Original article

Machine learning approach for automated predicting of COVID-19 severity based on clinical and paraclinical characteristics: Serum levels of zinc, calcium, and vitamin D

Azam Jahangirimehr [a], Elham Abdolahi Shahvali [f], Seyed Masoud Rezaeijo [b], Azam Khalighi [c], Azam Honarmandpour [d], Fateme Honarmandpour [c], Mostafa Labibzadeh [a], Nasrin Bahmanyari [a], Sahel Heydarheydari [e, *]

[a] Shoushtar Faculty of Medical Sciences, Shoushtar, Iran
[b] Department of Medical Physics, Ahvaz Jundishapour University of Medical Sciences, Ahvaz, Iran
[c] Khatam Al-Anbia Hospital, Shoushtar Faculty of Medical Sciences, Shoushtar, Iran
[d] Department of Midwifery, Shoushtar Faculty of Medical Sciences, Shoushtar, Iran
[e] Department of Radiology Technology, Shoushtar Faculty of Medical Sciences, Shoushtar, Iran
[f] Department of Nursing, Shoushtar Faculty of Medical Sciences, Shoushtar, Iran

ARTICLE INFO

SUMMARY

Background & aims: Considering that no standard therapy has yet been found for the novel coronavirus disease (COVID-19), identifying severe cases as early as possible, and such that treatment procedures can be escalated seems necessary. Hence, the present study aimed to develop a machine learning (ML) approach for automated severity assessment of COVID-19 based on clinical and paraclinical characteristics like serum levels of zinc, calcium, and vitamin D.

Methods: In this analytical cross-sectional study which was conducted from May 2020 to May 2021, clinical and paraclinical data sets of COVID-19-positive patients with known outcomes were investigated by combining statistical comparison and correlation methods with ML algorithms, including Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM).

Results: Our work revealed that some patients' characteristics including age, gender, cardiovascular diseases as an underlying condition, and anorexia as disease symptoms, and also some parameters which are measurable in blood samples including FBS and serum levels of calcium are factors that can be considered in predicting COVID-19 severity. In this regard, we developed ML predictive models that indicated accuracy and precision scores >90% for disease severity prediction. The SVM algorithm indicated better results than other algorithms by having a precision of 95.5%, recall of 94%, F1 score of 94.8%, the accuracy of 95%, and AUC of 94%.

Conclusions: Our results indicated that clinical and paraclinical features like calcium serum levels can be used for automated severity assessment of COVID-19.

1. Introduction

The SARS-CoV-2 is the causative agent of COVID-19 pneumonia, which the WHO identified as a pandemic [1—4]. Considering that no standard therapy has yet been found for the novel coronavirus disease (COVID-19) [5], some supplements like zinc, calcium, and vitamin D can strengthen the immune response against the SARS-CoV-2 virus and may effectively reduce disease incidence and severity. Since a well-balanced diet is the best manner to get all the

* Corresponding author.
E-mail addresses: a.jahangirimehr@gmail.com (A. Jahangirimehr), eabdolahi31@yahoo.com (E. Abdolahi Shahvali), rezaei-sm@ajums.ac.ir (S.M. Rezaeijo), khalighiazam@gmail.com (A. Khalighi), honarmandpour.a@gmail.com (A. Honarmandpour), Amirmasuood439@gmail.com (F. Honarmandpour), mlabibzadeh@yahoo.com (M. Labibzadeh), n.bahmanyari18@gmail.com (N. Bahmanyari), sheydari7@gmail.com (S. Heydarheydari).

essential nutrients, nutrition has become a priority to ensure a normal immune response [6].

Due to the direct-acting antiviral and immunomodulatory effects, zinc supplement is considered possible supportive care against COVID-19. Since patients with hypozincemia are vulnerable to viral infection, zinc deficiency may be closely related to the severe states of COVID-19. Hence, serum zinc level can be a predictive factor for COVID-19 severity [8—11]. Calcium is one of the most abundant minerals in the body, which makes up 39% of the body's total salts, and plays a vital role in the immune system. Since there may be a relationship between calcium deficiency and inflammation, hypocalcemia is highly noteworthy in COVID-19 [16]. In this regard, serum calcium level can be a prognostic factor in determining the severity of the disease. Vitamin D has immune regulating properties and plays an influential role in allergic and respiratory disorders [17—19]. It was found that vitamin D reduces respiratory infections, prevents asthma attacks, and controls asthma [20]. Vitamin D deficiency increases the risk of viral infections such as RSV-induced bronchiolitis and seasonal influenza [21].

Computed Tomography (CT) of the chest plays a valuable role in detecting the progress and severity of COVID-19 pneumonia. For COVID-19-positive patients, Ground-Glass Opacity (GGO) is a radiological finding in chest CT scans which reflects the severity of pulmonary inflammation. GGO is a time-consuming and error-prone subjective evaluation which generally dependent on reader experience [24,25]. Laboratory test results, including hematological and biochemical parameters, could provide predictive information about the COVID-19 disease severity [26,27]. It is well-known that even the most knowledgeable and experienced physicians can extract only a minor fraction of information contained in the laboratory test results. In contrast, machine learning (ML) can recognize subtle patterns in data and is suitable for differentiating various patterns observed in blood parameters [28,29]. ML is a subfield of Artificial intelligence (AI) that focuses on algorithms that enable computers to describe a model for complex relationships or patterns from empirical data without being explicitly programmed [30].

Since the statistical/machine learning (ML) approach can improve the predictive values of the mentioned parameters [31], in this study, a comprehensive assessment was designed to investigate a range of statistical and ML approaches. Indeed, the current study was conducted in two sections: In the first section, the relationship between clinical and paraclinical data like serum levels of zinc, calcium, and vitamin D with COVID-19 severity was investigated using statistical analysis. In the second section, machine learning-based algorithm models were implemented and developed to predict disease severity automatically.

