



# Development and validation of prognostic scoring system for COVID-19 severity in South India

Vishnu Shankar<sup>1</sup> · Pearlsy Grace Rajan<sup>1</sup> · Yuvaraj Krishnamoorthy<sup>2</sup>  · Damal Kandadai Sriram<sup>3</sup> · Melvin George<sup>4</sup> · S. Melina I. Sahay<sup>4</sup> · B. Jagan Nathan<sup>3</sup>

Received: 30 October 2021 / Accepted: 29 November 2021  
© The Author(s), under exclusive licence to Royal Academy of Medicine in Ireland 2021

## Abstract

**Background** Development of a prediction model using baseline characteristics of COVID-19 patients at the time of diagnosis will aid us in early identification of the high-risk groups and devise pertinent strategies accordingly. Hence, we did this study to develop a prognostic-scoring system for predicting the COVID-19 severity in South India.

**Methods** We undertook this retrospective cohort study among COVID-19 patients reporting to Hindu Mission Hospital, India. Multivariable logistic regression using the LASSO procedure was used to select variables for the model building, and the nomogram scoring system was developed with the final selected model. Model discrimination, calibration, and decision curve analysis (DCA) was performed.

**Results** In total, 35.1% of the patients in the training set developed severe COVID-19 during their follow-up period. In the basic model, nine variables (age group, sex, education, chronic kidney disease, tobacco, cough, dyspnea, olfactory-gustatory dysfunction [OGD], and gastrointestinal symptoms) were selected and a nomogram was built using these variables. In the advanced model, in addition to these variables (except OGD), C-reactive protein, lactate dehydrogenase, ferritin, d-dimer, and CT severity score were selected. The discriminatory power (c-index) for basic model was 0.78 (95%CI: 0.74–0.82) and advanced model was 0.83 (95%CI: 0.79–0.87). DCA showed that both the models are beneficial at a threshold probability around 10–95% than treat-none or treat-all strategies.

**Conclusion** The present study has developed two separate prognostic-scoring systems to predict the COVID-19 severity. This scoring system could help the clinicians and policymakers to devise targeted interventions and in turn reduce the COVID-19 mortality in India.

**Keywords** Clinical decision rules · COVID-19 · Prognosis · SARS-CoV-2

## Introduction

The World Health Organization (WHO) has declared that COVID-19 is a “Public Health Emergency of International Concern (PHEIC)” on January 30, 2020, and on March 11, 2020; it was announced as a pandemic [1, 2]. As of August

25, 2021, a total of 214,241,311 laboratory-confirmed cases had been documented globally spanning over 220 countries and territories [3]. However, majority of these COVID-19 patients experienced only mild-to-moderate illness requiring no special treatment [4]. About 20% of patients develop severe-to-critical form of illness, at risk of progressing to acute respiratory distress syndrome (ARDS), acute respiratory failure (ARF), and/or multiple organ dysfunction [4]. Moreover, the risk of a fatal outcome is higher in patients with the severe form of the disease than in those with the non-severe form. Though there is no available treatment specific to COVID-19, early recognition of high-risk individuals and providing supportive management for these patients at risk of poorer prognosis might contribute to significantly reducing the mortality [5].

✉ Yuvaraj Krishnamoorthy  
yuvi.1130@gmail.com

<sup>1</sup> Department of General Medicine, Hindu Mission Hospital, Chengalpattu, India  
<sup>2</sup> Department of Community Medicine, ESIC Medical College & PGIMS, K.K. Nagar, Chennai, India  
<sup>3</sup> Hindu Mission Hospital, Chengalpattu, India  
<sup>4</sup> Clinical Research, Hindu Mission Hospital, Chengalpattu, India

Developing a reliable scoring system that can be applied during early clinical assessment of the newly diagnosed patients might help in allocation and prioritization of resources in the healthcare setting. In addition, the patients at high risk of developing COVID-19 severity should be monitored continuously using such reliable scoring system during the period of hospital stay. Likewise, this prediction tool can be used to reduce the patient admission rate especially those with significantly lower risk of developing any serious complications. These patients can be safely handled by self-quarantining, which can significantly minimize the burden on the healthcare facilities.

Previous studies have identified several sociodemographic, behavioral, biochemical, and radiological parameters as potential predictors of COVID-19 severity [6–15]. However, the extent to which these predictors impact COVID-19 severity and deaths is largely region-specific. Several studies exploring models for predicting these outcomes are available from varied study settings [6–8]. Nevertheless, work supporting the development of a prediction model for examining the severity for COVID-19 patients has not previously been attempted in an Indian setting. Furthermore, developing a prediction model using baseline parameters will aid us in early identification of the high-risk groups and devise pertinent strategies to reduce the risk of death among newly diagnosed COVID-19 patients. Hence, we did this study with an aim of developing and validating a statistical prediction model to predict the probability of COVID-19 severity in South India.

## Methods

### Study setting and study population

We conducted this study as a part of large-scale retrospective cohort study (COSMOS–COvid Severity and Mortality Scoring). We developed this prediction model by retrospectively analyzing the data among COVID-19 patients reporting to Hindu Mission Hospital, India. The Hindu Mission Hospital is a 220-bedded multispecialty tertiary care hospital located in Tamil Nadu, Southern part of India. During this pandemic, the hospital was actively providing outpatient, inpatient, teleconsultation, intensive care, and specialist services for COVID-19 patients.

