



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Original Article/Research

## Personalized stratification of hospitalization risk amidst COVID-19: A machine learning approach

Carson Lam, Jacob Calvert, Anna Siefkas, Gina Barnes, Emily Pellegrini, Abigail Green-Saxena<sup>\*</sup>, Jana Hoffman, Qingqing Mao, Ritankar Das

*Dascena, Inc., Houston, TX, United States*



## ARTICLE INFO

## Keywords:

Machine learning  
Algorithm  
COVID-19  
Prediction

## ABSTRACT

**Objective:** In the wake of COVID-19, the United States (U.S.) developed a three stage plan to outline the parameters to determine when states may reopen businesses and ease travel restrictions. The guidelines also identify subpopulations of Americans deemed to be at high risk for severe disease should they contract COVID-19. These guidelines were based on population level demographics, rather than individual-level risk factors. As such, they may misidentify individuals at high risk for severe illness, and may therefore be of limited use in decisions surrounding resource allocation to vulnerable populations. The objective of this study was to evaluate a machine learning algorithm for prediction of serious illness due to COVID-19 using inpatient data collected from electronic health records.

**Methods:** The algorithm was trained to identify patients for whom a diagnosis of COVID-19 was likely to result in hospitalization, and compared against four U.S. policy-based criteria: *age over 65; having a serious underlying health condition; age over 65 or having a serious underlying health condition; and age over 65 and having a serious underlying health condition.*

**Results:** This algorithm identified 80% of patients at risk for hospitalization due to COVID-19, versus 62% identified by government guidelines. The algorithm also achieved a high specificity of 95%, outperforming government guidelines.

**Conclusions:** This algorithm may identify individuals likely to require hospitalization should they contract COVID-19. This information may be useful to guide vaccine distribution, anticipate hospital resource needs, and assist health care policymakers to make care decisions in a more principled manner.

## Public interest summary

A machine learning algorithm was developed and evaluated for the prediction of serious illness due to COVID-19 using inpatient data collected from electronic health records. The algorithm identified 80% of patients at risk for hospitalization due to COVID-19, versus, at most, 62% that were identified by government guidelines. These results support the use of a machine learning algorithm for prediction of serious illness, and may help to predict use of hospital resources, guide vaccine distribution, and assist health care policymakers in making informed care decisions throughout the remainder of the COVID-19 pandemic.

## Introduction

SARS-CoV-2 continues to spread across the United States (U.S.).

Outbreak of the virus has resulted in numerous state and national protocols intended to contain its spread [1,2]. Despite these efforts, concerns remain about future COVID-19 surges. Researchers have noted the necessity of anticipating and minimizing strain on hospital resources through the remainder of the pandemic [3,4]. Additionally, as vaccines begin to become available, appropriately allocating vaccine doses to those most in need will be important for preventing future surges in hospitalization [5,6] and allowing for a full reopening of U.S. businesses [7]. To address these needs, it is imperative to understand how many—and which—individuals are at risk of significant complications should they become infected.

Despite the continued threat of COVID-19, local, state, and federal governments have faced pressure to kick-start the economy and reduce the economic burden of unemployment and underemployment. This pressure emerges from the significant economic impact of the pandemic.

<sup>\*</sup> Corresponding author: 12333 Sowden Rd Suite B PMB 65148, Houston, TX 77080, United States.

E-mail address: [abigail@dascena.com](mailto:abigail@dascena.com) (A. Green-Saxena).

<https://doi.org/10.1016/j.hlpt.2021.100554>

Available online 4 August 2021

2211-8837/© 2021 Fellowship of Postgraduate Medicine. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

More than \$16 trillion estimated costs are directly associated with COVID-19 in the United States [8] and over \$60 million Americans have filed for unemployment to date [8]. To determine when the community, as a whole, should commit to staying at home and when activities outside the home are safe, governments have implemented guidelines based on the Centers for Disease Control and Prevention (CDC)-guided action plan entitled *Opening Up America Again (OUAA)* [9]. This three-phase strategy identifies the minimum requirements that states must meet prior to easing restrictions, as well as the individual- and business-level practices recommended for limiting disease spread. Progression through each phase is data-driven by state reported incidences of three criteria: 1) symptomatology of individuals, 2) number of confirmed COVID-19 cases reported, and 3) the ability of clinical sites to provide sufficient COVID-19 testing. These criteria are intended to provide projections for frontline workers and to aid patient care management without patient health declining to the point of requiring crisis care. In all three phases, vulnerable populations must continue to practice safe distancing.

