

# A Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Prediction Model From Standard Laboratory Tests

Vafa Bayat,<sup>1,a</sup> Steven Phelps,<sup>2,a</sup> Russell Ryono,<sup>3</sup> Chong Lee,<sup>2</sup> Hemal Parekh,<sup>2</sup> Joel Mewton,<sup>2</sup> Farshid Sedghi,<sup>4</sup> Payam Etmnani,<sup>4</sup> and Mark Holodniy<sup>5,6,7</sup>

<sup>1</sup>Research and Development, Bitscopic Inc, Palo Alto, California, USA, <sup>2</sup>Data Science, Palo Alto, California, USA, <sup>3</sup>Clinical Applications, Bitscopic Inc, Palo Alto, California, USA, <sup>4</sup>Executive Management, Palo Alto, California, USA, <sup>5</sup>Public Health Surveillance and Research, Department of Veterans Affairs, Palo Alto, California, USA, <sup>6</sup>VHA Public Health Reference Laboratory, Palo Alto, California, USA, and <sup>7</sup>Division of Infectious Disease and Geographic Medicine, Stanford University School of Medicine, Stanford, California, USA

**Background.** With the limited availability of testing for the presence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus and concerns surrounding the accuracy of existing methods, other means of identifying patients are urgently needed. Previous studies showing a correlation between certain laboratory tests and diagnosis suggest an alternative method based on an ensemble of tests.

**Methods.** We have trained a machine learning model to analyze the correlation between SARS-CoV-2 test results and 20 routine laboratory tests collected within a 2-day period around the SARS-CoV-2 test date. We used the model to compare SARS-CoV-2 positive and negative patients.

**Results.** In a cohort of 75 991 veteran inpatients and outpatients who tested for SARS-CoV-2 in the months of March through July 2020, 7335 of whom were positive by reverse transcription polymerase chain reaction (RT-PCR) or antigen testing, and who had at least 15 of 20 lab results within the window period, our model predicted the results of the SARS-CoV-2 test with a specificity of 86.8%, a sensitivity of 82.4%, and an overall accuracy of 86.4% (with a 95% confidence interval of [86.0%, 86.9%]).

**Conclusions.** Although molecular-based and antibody tests remain the reference standard method for confirming a SARS-CoV-2 diagnosis, their clinical sensitivity is not well known. The model described herein may provide a complementary method of determining SARS-CoV-2 infection status, based on a fully independent set of indicators, that can help confirm results from other tests as well as identify positive cases missed by molecular testing.

**Keywords.** machine learning; human coronavirus; polymerase chain reaction; viral pneumonia.

The rapid emergence and spread of the virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and its associated disease coronavirus disease 2019 (COVID-19), combined with the limited availability of testing as exemplified by the recently promoted use of pooled testing [1], have driven the search for alternative screening methods [2–4]. Computed tomography (CT) scans are a proven alternative, but unnecessary irradiation and overuse of a limited resource for the purpose of screening are significant drawbacks [5]. And although protein-based antibody and antigen tests that deliver results on shorter timescales are now emerging, concerns about their accuracy persist [2–4].

The combination of vital signs with common laboratory tests presents a promising alternative means of SARS-CoV-2 diagnosis, particularly for those with active symptoms. In public settings such as airports, for example, the use of body temperature alone is a convenient if blunt diagnostic tool but has high false negative rates [6]. Meanwhile, various studies have noted that COVID-19 patients have characteristically low counts of white blood cells, lymphocytes, and platelets [1, 7–9] and elevated measures of serum ferritin and C-reactive protein (CRP) [10–12]. Therefore, we hypothesize that with the assistance of machine learning, small differences across a suite of commonly administered laboratory tests can, in aggregate, carry enough information to accurately infer active SARS-CoV-2 infection.

Such a method would have significant advantages. In many cases it would require no additional material outlay, as the necessary data are already present in the patients' medical records. In other cases, the data could be readily obtained using existing laboratory equipment and reagents with the same cost and turnaround time as standard blood panels. This would allow large numbers of patients to be rapidly analyzed. For inpatients, results could be generated in near real-time to compute a probability score that could be automatically updated as additional data become available. Entire inpatient

Received 30 May 2020; editorial decision 4 August 2020; accepted 20 August 2020; published online August 12, 2020.

<sup>a</sup>V. B. and S. P. contributed equally to this manuscript.

Correspondence: V. Bayat, Bitscopic, Inc, 715 Colorado Ave, Ste B, Palo Alto, CA 94303 (vafa@bitscopic.com).