We collected data from the pre-existing records of all the adult COVID-19 patients ( $\geq 18$  years) who have reported to the facility between April 26 and June 25, 2021 (2-month period coinciding with the second wave of COVID-19 pandemic in India).

### Sample size and sampling technique

Sample size was calculated based on a previous similar study on prediction model for risk of COVID-19 severity. To perform a prediction model, a minimum of 20 outcome events per variable is required. We were planning to make a prediction model with at least eight variables, and hence, a minimum of 160 events should occur in our study. Based on the previous evidence, 35.4% of COVID-19 patients had developed a severe form of condition within the follow-up period [6]. Based on this proportion and number of predictors to be included in the model, minimum sample size required will be 452. However, we will be including all the participants satisfying the inclusion criteria during the study period.

In total, 642 patients met the eligibility criteria during the study period. However, 608 patients had all the information required to develop the prediction model and included in the analysis.

### Study procedure

Data collection was started after obtaining ethical approval to retrieve the pre-existing medical records of the COVID-19 patients during the study period. The data was anonymized and used a pretested semi-structured questionnaire to gather the sociodemographic details (age, sex, education, occupation), behavioral habits (smoking, alcohol), self-reported comorbidities such as diabetes mellitus (DM), hypertension (HTN), chronic kidney disease (CKD), coronary artery disease (CAD), chronic lung conditions (chronic obstructive pulmonary disease [COPD], tuberculosis [TB], bronchial asthma [BA]), hypothyroidism, old cerebrovascular accidents (CVA), cancer, and HIV-COVID-19–related symptoms such as fever/cough/sore throat/loss of taste/loss of smell/fatigue/dyspnea/gastrointestinal (GI) symptoms (vomiting and diarrhea).

We also collected biochemical parameters such as C-reactive protein (CRP), D-dimer, lactate dehydrogenase (LDH), ferritin, and radiological parameter, i.e., CT severity score. All the CT images were examined and reviewed by a radiology specialist. The spectrum of CT findings consisted of ground glass opacities, consolidation, subpleural bands, vascular dilatation, architectural distortion, and crazy paving. After the baseline assessment, the patients were followed-up until their discharge to check for severity of the condition.

### Study definitions

#### Smoking and alcohol use

History of smoking and alcohol use was self-reported by the study participants.

### Olfactory-gustatory dysfunction

Patients having loss of taste and/or loss of smell during the baseline assessment were considered having olfactory-gustatory dysfunction (OGD) [9].

### High CRP

Participants with CRP level  $\geq 100$  mg/L were considered having high CRP values [10–14].

### High D-dimer

Participants with D-dimer  $\geq 1000$  ng/mL were considered having high D-dimer values [10–14].

### High ferritin

Participants with ferritin  $\geq 500$  mcg/L were considered having high ferritin values [10–14].

### High LDH

Participants with LDH  $\geq 400$  units/L were considered having high LDH values [15].

### CT severity score

The severity score was calculated based on the percentage of lung involvement. Scores were assigned based on the percentage of involvement in individual lobes and each lobe was given a score of 1 to 5:

- Score 1: < 5% involvement
- Score 2: 5–25% involvement
- Score 3: 26–50% involvement
- Score 4: 51–75% involvement
- Score 5: > 75% involvement

The final score was computed by summing the individual lobar scores; it was scored out of 25 (total score) [16]. The total lung involvement was obtained by multiplying this total score times four. Participants with CT severity score  $\geq 13$  (i.e., more than 50% lung involvement) were considered to have high CT severity score [17].

### COVID-19 severity

It was graded based on any of the following criteria: patients with respiratory rate (RR) > 30 breaths/min, oxygen saturation (SpO<sub>2</sub>) < 93%, oxygenation index (PaO<sub>2</sub>/FiO<sub>2</sub>)  $\leq 300$  mmHg, or requirement of artificial ventilation

[18]. Artificial ventilation includes the non-invasive form of mechanical ventilation, i.e., continuous positive airway pressure and/or high-flow nasal cannula and/or invasive mechanical ventilation [6].

### Statistical analysis

The completed questionnaires were entered into Microsoft Excel, and the data entry process was reviewed for errors and due correction was made by the data validation team. Descriptive analysis was performed by summarizing the continuous variables as mean and standard deviation (SD) and categorical variables as proportions. The entire dataset was split into training set and validation set at an 8:2 ratio. We developed two prediction models (basic and advanced) for identifying the patients with risk of COVID-19 severity using the training set. The basic model was built using only sociodemographic characteristics, behavioral habits, comorbidities, and COVID-19 symptoms. The advanced model was built using the baseline biochemical and radiological parameters in addition to the predictors identified in the basic model.

For developing the model, first we performed the “least absolute shrinkage and selection operator (LASSO)” regression for selection of predictor variables [19]. LASSO is a penalized regression method where the estimated coefficients are shrunken towards zero by adding a tuning parameter (lambda). Optimal value of lambda was calculated using data-driven K-fold cross-validation approach (3-folds) [19]. The largest lambda for which mean squared prediction error (MSPE) is within one standard error of the minimum loss (MSPE) was considered the optimal value of lambda ( $\lambda_{LSE}$ ). For the identified optimal lambda value, variables with non-zero coefficients were selected for the predictor model.