2. Methods

2.1. Study type and patient selection

This analytical cross-sectional study was conducted from May 2020 to May 2021 on 93 COVID-19-positive patients (41 male and 52 female) who were admitted to Khatam Al-Anbia Hospital. Regarding the research objectives, a previous study [32], the ratio of individuals who suffered from the severe form of the disease and had low vitamin D levels (P = 0.77, α = 0.05, d = 0.09, considering the 10% dropout rate in the study, and using the sample size ratio formula, the sample size was estimated as 93. The subjects were enrolled in the study via a purposive sampling method. The sample of Nasopharyngeal and/or oropharyngeal swabs were used to detect SARS-CoV-2 nucleic acid by real-time reverse-transcriptase-polymerase-reaction (RT-PCR). At the beginning of admission,

blood samples were taken from patients for the desired tests, and the severity of the disease was determined. Blood samples were taken from the patients by a nurse (a project colleague) and sent to a private laboratory for more investigation. COVID-19 severity was determined based on the severity of pulmonary involvement using CT chest images. Inclusion criteria were as follows: being infected with COVID-19 based on PCR test; being over 15 years old, and fasting during sampling. Exclusion criteria were as follows: consuming Zinc, calcium, and vitamin D supplements over the past 24 h; consuming corticosteroids, cholesterol-reducing drugs such as cholestyramine, barbiturates, and phenytoin; being with kidney disease, chronic liver disease, bone disease, parathyroid disease, and cancer; undergoing treatment with zinc, calcium, and vitamin D. A checklist was utilized to record patients' data. All the data was kept confidential in compliance with the principles of the Declaration of Helsinki. The institutional ethics committee approved the study (No: 1399.017). Written informed consent was obtained from all individual participants. All procedures performed within this research that involved human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments.

2.2. Data collection

Data included demographic information (age and gender), clinical information and vital signs (history of underlying diseases, fever, cough, dyspnea, etc.), and paraclinical information (laboratory results). The laboratory test results were recorded as follows: serum levels of zinc, calcium, and vitamin D, Fasting Blood Sugar (FBS), Blood Urea Nitrogen (BUN), Creatinine (Cr), Sodium (Na), Potassium (K), White Blood Cells (WBC), Red Blood Cells (RBC), Hemoglobin (Hb), Hematocrit (HCT), Platelet (PLT), Lymphocyte (Lym), Neutrophils (Neut), Arterial Blood Gas Test (PH, pO2, pCO2, HCO3), Erythrocyte Sedimentation Rate (ESR). Serum zinc levels were measured by Atomic Absorption Spectrophotometry (AAS) with a reference range of 70—127 μg/dl. Serum calcium levels were measured with a reference range of 8.6—10.3 mg/dl. Serum vitamin D levels were measured by Enzyme-Linked Immunosorbent Assay (ELISA). Serum levels of vitamin D were divided into 30—100 ng/ml as sufficient, 10—30 ng/ml as insufficient, and <10 ng/ml as deficient.

2.3. Chest CT acquisition

COVID-19 severity was determined based on the severity of pulmonary involvement using CT chest images. CT images were subjectively assessed by an emergency medicine specialist, who was blinded to clinical and paraclinical (laboratory) results. Disease severity in confirmed cases was classified into three categories, including mild (n = 26), moderate (n = 30), and severe (n = 37). All chest CT images were acquired with a multi-slice CT scanner (GE Healthcare, USA). Patients were scanned in a supine position during breath-holding. Imaging parameters were as follows:

kVp = 120, mAs = 115, matrix size = 512 × 512, slice thickness = 7.5 mm.

2.4. Statistical analysis to identify the most significant and associative parameters

Patients' characteristics (demographical data, clinical data, and vital signs) and laboratory results (paraclinical data) were recorded on the specific checklist that was specifically designed for this study. The collected data were analyzed using IBM SPSS software for Windows (version 18). Descriptive statistics for the quantitative

variables were presented as mean, standard deviation, and median (IQR: Q1-Q3), and for the qualitative (categorical) variables were reported as the frequency with percentage. Independent t-test or Mann−Whitney test was employed to compare means, ANOVA or Kruskal−Wallis test to compare means between levels of variables, and Chi-square test to compare qualitative variables. Selecting a statistical test was dependent on data normality, which was assessed using the Kolmogorov−Smirnov test. Another statistical test including the Spearman rank-order correlation coefficient was employed to examine the relationship between quantitative variables and to predict some factors affecting the severity of COVID-19, respectively. The significance level was considered less than 0.05.

2.5. ML models to classify COVID-19 severity

Some patients' characteristics and blood parameters including age, gender, cardiovascular diseases as an underlying condition, anorexia as a disease symptom, FBS, and serum levels of calcium which have been shown statistically significant relationships with COVID-19 severity were entered into the ML proposed models as selected features. Three classification algorithms including Support Vector Machines (SVMs) [33], Decision Tree (DT) [34], and Random Forest (RF) [35] were employed in this study. The SVM transforms a training dataset to a higher dimension and optimizes a hyperplane to separate the two classes with minimum classification errors. The DT creates a tree-structured model to describe the relationships between features and a class label. The RF (DT ensemble algorithm) creates multiple trees through a re-sampling process called bootstrap aggregation [36]. We now address the hyperparameters of the ML procedures used when they were run. In the SVM procedure, we used a penalty of l2, a loss of hing, a c of 2, a random state of 0, and a max_iter of 2000. In the DT procedure, we used a random state of 0, entropy criterion, a minimum sample split of 2, and min_samples_leaf of 1. In the RF algorithm, n_estimators of 2000, a criterion of entropy, a min_samples_split of 4, and a random state of 0 were used. In the current study, the performance of the ML algorithms was acquired with five-fold cross-validation. In each fold, 4 divisions were employed for the training and validation process [37]. All experiments including data preprocessing and analysis were conducted on the Google Cloud computing service "Google Colab" (colab.research.google.com) using Python 3.7.