The OUAA plan identifies vulnerable populations as the elderly and those with serious underlying health conditions [9].

However, this macro-level framework does not consider individual-level health data to determine the risk that individuals face if COVID-19 is contracted. Per an estimate by the CDC, 21% to 31% of people who have tested positive for COVID-19 have severe reactions that require hospitalization [10]. Adults over the age of 65 constitute the largest population to require hospitalization due to COVID-19, with estimated age-specific hospitalization rates between 28% and 70% [10]. However, individuals younger than 65 have also experienced severe disease outcomes and hospitalization without serious underlying health conditions, and some individuals over the age of 65 do not have severe disease when they contract COVID-19 [10]. Guidelines, such as those presented by the CDC, are therefore likely to be insufficient for determining individual risk of hospitalization, for accurately predicting hospital resource needs throughout the pandemic, or for prioritizing individuals for early receipt of COVID-19 vaccination. Reliance on these demographics-based criteria may leave both governments and individuals with an inadequate estimate of likely hospitalization needs.

The study presented here used machine learning (ML) technology to identify an individual's risk of hospitalization because of high physiological severity of COVID-19 symptoms. This algorithm has the potential to support clinical recommendations, guide vaccine distribution, and estimate hospital resource use by predicting which patients will experience severe outcomes from COVID-19 using existing and readily accessible patient data.

## Methods

### Patient selection

Patients were selected from five community hospitals in the United States. Selection began with 2,412 patients who were tested for COVID-19 from January 1, 2020. Out of these patients, 289 had a prior inpatient or emergency department (ED) encounter within the past 24 months. Among these patients with prior encounters, 188 were hospitalized at their return encounter for COVID-19. Hospitalization was determined by the transfer to a floor bed, Intensive Care Unit (ICU) bed, or unit other than the ED in the patient record and having more than 24 hours of patient data, such as vital signs or laboratory tests. The remaining 101 COVID-19 positive patients were not hospitalized after evaluation. Supplementary Figure 1 depicts the patient inclusion criteria. Using only information collected at the previous hospital visit, an algorithm was trained to predict whether patients who tested positive for COVID-19 at a subsequent visit would go on to require hospitalization for COVID-19.

### Data processing and measures

Health information was obtained from patient Electronic Health Records (EHRs). Features used from the initial visit for the algorithm-based prediction included: age, gender, diabetes diagnosis, hypertension diagnosis, respiratory diagnosis, cardiovascular diagnosis, chemotherapy, cancer diagnosis, obesity, systolic and diastolic blood pressure, respiratory rate, heart rate, temperature, oxygen saturation (SpO<sub>2</sub>), glucose, international normalized ratio (INR), lactate, blood urea nitrogen (BUN), creatinine, bilirubin, calcium, white blood cell (WBC), platelet count, red blood cell (RBC), red blood cell distribution width (RDW), hemoglobin, hematocrit, mean corpuscular volume (MCV), mean corpuscular hemoglobin concentration (MCHC), counts of lymphocytes, monocytes, eosinophils, basophils and neutrophils. These features served as the input variables for the algorithm. Though not all features were available for all patients, as described in later sections and the supplement, the algorithm effectively compensates for missing inputs.

For each patient included in the analysis (all of whom had an encounter in which they tested positive for COVID-19 and a prior encounter), features were constructed from data from the previous two hours of the prior encounter. We binned the measurements by averaging all measurements taken in the last hour of the prior encounter and averaging all measurements taken from the second to last hour of the prior encounter. The results from this binning process were two features from each input variable for each encounter. We further incorporated information about the trends in each measurement by calculating differences between these two bins.

Using information from the initial non-COVID related encounter, we predicted whether each patient would be hospitalized upon testing positive for COVID-19 at a return visit. Hospitalization was defined as transfer from the ED, ICU or inpatient floor during the same encounter, with length of stay greater than 24 hours. Encounters with only the ED as a location, or patient data less than 24 hours, did not meet this definition of being hospitalized. The classifier that made these predictions was created using the XGBoost method for fitting gradient boosted decision trees. Further technical specifications on the gradient boosted decision tree classifier design are described in the Supplementary Materials.