Clinical Infectious Diseases® 2021;73(9):e2901–7

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com  
DOI: 10.1093/cid/ciaa1175

populations could be thus monitored, with alerts issued when the probability of SARS-CoV-2 infection crosses a predefined threshold. Outpatients would also benefit from rapid predictive analysis. Note that we believe the sensitivity and specificity of our method qualifies it as a complement to, rather than a replacement for, molecular diagnosis. Below we describe a model that can identify patients with SARS-CoV-2 based on patient temperature and a suite of commonly ordered laboratory tests.

## METHODS

### Data

The US Department of Veterans Affairs (VA) healthcare system serves over 9 million veterans at over 1200 Veterans Health Administration sites of care throughout the United States and US territories [13]. All VA facilities were included in this analysis. We collected data from 75 991 US veteran inpatients and outpatients who had received at least 1 SARS-CoV-2 RT-PCR or rapid antigen test during the period 8 March through 22 July 2020. Relevant data sources from VA sites were maintained and integrated using the Bitscopic Praedico® platform [14].

### Normalization

As a preliminary step, we normalized vital sign names, laboratory test names, specimen site locations, and measurement units to ensure that test results were commensurate between all facilities from which our cohorts were derived. This resulted in an initial list of 70 features comprising age and sex (2), vital signs (6), hematology (7), blood chemistries (49), and 4 composite features believed to have potential predictive value: the platelet: lymphocyte, neutrophil: lymphocyte, CRP: albumin, and PaO<sub>2</sub>:FIO<sub>2</sub> ratios [15, 16]. The initial suite of predictors for the machine learning algorithm was later reduced to 20 features as described below.

### Cohort

Our initial cohort comprised 188 132 patients tested for SARS-CoV-2 using emergency use authorization (EUA) approved molecular SARS-CoV-2 tests between 8 March and 22 July 2020. This was immediately reduced to 75 991 patients after imposing a constraint of possessing any 15 or more completed lab tests from the list described below in a window period around the SARS-CoV-2 test. The latter tests were primarily RT-PCR tests performed on nasopharyngeal swabs but also included serum PCR tests, rapid tests such as those performed on Cepheid machines, and antigen-based tests. As a result, the SARS-CoV-2 tests utilized for our analysis comprised a diverse and inclusive set. As some individuals received multiple tests, the total number of testing encounters available for training and testing our machine learning algorithm were 7335 positive and 84 919 negative cases. In total, 7191 (7.8%) of the tests were from female patients, and 85 063 (92.2%) of the tests were from male patients; 7.1% of the tests from female patients were positive and 8.0% of the tests from male patients were positive. The mean ages for tested patients were 66 for male patients and 53 for female patients (Table 1). We also examined comorbidities in patients using ICD10 codes collected over the prior 3 years to calculate Charlson and Elixhauser comorbidity indices [17].

### Time Windowing and Completeness Requirements

Laboratory data were collected against each SARS-CoV-2 testing encounter within a 48-hour window ( $\pm 24$  hours) of the SARS-CoV-2 testing date. Where multiple laboratory test results or vital signs were present within the time window, the median value was taken.

### Outlier Exclusion

Extreme outliers within each feature were eliminated based on empirical analysis of test result distributions. Data excluded in this way, which may have included erroneous or improperly

**Table 1. Summary of the Numbers of Patient Encounter Test Results and Demographic Information Relating the Number and Median Ages Broken Out by Sex of Unique Patients and Tests for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Positive and Negative Patients**

	Unique Patient Tests/Encounters			
	Tested	SARS-CoV-2-positive	SARS-CoV-2-negative	SARS-CoV-2-positive and negative
Total	92 254	7335	84 919	N/A
	Unique patients			
	Tested	SARS-CoV-2 only positive	SARS-CoV-2 only negative	SARS-CoV-2 positive and negative
Male	69 634	5003	63 841	790
Female	6357	437	5881	39
Total	75 991	5440	69 722	829
	Demographic information			
	Mean age (male)	65.82	65.72	68.80
	Mean age (female)	52.77	52.62	56.26
	Mean age (all)	64.73	64.67	68.21

Abbreviation: N/A, not applicable.

calibrated measurements, represented <0.2% of the total number of laboratory measurements.