2.6. Evaluation matrices for the ML models

To evaluate the proposed models' performance, an area under the receiver operating characteristic (ROC) curve (AUC), accuracy, precision, recall, and f1-Score were calculated as follows:

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN) \qquad (1)$$

$$\text{Precision} = TP / (TP + FP) \qquad (2)$$

$$\text{Recall} = TP / (TP + FN) \qquad (3)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad (4)$$

TP, FP, TN, and FN represent the number of True Positive, False Positive, True Negative, and False Negative, respectively.

Study flowchart is presented in Fig. 1.

3. Results

In this study, 93 COVID-19 patients were examined with a mean and median age of 51.38 ± 15.75 and 51 (40−61), respectively. 55.9%

(n = 52) of patients were female and 44.1% (n = 41) were male. 47.3% of patients (n = 44) had no underlying conditions and 21.5% (n = 20) had cardiovascular diseases. The most common symptoms in cases were cough (60.2%) and dyspnea (40.9%). Disease severity in confirmed cases was classified into three categories, including mild (n = 26, 28%), moderate (n = 30, 32%), and severe (n = 37, 40%). The results of comparing patients' demographical and clinical data according to the COVID-19-related severity are summarized in Table 1. The median [interquartile range, (IQR)] length of hospital stay was four days [3−5].

As shown in Table 1, there was a significant and direct relationship between the severity of the disease with the age factor ($\rho = 0.457$, p < 0.001). 53.7% of patients who suffered from a severe form of COVID-19 were male (p = 0.032). According to the results, 38.6% of subjects who had no history of underlying conditions experienced a mild form of COVID-19; 25.0% of them suffered from a severe form of COVID-19 (p = 0.039). Whilst, 60% of patients with a history of cardiovascular diseases suffered from a severe form of COVID-19 (p = 0.024). Anorexia was inversely associated with the severity of the disease; 72.3% of patients with anorexia had a non-severe form of the disease (p = 0.014). Other symptoms and vital signs had no effect on the severity of COVID-19 (p > 0.05).

The results of comparing patients' paraclinical data (laboratory test results) according to the COVID-19-related severity are summarized in Table 2. As presented in Table 2, FBS levels were higher in patients with a severe form of COVID-19. In other words, there was a significant and direct relationship between FBS levels and COVID-19 severity ($\rho = 0.36$, p = 0.028). This is while the levels of calcium were lower in these cases. In other words, there was a significant and inverse relationship between calcium serum levels and COVID-19 severity ($\rho = -0.292$, p = 0.043). The mean serum level of calcium in the subjects was 9.14 ± 0.39 mg/dl with a range of 8.40−10.30 mg/dl. Thirty-nine subjects (42%) had calcium deficiency (<9 mg/dl), and 54 of them (58%) had normal calcium levels (9−10.5 mg/dl). The mean serum level of vitamin D in the subjects was 11.21 ± 21.40 ng/ml with a range of 8−56 ng/ml. Sixty-eight subjects (73%) had vitamin D deficiency (<30 ng/ml), and 25 of them (27%) had normal serum levels of vitamin D (30−100 ng/ml). The mean serum level of zinc in the subjects was 67.61 ± 15.10 μg/dl with a range of 41−112 μg/dl. Forty-nine subjects (52.7%) had zinc deficiency (<70 μg/dl), and 44 of them (47.3%) had normal serum levels of zinc (70−114 μg/dl).

Although serum calcium levels were not significantly different between the male and female subjects (P = 0.059), it was lower in female cases than in male ones. In this regard, serum vitamin D levels were also not statistically significant between the male and female subjects (P = 0.182); however, serum levels of vitamin D were lower in female cases than in male ones. No significant difference in serum zinc levels was observed between the male and female subjects (P = 0.591).