### Statistical analysis

Algorithm performance was compared to four OUAA policy-based criteria that identifies vulnerable individuals: *age over 65; having a serious underlying health condition; age over 65 or having a serious underlying health condition; and age over 65 and having a serious underlying health condition*. Serious underlying health conditions were identified by International Classification of Diseases (ICD)9 or ICD-10 codes, and were defined as any of a respiratory diagnosis (e.g. chronic obstructive pulmonary disease), a cardiovascular diagnosis (e.g. heart failure), diabetes, high blood pressure, a cancer diagnosis, immunosuppressed status (e.g. concurrent chemotherapy), or obesity.

All algorithm results were obtained on a hold-out test set not seen by the model during the training process. The algorithm was evaluated using the area under the receiver operator characteristic (AUROC), sensitivity, specificity, F1, diagnostic odds ratio (DOR), and positive and negative likelihood ratios (LR+ and LR-). The algorithm predicted whether or not a patient was likely to be hospitalized upon testing positive for COVID-19 in a subsequent visit, or not likely to be hospitalized. For each policybased criteria, patients either met the criteria, and thus were predicted to be hospitalized upon testing positive for COVID-19 in a subsequent visit, or do not meet the criteria.

To estimate the stay at home rate in the general population for the age-based policy, we first identified the proportion of the population that is elderly based on census data, which indicates that 16% of the population is over age 65 [11]. By that estimate, 16% of the population should continue to stay at home per the *age over 65* OUAA-based criteria.

We then analyzed our dataset and determined that 38% of these patients were over age 65. To normalize our estimates against the general population, we calculated a “normalization factor” by dividing 38% by 16%, which is 2.375. This “normalization factor” is our estimate of how biased our dataset was towards a sickly population. The “normalization factor” indicates that our dataset had 2.375 times as many patients likely to have complications should they contract COVID-19 as compared to the general population. We calculated the proportion of patients who met the various OUAA policies and determined estimated stay at home rates by scaling the calculated rates in our dataset by this “normalization factor”.

To estimate a stay at home rate for the algorithm, we applied the algorithm to a randomly sampled selection of patients who had been hospitalized at these five sites in the last 24 months. We scaled the proportion of patients predicted to be hospitalized should they later contract COVID-19 by a factor of 0.079 to obtain the estimated stay at home rate. This scaling factor was used because U.S. Census data has found that approximately 7.9% of the general U.S. population has been hospitalized [12].

Feature importance was assessed with SHapley Additive exPlanations (SHAP). Higher absolute SHAP values were associated with features that had a larger impact on model prediction scores. Additionally, we assessed the distribution of feature values, as well as correlation of feature values with model predictions.

## Results

In total, we included 289 patients who received a positive COVID-19 diagnosis between 1/1/2020 and 5/7/2020. Of those patients 188 (65%) were hospitalized after their initial COVID-19 evaluation. Those who were hospitalized for COVID-19 were likely to be older and slightly more likely to be male as compared to those who were not hospitalized. Patient demographic characteristics are presented in Table 1.

The algorithm achieved an AUROC of 0.88 (95% Confidence Interval: 0.79 - 0.97) for predicting hospitalization among COVID-19 patients based on data from a prior, non-COVID-19 related inpatient encounter. At all operating points, the algorithm provided higher sensitivity and specificity for hospitalization prediction than any of the OUAA policy-based criteria (*over age 65; presence of serious underlying health conditions; either condition; and both conditions*) (Fig. 1). Sensitivity was extremely low (< 0.20) for both the *serious underlying health conditions* and those *over age 65 and with a serious underlying health conditions* criteria.

The algorithm demonstrated improved performance across all measured test characteristics as compared to the policy-based criteria. Based on our estimated stay at home rates, the OUAA criterium of *age over 65 or having a serious underlying health condition* would designate 20% of the general population as high-risk for hospitalization due to COVID-19, with a sensitivity and specificity of 62% and 79%, respectively. The machine learning algorithm (MLA) would designate a smaller number of the general population as high-risk for hospitalization (4.3%), with an improved sensitivity and specificity of 80% and 95%, respectively (Table 2).

**Table 1**

Demographic characteristics of COVID-19 positive patients included in the analysis.