Then, we performed the multivariable logistic regression with the selected variables and the final model was obtained by retaining only the significant variables with the highest  $R^2$  and the least AIC and log-likelihood values. All the predictor variables were interpreted as adjusted odds ratio (aOR) with 95% confidence interval (CI). We constructed a nomogram scoring system to present the model. Receiver operator characteristic (ROC) analysis was performed to estimate the area under the curve (AUC) or c-statistic. It was used to explain the discriminatory capacity of the prediction model. We plotted calibration curve with observed frequency versus predicted probability of COVID-19 severity [20]. Hosmer–Lemeshow goodness-of-fit test was also applied to evaluate the calibration of model [21]. A  $P$  value more than 0.05 indicates that there is good model calibration. Discrimination and calibration analysis was repeated with the same set of

predictor variables in the validation set. Decision curve analysis was performed to quantify the net benefits at various threshold probabilities, which in turn determines the clinical usefulness of the final predictor model [22]. All the analysis was performed using STATA software version 14.2 (College Station, TX: StataCorp LP).

## Results

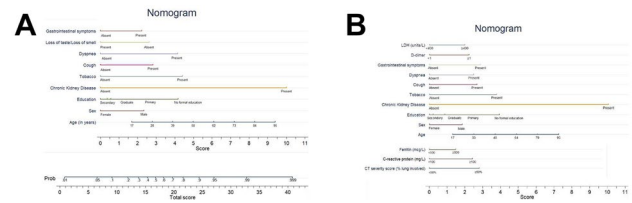
In total, 608 patients were recruited into the cohort. It was split into training set (490 patients) and validation set (118 patients). Sociodemographic characteristics of the participants in training and validation set are provided in Supplementary Table 1. In total, 35.1% of the patients in the training set developed COVID-19 severity during their follow-up period.

### Development of basic predictor model

K-fold cross-validation revealed the optimal value of lambda as 8.97 (Fig. 1A). Based on these lambda values, 14 variables were selected (age, sex, education, tobacco, alcohol, DM, HTN, CKD, CAD, fever, cough, dyspnea, OGD, and GI symptoms). Multivariable logistic regression of the final selected model is provided in Supplementary Table 2. It had the following variables: age, sex, education, CKD, tobacco, cough, dyspnea, OGD, and GI symptoms.

### Nomogram scoring system

The prediction model was presented as a nomogram (Fig. 2A), which could be used conveniently to predict COVID severity. Each of the nine variables included in the predictive models were arranged one by one on a horizontal plane with its scoring system, ranging from 0 to 10, at the bottom. The overall score can be obtained by the summation of these individual scores. Total score ranged from 0 to 42 and each of the components had the following scores (in descending order):



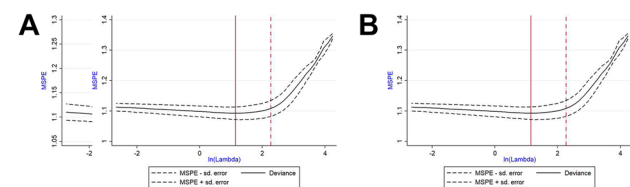
**Fig. 2** Nomogram for prediction of COVID-19 severity. **A** Basic prediction model. **B** Advanced prediction model. Prob=probability of COVID-19 severity

- I: CKD=10
- II: Age > 80 years=9.4
- III: Tobacco use=4.4
- IV: Dyspnea=4.2
- V: No formal education=4.2
- VI: Cough=2.8
- VII: OGD=2.6
- VIII: Male sex=2.3
- IX: GI symptoms=2.3

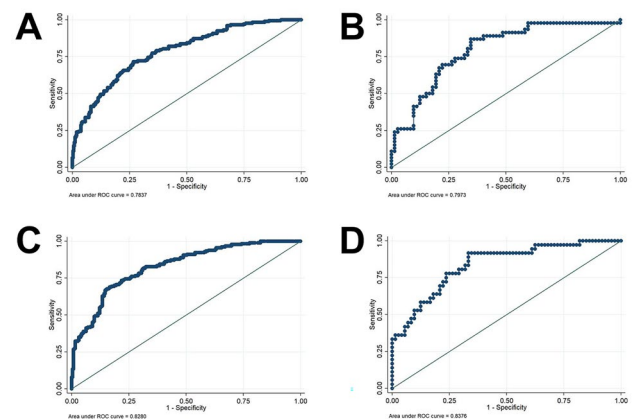
The optimal cutoff for this scoring system was 14.9. We found that patients having a score  $\geq 14.9$  at the time of diagnosis have sixfold higher risk of having COVID-19 severity when compared to those patients having a score < 14.9.

