	0.9677	0.7895	0.9896	0.8696	0.8327	0.8
		Severe	1	1	1	1	1	1	1
		Critical	0.9921	0.9677	1	0.9896	0.9836	0.9786	0.9
Ensemble CWVS	**Training**	**Mild**	0.9976	0.9905	1	0.9968	0.9952	0.9936	0.9
		Moderate	0.9976	1	0.9899	1	0.9949	0.9934	0.9
		Severe	1	1	1	1	1	1	1
		Critical	1	1	1	1	1	1	1
	Testing	**Mild**	0.9389	0.8205	0.9697	0.9286	0.8889	0.8528	0.9
		Moderate	0.9179	0.9655	0.7368	0.9896	0.8358	0.7951	0.8
		Severe	0.984	0.9429	1	0.9783	0.9706	0.9604	0.9
		Critical	0.9919	0.9677	1	0.9894	0.9836	0.9785	0.9
Ensemble HCWS	**Training**	**Mild**	0.9951	0.9811	1	0.9935	0.9905	0.9873	0.9
		Moderate	0.9951	1	0.9798	1	0.9898	0.9867	0.9
		Severe	0.9951	1	0.9808	1	0.9903	0.9871	0.9
		Critical	0.9951	0.9817	1	0.9934	0.9907	0.9875	0.9
	Testing	**Mild**	0.9398	0.8205	0.9697	0.93	0.8889	0.8536	0.9
		Moderate	0.9328	0.9677	0.7895	0.9896	0.8696	0.8327	0.8
		Severe	1	1	1	1	1	1	1
		Critical	0.9921	0.9677	1	0.9896	0.9836	0.9786	0.9

NN: Neural Network, **SVM:** Support Vector Machine,**VS:** Voting strategy, **QUEST** : Quick, unbiased, efficient statistical tree **CWVS:** Confidence weighted voting strategy,**HCWS:** Highest confidence wins strategy, **MCC:** The Matthews correlation coefficient

Table 4: Predictor importance values for each separate model

Predictor	Importance Values	Importance Values	Importance Values
	NN	**SVM**	**QUEST**
Age	0.24	0.22	0.20
Favipravir	0.24	0.15	0.27
DysneaPresence	0.22	0.13	0.15
CoughPresence	0.09	0.08	0.05
SmellLossPresence	0.04	0.06	0.06
LopinavirRitonavir	0.04	-	0.08
FatiguePresence	0.03	-	0.05
FeverPresence	0.03	-	-
FrontalTypeHeadachePresence	0.02	0.04	0.03
Gender	0.01	0.07	-
MyalgiaArthralgialPresence	-	0.10	0.07
Oseltamivir	-	0.07	-
NasalCongestionPresence	-	0.06	0.03

NN: Neural Network, **SVM:** Support Vector Machine.