We had clinical reports of 93 patients with confirmed COVID-19 that were labeled into three classes. The classification was done using ML procedures by supplying them with the selected features (parameters) that were extracted in the pre-processing step. Table 3 shows an analysis of the ML procedures that were used for performing this task. The results showed that the SVM algorithm indicates better results than other algorithms by having a precision of 95.5%, recall of 94%, F1 score of 94.8%, the accuracy of 95%, and AUC of 94%. The lowest performance values have yielded for the DT classifier with a precision of 93.5%, recall of 91%, F1 score of 92.3%, and AUC of 92%. The ROC curves for the models are shown in Fig. 2. According to the mean ROC (blue curve), as the mean AUC of folds gets closer to the upper and left-hand boundaries, the model indicates a smaller error.
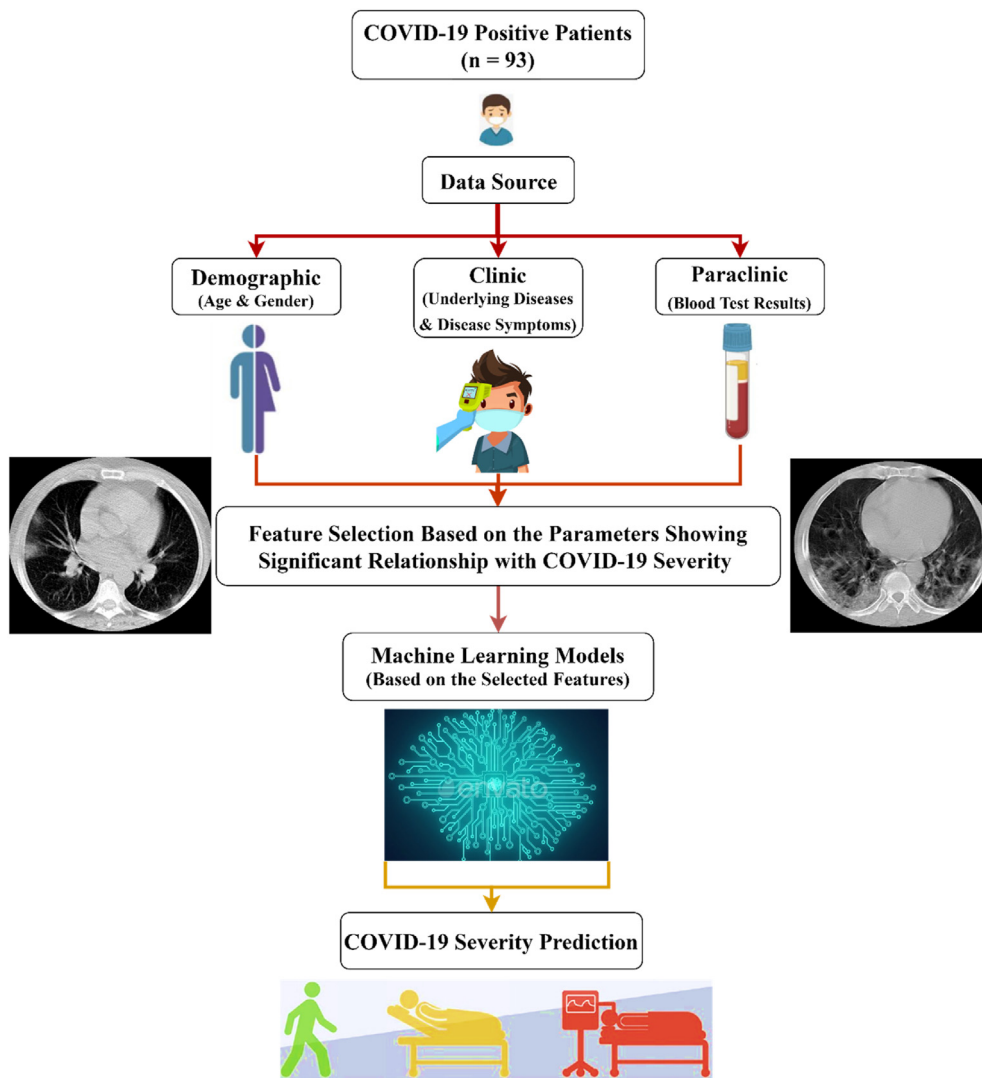
A. Jahangirimehr, E. Abdolahi Shahvali, S.M. Rezaeijo et al.

Clinical Nutrition ESPEN 51 (2022) 404–411

Fig. 1. Study flowchart.

Table 1
Comparison of patients' characteristics (demographical and clinical data) according to COVID-19-related severity.

Patient Characteristics		Number of Patients/ Mean ± Std dev	Disease Severity			p-value
			Mild (n = 26)	Moderate (n = 30)	Severe (n = 37)	
Age (year)		51.38 ± 15.75	42.2 ± 13.7	49.7 ± 15.4	59.1 ± 13.5	0.000*
Gender	Female	52 (55.9%)	19 (36.5%)	18 (34.6%)	15 (28.8%)	**0.032***
	Male	41 (44.1%)	7 (17.1%)	12 (29.3%)	22 (53.7%)	
Underlying Diseases	Cardiovascular Disease	20 (21.5%)	1 (5.0%)	7 (35.0%)	12 (60.0%)	**0.024***
	Diabetes	15 (16.1%)	3 (20.0%)	2 (13.3%)	10 (66.7%)	0.06
	Hypertension	10 (10.8%)	1 (10.0%)	3 (30.0%)	6 (60.0%)	0.292
	Pulmonary Disease	8 (8.6%)	3 (37.5%)	4 (50.0%)	1 (12.5%)	0.249
	No Underlying Diseases	44 (47.3%)	17 (38.6%)	16 (36.4%)	11 (25.0%)	**0.039***
Disease Symptoms	Cough	56 (60.2%)	15 (26.8%)	19 (33.9%)	22 (39.3%)	0.905
	Dyspnea	38 (40.9%)	11 (28.9%)	11 (28.9%)	16 (42.1%)	0.849
	Fever	27 (29.0%)	10 (37.0%)	5 (18.5%)	12 (44.4%)	0.196
	Muscle Pain	18 (19.4%)	7 (38.9%)	5 (27.8%)	6 (33.3%)	0.515
	Anorexia	18 (19.4%)	10 (55.6%)	3 (16.7%)	5 (27.8%)	**0.014***
	Fatigue	12 (12.9%)	5 (41.7%)	3 (25.0%)	4 (33.3%)	0.523
	Headache	11 (11.8%)	4 (36.4%)	4 (36.4%)	3 (27.3%)	0.647
	Digestive Symptoms	8 (8.6%)	2 (25.0%)	4 (50.0%)	2 (25.0%)	0.486
	Sore Throat	6 (6.5%)	2 (33.3%)	3 (50.0%)	1 (16.7%)	0.460
	Anosmia	2 (2.2%)	0 (0.0%)	2 (100.0%)	0 (0.0%)	0.117

*significance level: p < 0.05.