### Feature Reduction

To simplify the analysis and facilitate the potential creation of an inexpensive and scalable test, it would be desirable to reduce the initial list of 70 features as much as possible without severely impacting prediction accuracy. Toward this end, pairwise correlations between all features were computed, and features with correlations at the  $\geq 90\%$  level were eliminated. Among the remaining features, a missingness of up to 93% was tolerated before a feature was excluded from further consideration. For the remaining 54 features, an initial run of our algorithm described below produced a list ranked by importance, from which we further reduced our included feature set to the top 20 [18]. The Charlson and Elixhauser comorbidity indices had minimal or even negative importance and were therefore not included as part of the final set of features.

### Machine Learning: Algorithm Selection, Training, Tuning, and Prediction

From the general nature of the problem (binary classification, with a moderate number of mostly normally distributed predictors), a number of standard machine learning approaches could be deployed. We chose XGBoost because of its accuracy in the test set as well as its high tolerance for missing data without the need for imputation (this was particularly significant for features such as ferritin with both high missingness and high importance) [19]. An initial round of hyperparameter tuning led to the selection of a learning rate ( $\eta$ ) of 0.1 and maximum tree depth of 4, with 500 maximum iterations. Three-quarters of the data in the minority class were used for training and cross-validation, and one-quarter were held out for testing. The dominant negative class was undersampled to balance the training set, whereas the test set was likewise undersampled to preserve the original response class ratio (which was moderately imbalanced at 8.0% SARS-CoV-2 positive). The training cohort comprised 5501 positive and 5501 negative encounters, and the testing cohort comprised 1834 positive and 20 929 negative encounters. Due to the undersampling, not all of the 84 919 eligible negative encounters were included in the training and testing. All statistical analyses were performed using R (version 3.6.0) [20].

## RESULTS

Our model was able to reproduce the molecular SARS-CoV-2 test results with a sensitivity of 82.4% and a specificity of 86.8%, and an overall test accuracy of 86.4% (with a 95% confidence interval of [86.0%, 86.9%]). The results are summarized in Table 2. It must be noted that the accuracy of our method can at this time only be measured against reference tests whose overall accuracies are themselves not well known; thus the

**Table 2. Summary of the XGBoost Machine Learning Prediction Model Results**

	SARS-CoV-2 (+) vs SARS-CoV-2 (-)
Min/Max features	15/20
Number of total tests/encounters	92 254
Number of unique patients	75 991
Number of SARS-CoV-2 (+) encounters	7335
Number of unique SARS-CoV-2 (+) patients	6269
Number of encounters in training set	11 002
Number of encounters in test set	22 763
Specificity (%)	86.77
Sensitivity (%)	82.39
Overall accuracy (%)	86.40
Positive predictive value (%)	35.30
Negative predictive value (%)	98.25

Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

potential accuracy of our method, should it become possible to train it against a highly accurate reference test, may be higher. A summary profile of the 20 selected features, including their missingness both before and after imposing a minimum test completion rate of 15, is given in Table 3, with the top 10 features in descending order of importance being serum ferritin, white blood cell count, eosinophil count, patient temperature, CRP, serum lactate dehydrogenase (LDH), D-dimer, basophil count, monocyte %, and serum aspartate aminotransferase (AST). The difference in distributions between SARS-CoV-2 positive and negative patients for each feature is shown in Figure 1. Of note, our requirement for 15 out of 20 of these commonly tested features permitted our eligible training set to cover 40% of the addressable VA SARS-CoV-2 tested population.