### Discrimination and calibration

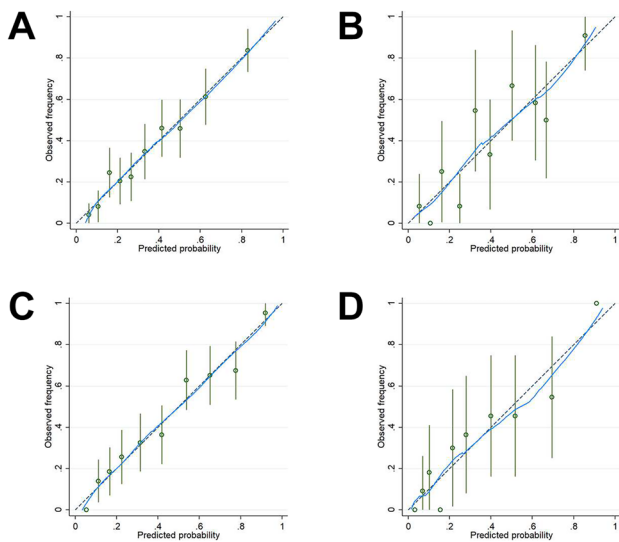
The discriminatory power (c-index) of the predictor model in the training set for COVID-19 severity was 0.78 (95%CI: 0.74–0.82) (Fig. 3A). The c-index using the validation set also revealed similar c-index (0.80) (Fig. 3B). The proposed



**Fig. 1** K fold cross-validation—MSPE versus natural log of lambda (minimum MSPE). **A** Basic prediction model. **B** Advanced prediction model. sd.error=standard error; ln(lambda)=natural logarithm of lambda; MSPE=mean squared prediction error



**Fig. 3** Discriminatory capacity (ROC curve) for the prediction of COVID-19 severity. **A** Basic prediction model—training set. **B** Basic prediction model—validation set. **C** Advanced prediction model—training set. **D** Advanced prediction model—validation set



**Fig. 4** Calibration plot for the model predicting COVID-19 severity. **A** Basic prediction model—training set. **B** Basic prediction model—validation set. **C** Advanced prediction model—training set. **D** Advanced prediction model—validation set. Blue dotted line = ideal line; blue solid line = observed; green circles = variables included in the prediction model

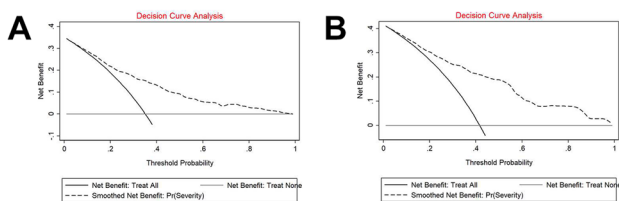
model was also well-calibrated in both training and validation as depicted by the calibration plot (Fig. 4A and B).

### Decision curve analysis

Figure 5 A shows that if the threshold probability is between 10 and 95%, application of our nomogram prediction model would add more benefit than either treat-none or treat-all strategies

### Development of advanced predictor model

For developing the advanced predictor model, we included the biochemical and radiological parameters in addition to the predictors identified in the simple model. However, data on all these variables were able among 540 patients. It was



**Fig. 5** Decision curve analysis. **A** Basic prediction model. **B** Advanced prediction model

split into training and validation set at 8:2 ratio (432 in training set and 108 in validation set).

K-fold cross-validation revealed the optimal value of lambda as 9.68 (Fig. 1B). Based on these lambda value, 13 variables were selected (age, sex, education, tobacco, CKD, cough, dyspnea, GI symptoms, CRP, LDH, ferritin, d-dimer, and CT severity score). Multivariable logistic regression was run with the selected variables, and all the variables were statistically significantly selected for final prediction model (Supplementary Table 3).

### Nomogram scoring system

The prediction model was presented as a nomogram (Fig. 2B). Each of the 13 variables included in the predictive models were arranged one by one on a horizontal plane with its scoring system, ranging from 0 to 10, at the bottom. Age, CKD, education, tobacco use, and CT severity score had the widest range of individual scores. Total score ranged from 0 to 46.6, and each of the components had the following scores (in descending order):

- I: CKD = 10
- II: Age > 80 years = 7.2
- III: No formal education = 3.9
- IV: Tobacco use = 3.8
- V: CT severity score = 3.2
- VI: CRP = 2.8
- VII: GI symptoms = 2.8
- VIII: Cough = 2.7
- IX: Dyspnea = 2.5
- X: D-dimer = 2.2
- XI: LDH = 2
- XII: Male sex = 1.8
- XIII: Ferritin = 1.7

The optimal cutoff for this scoring system was 16 points. We found that patients having a score  $\geq 16$  at the time of diagnosis have eightfold higher risk of having COVID-19 severity when compared to patients having a score < 16.

### Discrimination and calibration

The discriminatory power (c-index) of the predictor model in the training set for COVID-19 severity was 0.83 (95%CI: 0.79–0.87) (Fig. 3C). The c-index using the validation set also revealed similar value (0.84) (Fig. 3D). The proposed model was also well-calibrated in both training and validation as depicted by the calibration plot (Fig. 4C and D).

## Decision curve analysis:

Figure 5 B shows that if the threshold probability is between 10 and 100%, application of our nomogram prediction model would add more benefit than either treat-none or treat-all strategies

## Discussion

We conducted this study as a public health and clinical contribution to the low- and middle-income countries with an easy-to-apply and inexpensive prediction tool using baseline characteristics, to identify the COVID-19 severity. We have also created an advanced tool consisting of biochemical and radiological parameters applicable in high resource setting.