		Covid+: Non-hospitalized N (%)	Covid+: Hospitalized N (%)
Age	<30	16 (15.84)	11 (5.85)
	30-49	27 (26.73)	21 (11.17)
	50-59	21 (20.79)	32 (17.02)
	60-69	16 (15.84)	33 (17.55)
	70-79	10 (9.9)	46 (24.47)
	80+	11 (10.89)	45 (23.94)
Gender	Female	55 (54.5)	92 (48.9)

An analysis of feature importance and correlations indicated that the model relied heavily on patient age and information related to cardio-pulmonary functioning to generate predictions. Features showed a wide range of distributions and correlations with model predictions (Fig. 2a). The correlation analysis showed that higher age and serum glucose, as well as lower heart rate and mean corpuscular volume, were all associated with increased model prediction scores. An examination of model feature importance found that age, SpO<sub>2</sub>, and heart rate made the strongest contributions to model predictions (Fig. 2b), with age making the strongest contribution of any model feature. Fig. 2c presents an individual-level breakdown of how these model features cause changes in an individual's risk prediction score for three example patients, illustrating the risk prediction procedure by identifying which feature values had strong impacts on the final risk prediction score by shifting it higher or lower.

## Discussion

This study evaluates an MLA designed to identify patients at high risk of hospitalization due to COVID-19. Of particular importance is the potential for this algorithm to identify patients who should not return to work due to risk for severe COVID-19 symptoms with better accuracy (85%, see Table 2) than the OUAA guideline accuracy (60% for individuals over age 65 or 67% for individuals over 65 or with a serious underlying health condition). Further, results from the algorithm estimated that 4.3% of the general population is at risk of hospitalization due to COVID-19 versus the OUAA's suggested 16% for those aged over 65, and approximately 20% for those aged over 65 or those who have serious underlying health conditions such as a respiratory diagnosis, cardiovascular diagnosis, diabetes, high blood pressure, cancer diagnosis, concurrent chemotherapy, or obesity. Although recent reported rates of hospitalization due to COVID-19 are higher than our 4.3% estimate, reported rates should be interpreted with caution. Testing shortages and the presence of asymptomatic cases have likely led to an undercount of COVID-19 cases. Therefore, the true hospitalization rate is likely lower than commonly reported estimates.

For all patients, retrospective baseline data was collected at partner sites for patients who had been hospitalized within the past two years and again upon a second hospitalization, during which time the patients were positively diagnosed with COVID-19. The algorithm had higher sensitivity and specificity than the OUAA criteria; the algorithm had 80% sensitivity and 95% specificity. In contrast, OUAA criteria had <50% sensitivity for those aged over 65 and 62% for those aged over 65 or with a serious underlying health condition. The overall accuracy of the algorithm was 85% for versus 67% for the best OUAA criteria. The DOR of the algorithm was meaningfully higher than the OUAA DOR for age over 65, or age over 65 or serious underlying health condition, at 69.8 versus 5.1 and 6.0, respectively. These results provide initial evidence that an algorithm like the one presented in this study may be able to complement existing guidelines and provide more accurate individual-level risk assessment.

An analysis of algorithm model feature importance indicated that age, SpO<sub>2</sub>, and heart rate made the strongest contributions to model predictions. Many of the features identified as important for model prediction correspond with what is already known about the disease. For example, in our data, increased age is associated with increased risk of future hospitalization due to COVID-19.

Prior research utilizing MLAs to predict the risk of symptom escalation and identify populations at risk of hospitalization is sparse, despite their potential utility for identifying such patients for purposes of early vaccination and understanding likely hospital resource requirements [13]. A study by Yan et al. used ML to identify biomarkers associated with COVID-19 death in hospitalized patients [14]. While this may help to guide resource allocation once patients are hospitalized, this system does not anticipate likely COVID-19 severity in currently uninfected patients. Similarly, a study by Assaf et al. used ML to predict