Table 2
Comparison of patients' characteristics (paraclinical data) according to COVID-19-related severity.

Patient Characteristics		Mean ± Std dev	Disease Severity			p-value
			Mild (n = 26)	Moderate (n = 30)	Severe (n = 37)	
Blood Test Results[a]	Zinc (μg/dL)	67.61 ± 15.10	69.1 ± 15.2	69.4 ± 17.8	65.0 ± 12.4	0.417
	Calcium (mg/dL)	9.14 ± 0.39	9.2 ± 0.3	9.1 ± 0.4	9.0 ± 0.3	**0.043***
	Vitamin D (ng/mL)	11.21 ± 21.40	20.0 ± 10.4	19.1 ± 11.2	24.5 ± 11.4	0.077
	FBS (mg/dL)	151.56 ± 80.78	105.5 ± 36.4	140.9 ± 78.5	173.1 ± 86.6	**0.028***
	BUN (mg/dL)	17.97 ± 12.89	13.3 ± 3.6	21.0 ± 18.5	18.4 ± 11.0	0.301
	Cr (mg/dL)	1.08 ± 0.48	0.9 ± 0.1	1.1 ± 0.6	1.1 ± 0.4	0.169
	NA (mmol/L)	138.29 ± 12.43	140.4 ± 1.8	139.3 ± 4.4	136.3 ± 18.2	0.344
	K (mmol/L)	4.92 ± 4.77	6.6 ± 9.6	4.2 ± 0.3	4.4 ± 0.5	0.236
	WBC (10^9/L)	6.81 ± 2.63	6.9 ± 2.3	6.4 ± 3.3	6.9 ± 2.3	0.814
	RBC (10^{12}/L)	4.53 ± 0.64	4.6 ± 0.4	4.3 ± 0.5	4.5 ± 0.7	0.329
	Hb (g/dL)	12.46 ± 1.55	13.1 ± 1.4	12.4 ± 1.6	12.1 ± 1.5	0.06
	HCT (%)	37.96 ± 4.16	39.0 ± 4.6	36.8 ± 4.3	38.1 ± 3.7	0.280
	PLT (10^9/L)	224.96 ± 84.71	232.2 ± 61.7	193.6 ± 71.0	240.1 ± 98.4	0.140
	Lym (10^9/L)	27.65 ± 18.19	27.4 ± 16.5	29.2 ± 16.6	26.7 ± 20.2	0.520
	Neut (10^9/L)	67.24 ± 12.53	68.0 ± 12.2	62.9 ± 15.2	69.5 ± 10.2	0.174
	PH	7.37 ± 0.06	7.3 ± 0.04	7.3 ± 0.04	7.3 ± 0.08	0.869
	pCO2 (mmHg)	37.54 ± 6.68	38.2 ± 4.5	37.2 ± 7.0	37.2 ± 8.0	0.947
	pO2 (mmHg)	34.56 ± 11.83	35.0 ± 12.0	34.4 ± 12.6	34.5 ± 11.1	0.101
	HCO3 (meq/L)	22.03 ± 2.35	21.5 ± 0.6	21.6 ± 2.5	22.6 ± 2.9	0.554
	ESR (mm/hr)	37.42 ± 24.87	33.8 ± 17.7	22.2 ± 12.1	47.0 ± 29.2	0.097

*significance level: p < 0.05.
   [a] FBS: Fasting Blood Sugar, BUN: Blood Urea Nitrogen, Cr: Creatinine, Na: Sodium, K: Potassium, WBC: White Blood Cells, RBC: Red Blood Cells, Hb: Hemoglobin, HCT: Hematocrit, PLT: Platelet, Lym: Lymphocyte, Neut: Neutrophils, PH, pO2, pCO2, and HCO3: Arterial Blood Gas Test, ESR: Erythrocyte Sedimentation Rate.

Table 3
Comparison of the ML models based on the performance metrics.

Model[a]	Performance Metrics (%)				
	Precision	Recall	F1 score	Accuracy	AUC
SVM	**95.5**	**94**	**94.8**	**95**	**94**
DT	93.5	91	92.3	93	92
RF	94	93	93.5	93	93

   [a] SVM: Support Vector Machine, DT: Decision Tree, RF: Random Forest.

4. Discussion

COVID-19 has become a serious health concern worldwide. The validity and availability of early predictors of disease severity is an inevitable need in confirmed COVID-19 patients. In this regard, classifications of disease severity are of very great significance in prevention and treatment allocation during the worldwide outbreak of COVID-19. In the current study, demographical, clinical, and paraclinical data like serum levels of zinc, calcium, and vitamin D were assessed as prognostic factors to predict COVID-19 severity using ML models.

Our findings showed that serum levels of zinc and vitamin D did not affect the COVID-19 severity. This may be due to the fact that serum levels of both were lower in all three groups than normal. In this regard, Bansal et al. concluded that zinc supplementation does not affect the severity of pneumonia [38]. However, Wessels et al. stated that zinc supplements are associated with shortening symptoms, reducing disease severity, and, most importantly, decreasing morbidity and mortality in COVID-19 patients [39]. Previous studies reported that there is a significant inverse relationship between serum levels of vitamin D and asthma severity [21,40]. In another study, which was conducted on children with asthma in North America, low levels of vitamin D were directly associated with an increased risk of hospitalization [41]. Amrein et al. stated that vitamin D3 deficiency is one of the causes of elevated plasma glucose and dyslipidemia. This is while, similar to the zinc serum levels, no significant difference was observed in the disease severity concerning vitamin D serum levels in our study. This may be due to the fact that vitamin D serum levels were lower