### Summary of findings and comparison with previous literature

In the training set, 35.1% (95%CI: 31.0–39.4%) had COVID-19 severity which is way higher than the severity burden during the first wave of pandemic [23]. While exploring the factors responsible for such fluctuations and higher burden of COVID-19 severity, one of the commonly reported factors is the wider circulation of newer COVID-19 variants such as B.1.1.7 and B.1.617 (delta variant), which are reported to have higher transmissibility and severity compared to the variants in the first wave of the epidemic [24]. It has also been reported that the delta variant increases the risk of reinfection and also breakthrough infections [25]. However, this cannot be the only reason driving this explosive surge of severe COVID-19 cases. This calls for a need to understand the predictors for severity among the newly diagnosed COVID-19 patients, so that focused interventions can be delivered.

Severity among COVID-19 patients is influenced by several sociodemographic, behavioral, clinical, biochemical, and radiological characteristics. These factors, by interacting with each other, finally affect the outcome directly or indirectly. In the basic prediction tool, age, sex, education, CKD, tobacco, cough, dyspnea, OGD, and GI symptoms had higher scores for predicting COVID-19 severity, while the advanced tool had CRP, LDH, ferritin, D-dimer, and CT severity score apart from these factors for predicting COVID-19 severity. This statement is supported by several studies done in varied study settings [6–8, 14–17, 26–30]. We also found that age, CKD, CRP, D-dimer, and CT severity score are the factors that are consistently reported across various nomogram developed in the varied study settings [6–8]. The accuracy of our basic (AUC = 0.78) and advanced prediction tool (AUC = 0.83) was also high. The

model was also well-calibrated and found to have additional clinical benefit as per decision curve analysis.

Older age group, as expected, has been one of the important risk factors for predicting the COVID-19 severity. The aging of immune system undergoes the processes such as alterations in the T-cell diversity, immunosenescence, and chronic innate system activation (also known as inflammaging) [31]. These mechanisms cripple the ability of older adults to clear the virus, initiate cytokine storms, and sustain it, leading to acute injury to major organs, disseminated intravascular coagulation, and multiorgan failure [31]. Another possible mechanism could be the age-related decline in nicotinamide adenine dinucleotide + (NAD<sup>+</sup>) results in the activation of NLRP3 (major protein component of inflammasome) in older adults, further worsening the cytokine storm [31]. Coronaviruses possess an ADP-ribosyl hydrolase that depletes the already low levels of NAD<sup>+</sup> [31]. Combination of these processes contributes to the COVID-19 severity among older adults.

Male sex was a significant predictor in both the basic and advanced model, which was in line with previous studies [32, 33]. There is lack of evidence exploring the underlying mechanisms responsible for such observed sex differences in COVID-19 severity. Initially, the reason for such disparities was thought to be due to the presence of comorbidities and excessive high-risk behaviors such as smoking and alcohol use among males [34]. However, we found that the disparities exist despite controlling for these potential confounders. Possible reasons for such finding could be the sex differences in the immune response to foreign and self-antigens with males being more susceptible to the pathogens and with females mounting strong antigenic to infections, vaccine, and self-antigen at the expense of higher burden of autoimmune diseases [35, 36]. In addition, hospitalized patients have higher level of inflammatory chemokine and cytokine, with sex differences in these immune responses [37]. In further investigations, we found that poor T-cell response and kynurenic acid changes are potential drivers for sex-specific differences in COVID-19 severity [38].

Educational status was more of a public health tool, rather than a clinical tool to identify the high-risk patients as patients with no formal education tend to ignore their symptoms and have delayed health-seeking behavior [39]. This will in turn make them more susceptible to COVID-19 severity. We were also able to identify the patients at higher risk of severity based on the presentation of symptoms. Patients presenting with symptoms such as cough, dyspnea, and GI symptoms were at significantly higher risk of developing severe COVID-19. These findings were also in line with the previous studies [40–42]. Possible reasons for identifying dyspnea as a predictor of severity could be the delayed presentation of patients ignoring the initial warning signs like fever, sore throat, or fatigue. This will attenuate the severity

making the patients present with severe symptoms like dyspnea. The pathophysiology for the correlation between GI symptoms and COVID-19 severity is not well-understood. However, it has been proposed that COVID-19 can enter into the digestive system through the ACE2 expressing cells, and the cytokine storm responsible for the severe COVID-19 infection can lead to hypoxia-induced bowel ischemic changes and promote GI symptoms like diarrhea [43].

CKD was the only comorbidity significantly predicting the COVID-19 severity. This was not an unforeseen finding, given that the recent Global Burden of Disease (GBD) collaboration study reporting that CKD is the most prevalent risk factor for COVID-19 severity worldwide [44]. Though the exact mechanism responsible for this finding is not clearly understood, several plausible theories have been proposed like low-level inflammation associated with CKD leading to baseline lymphopenia, impaired innate and adaptive immune response, and increased production and reduced clearance of proinflammatory cytokines [45, 46]. Hence, more comprehensive studies are required to clarify the causal link between CKD and COVID-19 severity. However, it is rational to warn the health authorities on these factors, so that efforts can be made to ensure proper precautionary measures taken for CKD patients.