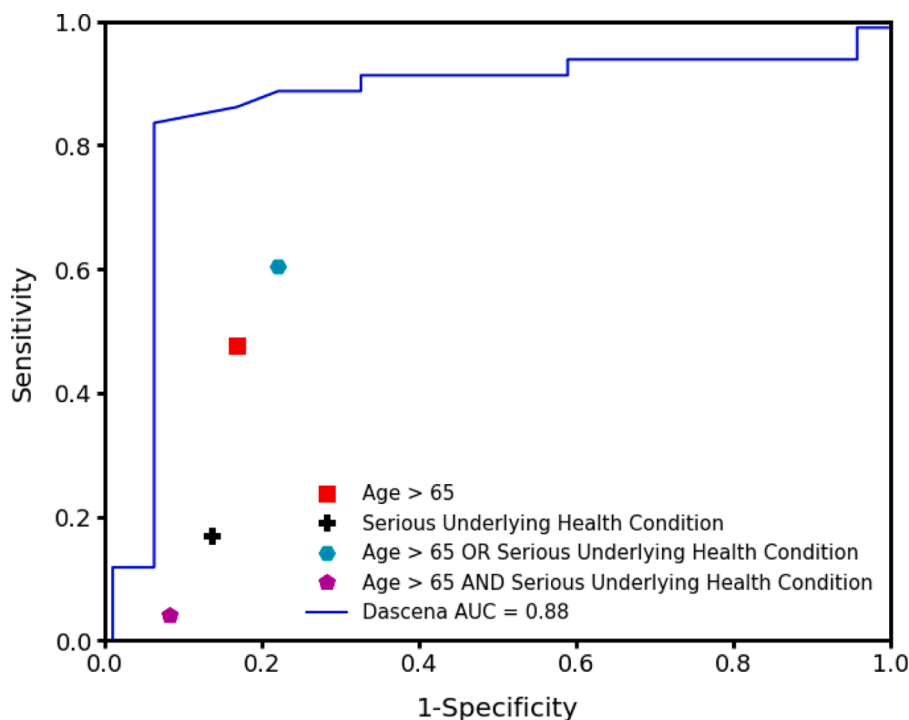


Fig. 1. Comparison of receiver operating characteristic (ROC) curve for the algorithm to the sensitivity and specificity of the operating points for the Age > 65, Age > 65 AND/OR Serious Underlying Health Condition, and Serious Underlying Health Condition policies.

**Table 2**  
Performance metrics for the machine learning algorithm (MLA) as compared to OUA policy-based criteria

	MLA	Age > 65	Serious Underlying Health Condition	Age > 65 OR Serious Underlying Health Condition	Age > 65 AND Serious Underlying Health Condition
Sensitivity	80% (67% - 92%)	49% (33% - 64%)	18% (6% - 30%)	62% (46% - 77%)	5% (0% - 12%)
Specificity	95% (85% - 100%)	84% (68% - 100%)	89% (76% - 100%)	79% (61% - 97%)	95% (85% - 100%)
Accuracy	85%	60%	41%	67%	34%
Estimated stay at home rate	4.3%	16%	6.7%	20%	2.1%
DOR	69.8	5.1	1.9	6.0	0.97
LR+	15.1	3.1	1.7	2.9	0.97
LR-	0.22	0.61	0.92	0.49	1.0
F1 score	0.87	0.62	0.29	0.72	0.095

DOR, diagnostic odds ratio; LR, likelihood ratio; MLA, machine learning algorithm.