in all three groups than normal. in agreement with our results, Radujkovic et al. stated that vitamin D levels did not differ significantly between the groups with different disease severity [42]. In a randomized clinical trial study, Murai et al. concluded that a high dose of vitamin D3, compared with a placebo, did not significantly reduce hospitalization [43]. Among hospitalized patients with COVID-19, a single high dose of vitamin D3, compared with a placebo, did not significantly reduce hospital length of stay. The findings do not support the use of vitamin D3 for the treatment of moderate to severe COVID-19. Nevertheless, serum calcium levels had a significant and direct effect on disease severity and can be helpful in predicting the severe forms of the COVID-19 disease. The results of this study showed that COVID-19 severity was significantly different in patients according to serum calcium levels; the lower the calcium level, the more severe the disease. Recently, a high prevalence of hypocalcemia in COVID-19 patients was reported in several studies. Sun et al. stated that serum levels of calcium were associated with disease severity and prognosis in COVID-19 patients [44]. The results of Yung et al. were consistent with those of the present study, as they reported that hypocalcemia might be an indicator for patients who are likely to have severe disease [45].

In this study, approximately 40% of confirmed COVID-19 patients suffered from severe pulmonary involvement. While Cao X et al. reported that most COVID-19 patients showed mild to moderate symptoms, and approximately 15% suffered from severe pneumonia [46]. Huang et al. indicated that the severe forms of disease resulted in mortality [47]. Existing studies suggest that nutrients including zinc, calcium, and vitamin D may play a role in the severity of COVID-19 disease. The results of our study showed that the severity of the disease was statistically significant and directly related to the patients' age. The severity of the disease increases with age. Lauc et al. stated that age is the most critical predictor of COVID-19 disease severity [48]. Huang et al. also named old age as one of the causes of COVID-19 mortality [26]. Targher et al. described older age as an independent variable with a higher risk of severe COVID-19 [49]. In Lix et al.'s study, elderly patients were associated with severe COVID-19 [50]. The results of the present study showed that the severity of the disease is higher
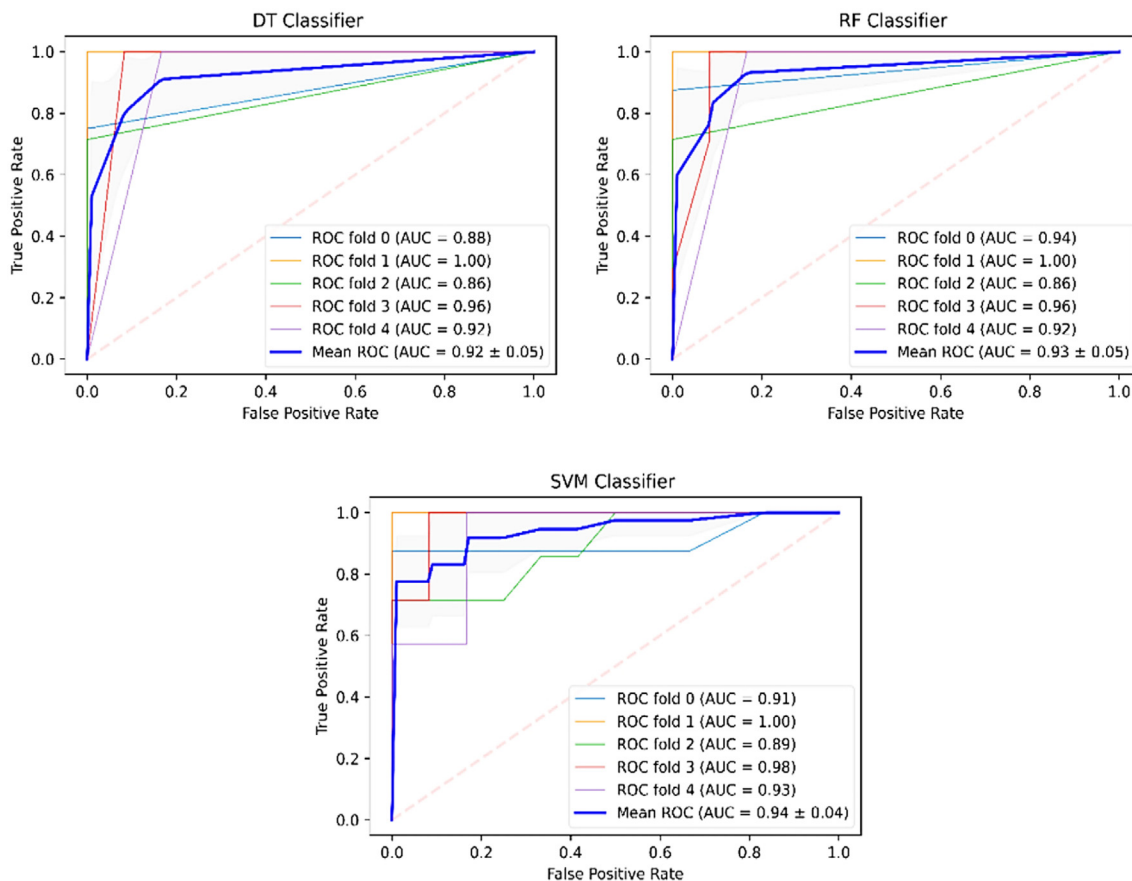
Fig. 2. The ROC curves of the ML models for the five-fold cross-validation.

in male than female subjects. Fei Zhou et al. showed that most of the patients were male [45]. Also, in the study of Targher et al., males were among the variables that were independently associated with a higher risk of severe COVID-19 disease [49]. In Lix et al.'s study, male patients were related to severe COVID-19 and death [50]. While in Shamsizadeh et al.'s study [51], no statistically significant relationship was found between disease severity and gender. According to our findings, there was a statistically significant and direct relationship between FBS and disease severity. The higher the patient's blood sugar, the higher the severity of the disease. Targher et al. reported that diabetes is associated with an approximately 4-fold increase in the risk of severe COVID-19 [49]. Results of Deng et al. [52], Ruan et al. [44], and Huang et al. [47] also linked diabetes to disease severity.