Radiological parameters like CT severity score were also found to significantly predict the COVID-19 severity. Thus, it gives an idea on the disease prognosis and ensures early, appropriate, and optimum management of the patient, thereby reducing the hospital admission and mortality rate. We have also identified biochemical markers such as CRP, LDH, ferritin, and d-dimer as significant prediction tools for COVID-19 severity in line with the previous study findings [26–30, 47, 48]. We tried to explore the mechanisms involved in the elevation of these biomarkers during COVID-19 severity through literature review.

The unregulated and extensive production of the interleukins (particularly IL-6) stimulates many downstream pathways, increasing acute-phase reactant production like CRP [48]. The coagulation cascade involves platelets, monocytes, neutrophils, endothelial cells, and macrophages. Both anti-inflammatory and anti-thrombotic properties are necessary for healthy vascular endothelium. However, during COVID-19, this protective barrier gets disrupted leading to thrombotic and inflammatory reactions driven primarily by the thrombin [49]. Macrophages generate plasmin, which makes the fibrin degraded to D-dimers. Hence, these macrophages can contribute to the elevation of D-dimer level among severe COVID-19 patients [49]. Hence, monitoring of D-dimer values has been recommended to be a part of risk stratification criteria to make decisions on the anticoagulation therapy [50].

Ferritin might stimulate the inflammatory pathways and act as an enhancer for the inflammatory processes in severe COVID-19 patients as observed in several other inflammatory

diseases such as sepsis and macrophage activation syndrome [51, 52]. The mechanism explaining the correlation between ferritin and COVID-19 severity is considered another piece of puzzle for the hyperferritinemic syndrome [53]. Finally, LDH elevation can occur as a result of inadequate tissue perfusion and multiorgan failure due to several mechanisms (like thrombosis) associated with severe COVID-19 [54].

## Strengths and limitations

Our study is one among the very few studies that has devised a prediction model to foresee severity risk in newly diagnosed COVID-19 patients from India. The strength of our model is that many of the factors present in basic model are already being collected routinely during patient review as compared to other biochemical or radiological approaches to predict such adverse outcomes. Thus, this prediction model is a major advantage to the resource-constrained settings as it is cheap and the data is widely available. In addition, we have also developed and provided an advanced model, which can be applied in high resource settings and ensure comprehensive assessment of COVID-19 patients before advising home isolation or hospitalization.

Our study has certain limitations too. Firstly, we built this model based on the data from retrospective cohort study. Secondly, this study was conducted at a single center with a reasonable sample size. Hence, we need to externally validate this scoring system based on the data from large-scale longitudinal multicentric studies. However, our model was internally validated with good discrimination and calibration. Finally, we have excluded those patients with incomplete data on the predictor variables, which might result in selection bias.

## Implications for clinical and public health practice

The nomogram will help the clinicians in identifying the probability of severe COVID-19 from the baseline characteristics of the patient at the “detect” stage of the COVID-19 care cascade itself. The presence of certain vital sociodemographic, comorbid, behavioral, biochemical, and radiological predictors at the point of diagnosis might alarm the clinicians to have a tailored follow-up, thereby decreasing such adverse outcome later. Our basic model will be of help to the policymakers to ensure application in various public healthcare facilities in a form of checklist, to ensure that the patients are risk-stratified properly, and advise on home or hospital management is given appropriately. The developed “cost-free” basic nomogram can be tested and used in the COVID-19 fever clinics around the country for further fine-tuning of the scoring system and the predicted probabilities. The same can be used in raising awareness among the general population. It will also help the policymakers in addressing challenges with regard to the implementation of advanced nomogram with certain high-cost biochemical tests and CT scan, as it will

ensure more accurate and comprehensive assessment of COVID-19 patients.

## Conclusion

The present study has developed two separate prognostic-scoring systems to predict the COVID-19 severity. In the basic prediction tool, age, sex, education, CKD, tobacco, cough, dyspnea, OGD, and GI symptoms had higher scores for predicting COVID-19 severity, while advanced tool had CRP, LDH, ferritin, D-dimer, and CT severity score apart from these factors for predicting COVID-19 severity. This scoring system could help the clinicians and policymakers to devise targeted interventions and in turn reduce the COVID-19 mortality in India.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s11845-021-02876-w>.

**Data availability** Data will be made available upon reasonable request from researchers.

## Declarations

**Ethics approval** This study was approved by the Institutional Ethics Committee of Hindu Mission Hospital, Chengalpattu, India (IEC Approval number: HMH/IEC/2021/EA73).