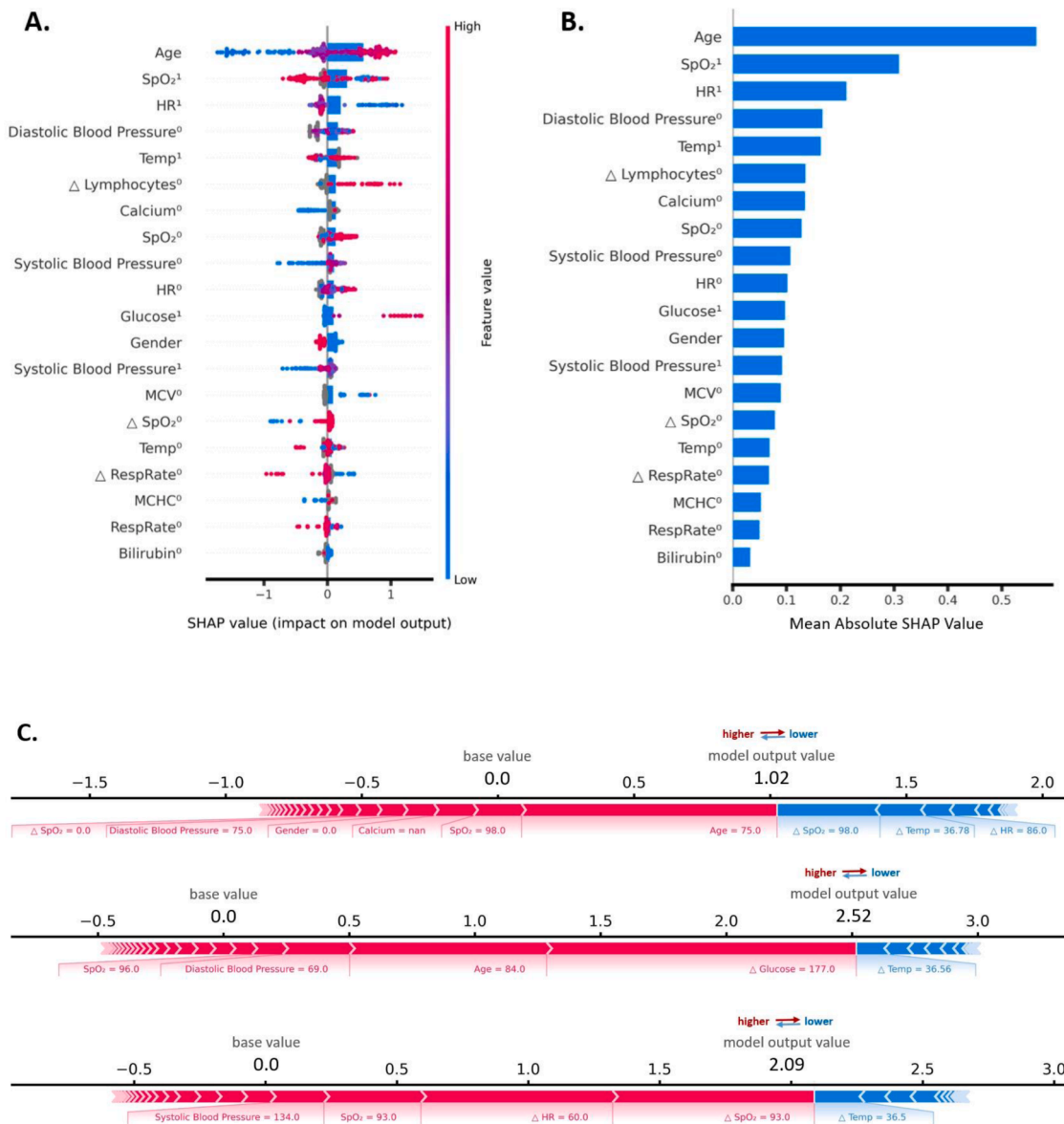
critical illness in hospitalized COVID-19 patients, but did not predict disease severity before the point of infection and hospitalization [15]. These systems therefore cannot be used to prioritize vaccine distribution or anticipate future hospital resource needs.

A small amount of prior literature exists for predicting more general disease severity in COVID-19 patients. A study by Decaprio et al. extracted data from the Medicare claims of 369,865 individuals to predict the potential for COVID-19 symptoms to result in development of comorbidities, such as bronchitis and pneumonia [16]. Their model yielded a 23.4% sensitivity rate. Because this data was selected from Medicare claims and represents a subset of the population living within the Federal Poverty Level (FPL) [17], this population may have a greater burden of preexisting health conditions that arise from health disparities

which include poverty [18]. The preexisting serious underlying health conditions within this population may result in increased severe outcomes with COVID-19. Data used to predict the potential of severe COVID-19 outcomes were not derived from the COVID-19 positive subset of the population, making it difficult to accurately assess model validity [16]. A second study in China conducted by Jiang et al. examined the efficacy of a predictive algorithm to identify COVID-19 positive patients' likelihood of developing pneumonia as a comorbidity. This predictive tool showed 70%-81% accuracy with a small sample of 53 [19]. Our research builds upon this study, taking advantage of a larger sample size to show similarly strong results for predicting hospitalization of COVID-19 patients.