In this investigation, we developed ML algorithms including DT, RF, and SVM that showed accuracy and precision scores >90% for predicting COVID-19 disease severity. We found that the SVM classifier achieved the highest performance. Patel et al. [53] used socio-demographic data, clinical data, and blood panel profile data at the time of initial presentation to develop ML algorithms for predicting the need for intensive care (ICU) and mechanical ventilation. Among the algorithms considered, the Random Forest (RF) classifier performed the best with AUC = 0.80 for predicting ICU need and AUC = 0.82 for predicting the need for mechanical ventilation. In another study, Marcos et al. [54] trained, validated, and externally tested an ML model to early identify patients who will die or require mechanical ventilation during hospitalization from clinical and laboratory features obtained at admission. In the development cohort, the model obtained an AUC of 0.85 (95% confidence interval [CI], 0.82 to 0.87) for predicting the severity of

disease progression. In the external testing cohort, the model performed an AUC of 0.83 (95% CI, 0.81 to 0.85). We obtained better results than the studies of Liu et al. and Marcos et al. (AUC of 94 by SVM).

5. Conclusion

Our results indicated that clinical and paraclinical features like calcium serum levels can be used for automated severity assessment of COVID-19. The ML predictive models may potentially help triage COVID-19-positive patients. With this approach, data from laboratory test results could be used to identify patients with COVID-19 who are at high risk of mortality. To improve the performance of the proposed ML model, a comprehensive study with a large sample size that should be conducted on data from other parts of the world is suggested.

Availability of data and materials

Data is available upon request.

Authors' contributions

AJ- Conception and design of study; Analysis and interpretation of data; Writing original draft, review and editing.

EA- Interpretation of data; Writing original draft, review and editing.

SMR- Analysis and interpretation of data; Writing original draft, review and editing.

AKH- Conception and design of study.

ML- Data collection.

NB- Data collection.

SH- Interpretation of data; Writing original draft, review and editing.

Consent for publication

All authors agree with the final revisions.

Ethical statement

All procedures performed within this research which involved human participants was in accordance with the ethical standards of the institutional research committee (No: 1399.017) and with the 1964 Helsinki declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Declaration of competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

Acknowledgments

References

[1] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. Addendum: a pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;588(7836):E6.

[2] Organization WH. World Health Organization coronavirus disease (COVID-19) dashboard. World Health Organization; 2020.

[3] Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19—studies needed. N Engl J Med 2020;382(13):1194—6.

[4] Garcia-Beltran WF, Lam EC, Astudillo MG, Yang D, Miller TE, Feldman J, et al. COVID-19-neutralizing antibodies predict disease severity and survival. Cell 2021;184(2):476—88. e11.

[5] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395(10223):497—506.

[6] Zupo R, Castellana F, Sardone R, Sila A, Giagulli VA, Triggiani V, et al. Preliminary trajectories in dietary behaviors during the COVID-19 pandemic: a public health call to action to face obesity. Int J Environ Res Publ Health 2020;17(19):7073.

[8] Yasui Y, Yasui H, Suzuki K, Saitou T, Yamamoto Y, Ishizaka T, et al. Analysis of the predictive factors for a critical illness of COVID-19 during treatment-relationship between serum zinc level and critical illness of COVID-19. Int J Infect Dis 2020;100:230—6.

[9] Fraker P, Gershwin ME, Good R, Prasad A, editors. Interrelationships between zinc and immune function. Federation proceedings; 1986.

[10] Fraker PJ, DePasquale-Jardieu P, Zwickl CM, Luecke RW. Regeneration of T-cell helper function in zinc-deficient adult mice. Proc Natl Acad Sci USA 1978;75(11):5660—4.

[11] Devaux CA, Rolain J-M, Raoult D. ACE2 receptor polymorphism: susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. J Microbiol Immunol Infect 2020;53(3):425—35.

[16] Alemzadeh E, Alemzadeh E, Ziaee M, Abedi A, Salehiniya H. The effect of low serum calcium level on the severity and mortality of Covid patients: a systematic review and meta-analysis. Immun Inflamm Dis 2021;9(4):1219—28.

[17] Mortazavi A. The effect of administration of oral single dose vitamin D on outcome pediatric pneumonia in Boali hospital of Ardabil. Med J Ardabil Univ; 2004. p. 10—3.

[18] Kliegman RMSB, St Geme JW. Nelson textbook of pediatric. 19th ed. Elsevier Saunders PA; 2011.

[19] Merewood A, Mehta SD, Grossman X, Chen TC, Mathieu JS, Holick MF, et al. Widespread vitamin D deficiency in urban Massachusetts newborns and their mothers. Pediatrics 2010;125(4):640—7.

[20] Ginde AA, Mansbach JM, Camargo CA. Vitamin D, respiratory infections, and asthma. Curr Allergy Asthma Rep 2009;9(1):81—7.

[21] Hollams E, Hart P, Holt B, Serralha M, Parsons F, De Klerk N, et al. Vitamin D and atopy and asthma phenotypes in children: a longitudinal cohort study. Eur Respir J 2011;38(6):1320—7.

[24] Rezaeijo SM, Ghorvei M, Alaei M, editors. A machine learning method based on lesion segmentation for quantitative analysis of CT radiomics to detect covid-19. 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). IEEE; 2020.