**Patient consent** Not applicable.

**Conflict of interest** The authors declare no competing interests.

**Permission to reproduce material from other sources** Not applicable.

## References

- World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) [Internet] 2020 Jan 30. [cited 2021 Aug 25] Available from: [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))
- World Health Organization (2020) WHO Director-General's opening remarks at the media briefing on COVID-19. [Internet] 2020 Mar 11. [cited 2021 Aug 25] Available from: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- World Health Organization (2019) Coronavirus disease 2019 (COVID-19) Situation report – 65. [Internet] 2020. [cited 2021 Aug 25] Available from: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200325-sitrep-65-covid-19.pdf?sfvrsn=2b74edd8\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200325-sitrep-65-covid-19.pdf?sfvrsn=2b74edd8_2)
- Shang Y, Pan C, Yang X et al (2020) Management of critically ill patients with COVID-19 in ICU: statement from front-line intensive care experts in Wuhan. *China Ann Intensive Care* 10:73. <https://doi.org/10.1186/s13613-020-00689-1>
- Wu Z, McGoogan JM (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 323(13):1239–1242. <https://doi.org/10.1001/jama.2020.2648>
- Yao Z, Zheng X, Zheng Z, Wu K, Zheng J (2021) Construction and validation of a machine learning-based nomogram: a tool to predict the risk of getting severe coronavirus disease 2019 (COVID-19). *Immun Inflamm Dis* 9(2):595–607
- Yann Nguyen, Félix Corre, Vasco Honsel, Sonja Curac, Virginie Zarrouk, Catherine Paugam Burtz et al (2020) A nomogram to predict the risk of unfavourable outcome in COVID-19: a retrospective cohort of 279 hospitalized patients in Paris area. *Ann Med* 2020;52:7 367–375. <https://doi.org/10.1080/07853890.2020.1803499>
- Yang Y, Zhu XF, Huang J, Chen C, Zheng Y, He W et al (2021) Nomogram for prediction of fatal outcome in patients with severe COVID-19: a multicenter study. *Mil Med Res* 8(1):21. <https://doi.org/10.1186/s40779-021-00315-6>
- Porta-Etessam J, Núñez-Gil IJ, García NG, Fernandez-Perez C, Viana-Llamas MC, Eid CM et al (2021) COVID-19 anosmia and gustatory symptoms as a prognosis factor: a subanalysis of the HOPE COVID-19 (Health Outcome Predictive Evaluation for COVID-19) registry. *Infect* 1:1–8
- Guan WY, Ni ZY, Hu Y et al (2020) Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*
- Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *China Lancet* 395:497
- Zhou F, Yu T, Du R et al (2020) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395:1054
- Wang D, Hu B, Hu C et al (2020) Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*
- Ruan Q, Yang K, Wang W et al (2020) Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med*
- Acar HC, Can G, Karaali R, Börekçi Ş, Balkan İİ, Gemicioğlu B et al (2021) An easy-to-use nomogram for predicting in-hospital mortality risk in COVID-19: a retrospective cohort study in a university hospital. *BMC Infect Dis* 21(1):1–2
- Al-Mosawe AM, Abdulwahid HM, Fayadh NAH (2021) Spectrum of CT appearance and CT severity index of COVID-19 pulmonary infection in correlation with age, sex, and PCR test: an Iraqi experience. *Egypt J Radiol Nucl Med* 52(1):40. <https://doi.org/10.1186/s43055-021-00422-3>
- Zayed NE, Bessar MA, Lutfy S (2021) CO-RADS versus CT-SS scores in predicting severe COVID-19 patients: retrospective comparative study. *Egypt J Bronchol* 15(1):1
- Cascella M, Rajnik M, Cuomo A et al (2021) Features, evaluation, and treatment of coronavirus (COVID-19) [Updated 2021 Jan 16]. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. [Available from: <https://www.ncbi.nlm.nih.gov/books/NBK554776/>]
- Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB (2014) Validation of prediction models based on lasso regression with multiply imputed data. *BMC Med Res Methodol* 14:116
- Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 162(1):55–63



21. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S (1997) A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 16(9):965–980
22. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ et al (2018) Reporting and interpreting decision curve analysis: a guide for investigators. *Eur Urol* 74(6):796–804
23. Jain VK, Iyengar KP, Vaishya R (2021) Differences between first wave and second wave of COVID-19 in India. *Diabetes Metab Syndr* 15(3):1047–1048. <https://doi.org/10.1016/j.dsx.2021.05.009>
24. Ranjan R, Sharma A, Verma MK (2021) Characterization of the second wave of COVID-19 in India. *medRxiv*
25. Pattabiraman C, Prasad P, George AK et al (2021) Importation, circulation, and emergence of variants of SARS-CoV-2 in the South Indian State of Karnataka. *medRxiv*. <https://doi.org/10.1101/2021.03.17.21253810>
26. Zhang B, Liu Q, Zhang X, Liu S, Chen W, You J et al (2020) Clinical utility of a nomogram for predicting 30-days poor outcome in hospitalized patients with COVID-19: multicenter external validation and decision curve analysis. *Front Med* 7:951
27. Henry BM, Aggarwal G, Wong J, Benoit S, Vikse J, Plebani M et al (2020) Lactate dehydrogenase levels predict coronavirus disease 2019 (COVID-19) severity and mortality: a pooled analysis. *Am J Emerg Med* 38(9):1722–1726
28. Xue G, Gan X, Wu Z, Xie D, Xiong Y, Hua L et al (2020) Novel serological biomarkers for inflammation in predicting disease severity in patients with COVID-19. *Int Immunopharmacol* 89:107065
29. Lin Z, Long F, Yang Y, Chen X, Xu L, Yang M (2020) Serum ferritin as an independent risk factor for severity in COVID-19 patients. *J Infect* 81(4):647–679
30. Lippi G, Favaloro EJ (2020) D-dimer is associated with severity of coronavirus disease 2019: a pooled analysis. *Thromb Haemost* 120(05):876–878
31. Mueller AL, McNamara MS, Sinclair DA (2020) Why does COVID-19 disproportionately affect older people? *Aging* 12(10):9959–9981. <https://doi.org/10.18632/aging.103344>
32. Vahidy FS, Pan AP, Ahnstedt H, Munshi Y, Choi HA, Tiruneh Y et al (2021) Sex differences in susceptibility, severity, and outcomes of coronavirus disease 2019: cross-sectional analysis from a diverse US metropolitan area. *PLoS One* 16(1):e0245556
33. Peckham H, de Grujter NM, Raine C et al (2020) Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ICU admission. *Nat Commun* 11:6317. <https://doi.org/10.1038/s41467-020-19741-6>
34. Global Health 50/50 (2020) The sex, gender and Covid-19 project. 2020 [cited 3 Oct 2020]. Available from: <https://globalhealth5050.org/covid19>
35. Klein SL, Flanagan KL (2016) Sex differences in immune responses. *Nat Rev Immunol* 16:626–638
36. Ahnstedt H, McCullough LD (2019) The impact of sex and age on T cell immunity and ischemic stroke outcomes. *Cell Immunol* 345:103960
37. Yale IMPACT Research Team, Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C et al (2020) Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature*
38. Cai Y, Kim DJ, Takahashi T, Broadhurst DI, Ma S, Rattray NJW et al (2020) Kynurenic acid underlies sex-specific immune responses to COVID-19. *medRxiv* [Preprint]. 2020. Available from: <https://www.medrxiv.org/content/https://doi.org/10.1101/2020.09.06.20189159v2>
39. Nayak N, Varambally KV (2017) Impact of autonomy on health-seeking behaviour: evidence from rural India. *J Health Manag* 19(1):109–120
40. Zhou Y, He Y, Yang H, Yu H, Wang T, Chen Z et al (2020) Development and validation a nomogram for predicting the risk of severe COVID-19: a multi-center study in Sichuan, China. *PLoS One* 15(5):e0233328
41. Shi L, Wang Y, Wang Y, Duan G, Yang H (2020) Dyspnea rather than fever is a risk factor for predicting mortality in patients with COVID-19. *J Infect* 81(4):647–679. <https://doi.org/10.1016/j.jinf.2020.05.013>
42. Livanos AE, Jha D, Cossarini F, Gonzalez-Reiche AS, Tokuyama M, Aydillo T et al (2020) Gastrointestinal involvement attenuates COVID-19 severity and mortality. *MedRxiv*
43. Ramachandran P, Onukogu I, Ghanta S et al (2020) Gastrointestinal symptoms and outcomes in hospitalized Coronavirus disease 2019 patients. *Dig Dis* 38(5):373–379. <https://doi.org/10.1159/000509774>
44. Clark A, Jit M, Warren-Gash C et al (2020) Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *Lancet Glob Health* 8:e1003–e1017
45. Fernández-Fresnedo G, Ramos MA, González-Pardo MC, de Francisco AL, López-Hoyos M (2000) Arias M : B lymphopenia in uremia is related to an accelerated in vitro apoptosis and dysregulation of Bcl-2. *Nephrol Dial Transplant* 15:502–510
46. Syed-Ahmed M, Narayanan M (2019) Immune dysfunction and risk of infection in chronic kidney disease. *Adv Chronic Kidney Dis* 26(1):8–15. <https://doi.org/10.1053/j.ackd.2019.01.004>
47. Huang I, Pranata R, Lim MA, Oehadian A, Alisjahbana B (2020) C-reactive protein, procalcitonin, D-dimer, and ferritin in severe coronavirus disease-2019: a meta-analysis. *Ther Adv Respir Dis* 14:1753466620937175. <https://doi.org/10.1177/1753466620937175>
48. Samprathi M, Jayashree M (2020) Biomarkers in COVID-19: an up-to-date review. *Front Pediatr* 8
49. Mitchell WB (2020) Thromboinflammation in COVID-19 acute lung injury. *Paediatr Respir Rev* 35:20–24. <https://doi.org/10.1016/j.prrv.2020.06.004>
50. Aggarwal M, Dass J, Mahapatra M (2020) Hemostatic abnormalities in COVID-19: an update. *Indian J Hematol Blood Transfus* 36:1–11. <https://doi.org/10.1007/s12288-020-01328-2>
51. Kappert K, Jahić A, Tauber R (2020) Assessment of serum ferritin as a biomarker in COVID-19: bystander or participant? Insights by comparison with other infectious and non-infectious diseases. *Biomarkers* 25(8):616–625
52. Ruscitti P, Berardicurti O, Di Benedetto P, Cipriani P, Iagnocco A, Shoenfeld Y et al (2020) Severe COVID-19, another piece in the puzzle of the hyperferritinemic syndrome. An immunomodulatory perspective to alleviate the storm. *Front Immunol* 11:1130
53. Ruscitti P, Berardicurti O, Barile A, Cipriani P, Shoenfeld Y, Iagnocco A et al (2020) Severe COVID-19 and related hyperferritinaemia: more than an innocent bystander? *Ann Rheumatic Dis* 79(11):1515–1516
54. Erez A, Shental O, Tchebiner JZ et al (2014) Diagnostic and prognostic value of very high serum lactate dehydrogenase in admitted medical patients. *Isr Med Assoc J* 16:439–443