There are several limitations to this study. Our algorithm identified patients who should consider remaining at home as the United States progresses through stages of reopening due to high risk of hospitalization. However, the model was developed using data from patients with a prior hospitalization within the last two years. Therefore, we cannot determine the extent to which the algorithm is capable of making accurate predictions on individuals with a prior hospitalization beyond the last two years, or on individuals with no record of prior hospitalization. Future studies utilizing more distant data or outpatient information would improve model generalizability and improve the ability of the algorithm to generate predictions on diverse populations. It is also possible that patients included in the study were hospitalized after the conclusion of the study period and their hospitalizations were not captured in our data. This may lead to some degree of misclassification of the outcome. While the most important features such as age and oxygen saturation are clinically intuitive, some features are more complex, such as change in lymphocytes. While this may represent an achievement in the algorithm's ability to find subtle correlations between multitudes of features, the same ability can lead MLAs to overfit. These findings may add to the candidate pool of potential mechanisms of disease to explore in the future. While the population included in this study was obtained from several institutions, clustering of patients within institutions may have influenced study results. The institutions included also do not form a nationally representative sample and may





**Fig. 2.** A) Feature correlations and distribution of feature importance for each patient. Input variables are ranked in descending order of feature importance. Red indicates a high feature value; blue indicates a low feature value. Dots to the right resulted in a higher score; dots to the left resulted in a lower score. The superscript denotes the number of hours prior to the time the algorithm was applied, and  $\Delta$  denotes change from the previous hour of measurement. For example,  $\Delta SpO_2^0$  represents the change in oxygen saturation from the previous hour to the current hour. B) Global feature importance reported by SHAP (SHapley Additive exPlanations). C) Feature Importance for Three Example Individuals. The Model Output Value is the algorithm score for these three individual example patients. The red colors indicate that a feature pushed the output to a higher score, and blue colored features reduced the output score.

limit generalizability of these results. Additionally, algorithm performance was assessed on retrospective data. Prospective validation on a diverse and nationally representative cohort of as-yet uninfected individuals is necessary for any claims of predictive performance in live settings. As the SARS-CoV-2 virus evolves over time, continued validation of the accuracy of the algorithm may be warranted to ensure that its use remains appropriate. Although the algorithm could be used to assist health care providers and policymakers in making more principled care decisions, we stress that the tool is intended to support, not to replace, clinical judgement and CDC guidance.

The global effort to stem the spread and impact of COVID-19 requires the use of cross-disciplinary ingenuity and tools that can adapt rapidly to meet a variety of symptoms and levels of disease severity. MLAs have the capability to analyze large amounts of patient data from EHRs. This algorithm identifying patients most at risk of hospitalization from COVID-19 can be quickly and simply implemented in any EHR and is

non-invasive. While prospective validation is warranted before implementation, these retrospective results support that the algorithm may serve as a powerful tool to supplement clinical evaluations of COVID-19 positive patients. These traits are especially valuable as hospitals continue to experience significant COVID-19 positive patient admissions [20]. Further, the algorithm emphasizes the benefit from incorporating individualized factors based on specific patient characteristics into models that will be used to guide the public about safe practices as stay at home orders are lifted. Beyond its immediate public health implications, the use of such a tool in these settings may serve to improve trust in ML and artificial intelligence (AI) applications to medicine. Despite the potential of AI to improve clinical practice and healthcare delivery, and despite FDA approval of numerous machine learning devices [21], trust in AI remains an important barrier to implementation [22]. Developing effective algorithms for use in public health settings and that complement established guidelines, such as the OUAA guidelines, may

complement efforts to develop explainable AI [23] and help to bolster support for such algorithms in both current and future infectious outbreaks.

## Conclusions

We have developed and evaluated an algorithm using EHR data from prior patient hospital visits that can accurately predict the likelihood of future hospitalization from COVID-19 complications. This algorithm demonstrates higher specificity, sensitivity, and accuracy than broad population-level categorization. This algorithm may serve as a valuable tool to assist clinical evaluations of COVID-19 positive patients and guide clinical decision making. The present study provides proof of concept that machine learning methods may be useful for understanding individual level risk in future infectious outbreaks. Such methods have the potential to complement traditional risk stratification tools and assist with resource allocation, disease prevention, and maintenance of important economic activity.

## Author statements

There is no funding to report for this work. This study was approved by the Pearl Institutional Review Board with a waiver of informed consent under study number 20-DASC-122. All authors who have affiliations listed with Dascena (Houston, Texas, USA) are employees or contractors of Dascena.

## Author roles

CL contributed to the data analysis of this work; JC contributed to the conception and drafting of this work; GB, EP, AS, AGS, and JH contributed to the drafting and revision of this work; and RD and QM contributed to the conception and revision of this work.

**Financial Disclosures:** Ritankar Das owns stock in Dascena. All other authors have no financial disclosures to report.