[25] Masoud Rezaeijo S, Abedi-Firouzjah R, Ghorvei M, Sarnameh S. Screening of COVID-19 based on the extracted radiomics features from chest CT images. J X Ray Sci Technol 2021;(Preprint):1—15.

[26] Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. Sci China Life Sci 2020;63(3):364—74.

[27] Wu J, Zhang P, Zhang L, Meng W, Li J, Tong C, et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv 2020. https://doi.org/10.1101/2020.04.02.20051136.

[28] Gunčar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M, et al. An application of machine learning to haematological diagnosis. Sci Rep 2018;8(1):1—12.

[29] Kukar M, Gunčar G, Vovko T, Podnar S, Černelč P, Brvar M, et al. COVID-19 diagnosis by routine blood tests using machine learning. Sci Rep 2021;11(1):1—9.

[30] Panesar A. Machine learning and AI for healthcare. Springer; 2019.

[31] 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. Comput Struct Biotechnol J 2021;19:3640—9.

[32] Maghbooli Z, Sahraian MA, Ebrahimi M, Pazoki M, Kafan S, Tabriz HM, et al. Vitamin D sufficiency, a serum 25-hydroxyvitamin D at least 30 ng/mL reduced risk for adverse clinical outcomes in patients with COVID-19 infection. PLoS One 2020;15(9):e0239799.

[33] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl 1998;13(4):18—28.

[34] Brijain M, Patel R, Kushik M, Rana K. A survey on decision tree algorithm for classification. 2014.

[35] Boulesteix AL, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdiscip Rev Data Min Knowl Discov 2012;2(6):493—507.

[36] Rezaeijo SM, Ghorvei M, Abedi-Firouzjah R, Mojtahedi H, Zarch HE. Detecting COVID-19 in chest images based on deep transfer learning and machine learning algorithms. Egyptian Journal of Radiology and Nuclear Medicine 2021;52(1):1—12.

[37] Advanced automatic segmentation of tumors and survival prediction in head and neck cancer. In: Salmanpour MR, Hajianfar G, Rezaeijo SM, Ghaemi M, Rahmim A, editors. 3D head and neck tumor segmentation in PET/CT challenge. Springer; 2021.

[38] Bansal A, Parmar VR, Basu S, Kaur J, Jain S, Saha A, et al. Zinc supplementation in severe acute lower respiratory tract infection in children: a triple-blind randomized placebo controlled trial. Indian J Pediatr 2011;78(1):33—7.

[39] Wessels I, Rolles B, Rink L. The potential impact of zinc supplementation on COVID-19 pathogenesis. Front Immunol 2020;11:1712.

[40] Barassi A, Pezzilli R, Mondoni M, Rinaldo RF, Davì M, Cozzolino M, et al. Vitamin D in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) patients with non-invasive ventilation support. Panminerva medica; 2021.

[41] Brehm JM, Schuemann B, Fuhlbrigge AL, Hollis BW, Strunk RC, Zeiger RS, et al. Serum vitamin D levels and severe asthma exacerbations in the Childhood Asthma Management Program study. J Allergy Clin Immunol 2010;126(1):52—8. e5.

[42] Radujkovic A, Hippchen T, Tiwari-Heckler S, Dreher S, Boxberger M, Merle U. Vitamin D deficiency and outcome of COVID-19 patients. Nutrients 2020;12(9):2757.

[43] Murai IH, Fernandes AL, Sales LP, Pinto AJ, Goessler KF, Duran CSC, et al. Effect of a single high dose of vitamin D3 on hospital length of stay in patients with moderate to severe COVID-19: a randomized clinical trial. JAMA 2021;325(11):1053—60.

[44] Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020;46(5):846—8.

[45] Yang C, Ma X, Wu J, Han J, Zheng Z, Duan H, et al. Low serum calcium and phosphorus and their clinical performance in detecting COVID-19 patients. J Med Virol 2021;93(3):1639—51.

[46] Cao X. COVID-19: immunopathology and its implications for therapy. Nat Rev Immunol 2020;20(5):269—70.

[47] Huang I, Lim MA, Pranata R. Diabetes mellitus is associated with increased mortality and severity of disease in COVID-19 pneumonia—a systematic review, meta-analysis, and meta-regression. Diabetes Metabol Syndr: Clin Res Rev 2020;14(4):395—403.

[48] Lauc GSD. Biomarkers of biological age as predictors of COVID-19 disease severity. Aging (Albany NY) 2020;12(8):6490.

[49] Targher G, Mantovani A, Wang X-B, Yan H-D, Sun Q-F, Pan K-H, et al. Patients with diabetes are at higher risk for severe illness from COVID-19. Diabetes & metabolism 2020;46(4):335.

[50] Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol 2020;146(1):110—8.

[51] Ahmad Shamsizadeh, Roya Nikfar MS, Ziaei Kajbaf Tahereh, Saberi-Demneh Amir, Karbalaei Reza. Assessment of serum 25(OH)D level in infants with bronchiolitis. Tehran Univ Med J 2017;75(12):888—93.

[52] Deng S-Q, Peng H-J. Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. J Clin Med 2020;9(2):575.

[53] Patel D, Kher V, Desai B, Lei X, Cen S, Nanda N, et al. Machine learning based predictors for COVID-19 disease severity. Sci Rep 2021;11(1):1—7.

[54] Marcos M, Belhassen-García M, Sánchez-Puente A, Sampedro-Gomez J, Azibeiro R, Dorado-Díaz P-I, et al. Development of a severity of disease score and classification model by machine learning for hospitalized COVID-19 patients. PLoS One 2021;16(4):e0240200.