**Competing Interests:** All authors who have affiliations listed with Dascena (San Francisco, California, USA) are employees or contractors of Dascena.

**Funding:** None

**Ethical Approval:** Study 20-DASC-122 approved by Pearl Institutional Review Board.

This study was considered to be of minimal risk for human subjects and was approved by the Pearl Institutional Review Board with a waiver of informed consent under study number 20-DASC-122.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.hlpt.2021.100554](https://doi.org/10.1016/j.hlpt.2021.100554).

## References

- [1] Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19): Situation Summary [Internet]. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/summary.html>.
- [2] Mortality Analyses [Internet]. Johns Hopkins Coronavirus Resource Center. [cited 2021 Jan 11]. Available from: <https://coronavirus.jhu.edu/data/mortality>.
- [3] CDC. Coronavirus Disease. (COVID-19) [Internet]. Centers for Disease Control and Prevention. 2019 [cited 2021 Jan 11]. Available from, <https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-hcf.html>.
- [4] Caring with Limited Resources During COVID-19 | AHA [Internet]. [cited 2021 Jan 11]. Available from: <https://www.aha.org/clinical-resources-covid-19>.
- [5] COVID-19 Vaccination Program Operational Guidance | CDC [Internet]. 2021 [cited 2021 Jan 11]. Available from: <https://www.cdc.gov/vaccines/covid-19/covid19-vaccination-guidance.html>.
- [6] Iboi EA, Ngonghala CN, Gumel AB. Will an imperfect vaccine curtail the COVID-19 pandemic in the U.S.? *Infect Dis Model* 2020;5:510–24.
- [7] Janiak A, Machado C, Turén J. Covid-19 contagion, economic activity and business reopening protocols. *J Econ Behav Organ* 2021;182:264–84. Feb.
- [8] Cutler DM, Summers LH. The COVID-19 Pandemic and the \$16 Trillion Virus. *JAMA* 2020 Oct 20;324(15):1495–6.
- [9] Opening Up America Again [Internet]. The White House. [cited 2021 Jan 11]. Available from: <https://trumpwhitehouse.archives.gov/openingamerica/>.
- [10] CDCMMWR. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep* [Internet]. 2020 [cited 2021 Jan 11];69. Available from: <https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm>.
- [11] U.S. Census Bureau QuickFacts: United States [Internet]. [cited 2021 Jan 11]. Available from: <https://www.census.gov/quickfacts/fact/table/US/PST045219>.
- [12] FastStats - Hospital Utilization [Internet]. 2020 [cited 2021 Jan 11]. Available from: <https://www.cdc.gov/nchs/fastats/hospital.htm>.
- [13] Naudé W. Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. *Ai Soc*; 2020. p. 1–5. Apr 28.
- [14] Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *medRxiv* 2020. Mar 17;2020.02.27.20028027.
- [15] Assaf D, Gutman Y, Neuman Y, Segal G, Amit S, Gefen-Halevi S, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med* 2020;15(8):1435–43. Nov.
- [16] Building a COVID-19 vulnerability index - DeCaprio - Journal of Medical Artificial Intelligence [Internet]. [cited 2021 Jan 11]. Available from: <http://jmai.amegroups.com/article/view/5930/html>.
- [17] Federal Poverty Level (FPL) - HealthCare.gov Glossary [Internet]. HealthCare.gov. [cited 2021 Jan 11]. Available from: <https://www.healthcare.gov/glossary/federal-poverty-level-fpl/>.
- [18] 2020 Topics and Objectives – Objectives A–Z | Healthy People 2020 [Internet]. [cited 2021 Jan 11]. Available from: <https://www.healthypeople.gov/2020/top-ics-objectives>.
- [19] Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, et al. Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. *Comput Mater Contin* 2020;62(3):537–51.
- [20] COVID-19 Hospitalizations [Internet]. [cited 2021 Jan 11]. Available from: [https://gis.cdc.gov/grasp/covidnet/COVID19\\_3.html](https://gis.cdc.gov/grasp/covidnet/COVID19_3.html).....
- [21] Benjamens S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *Npj Digit Med* 2020 Sep 11;3(1):1–8.
- [22] Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *J Med Internet Res* [Internet] 2020;22(6). Jun 19 [cited 2021 Apr 21] Available from, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7334754/>.
- [23] Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI. A Review of Machine Learning Interpretability Methods. *Entropy Basel Switz* 2020 Dec 25;23(1).