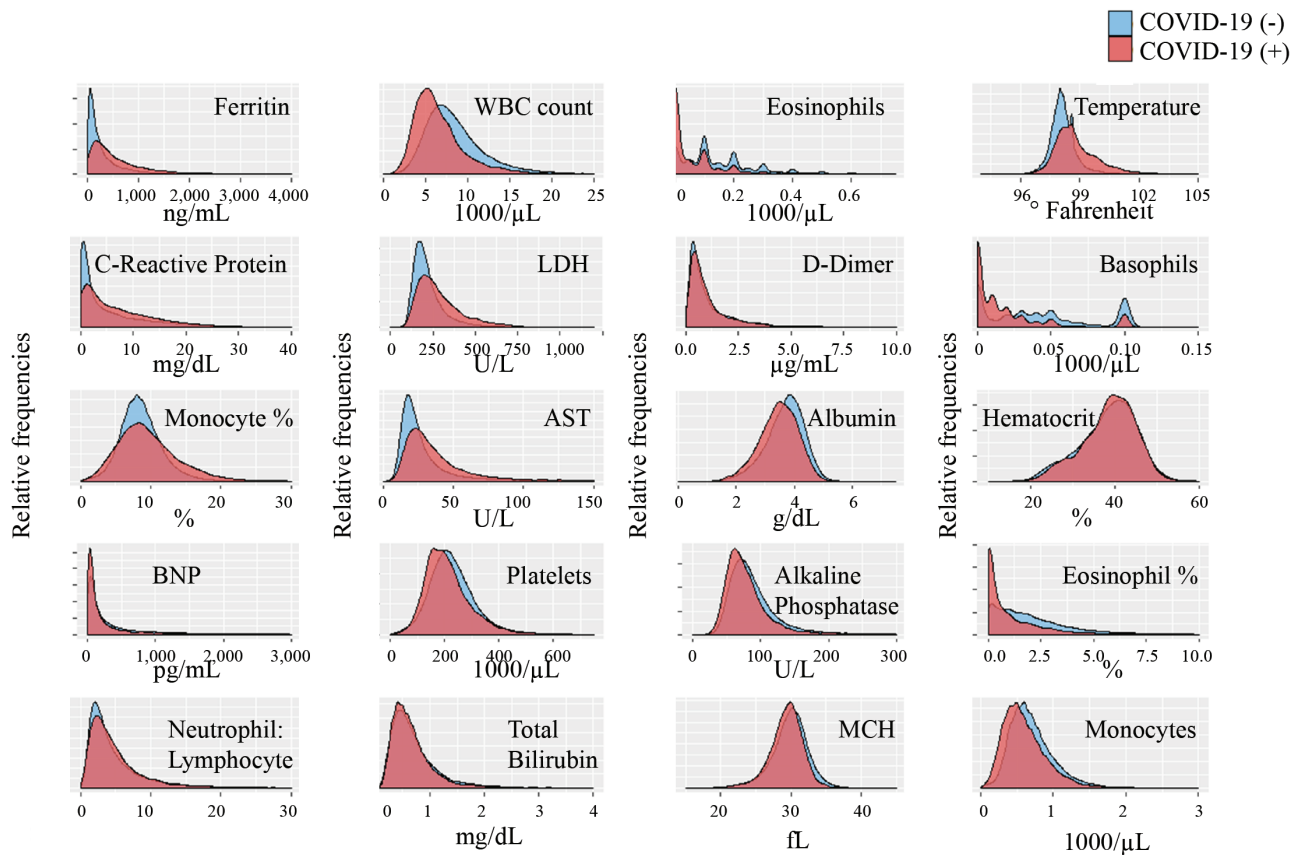
A particular area of concern in present-day SARS-CoV-2 testing is the false negative rate of molecular tests, with a meta-review estimating that the median false-negative rate was 38% on the day of symptom onset [21]. One way to analyze our model's potential as an independent source of information is to examine cases where multiple SARS-CoV-2 tests were administered to a single individual with mixed results: a negative test result followed within several days by a positive result may have been a false negative, especially because at this time the SARS-CoV-2 test is not typically administered in the absence of symptoms. Of the 77 instances where an individual in our cohort tested positive for SARS-CoV-2 subsequent to a negative test in the prior 3 days, our prediction model contradicted the negative molecular finding 54 times, or 70% of the time. In the converse case of instances where an individual tested negative for SARS-CoV-2 subsequent to a positive test in the prior 3 days, our prediction model contradicted the negative test finding 309 out of 788 times, or 39% of the time. The high percentage of positive predictions contradicting the negative reference tests (70% of all tested negatives soon followed by positives, and 39% of tested negatives soon

**Table 3. Summary of the 20 Features Discussed in the Manuscript, Listed in Descending Order of Importance.**

Features	Variable Importance	Missingness-COVID-19 Tested Patients (Pre-filter)	Missingness-COVID-19 Tested Patients (Post-filter)	Normal Range	COVID-19 Range (Present Study)	COVID-19 Range (ref. [1])	COVID-19 Range (ref. [7])	COVID-19 With ARDS (ref. [24])
<b>Ferritin (ng/mL)</b>	100.0	0.917	0.811	12–300	680.0 ± 823.0			
<b>White blood cell count (1000/uL)</b>	70.35	0.371	0.004	4–11	6.76 ± 3.26	7.0 ± 3.6	4.7 ± 1.2	7.2 ± 2.8
<b>Eosinophils (1000/uL)</b>	59.70	0.456	0.015	<0.5	0.07 ± 0.12			
<b>Patient temperature (°F)</b>	51.28	0.085	0.024	97–99	98.82 ± 1.12			98.2
<b>CRP (mg/dL)</b>	49.48	0.897	0.772	<0.8	7.40 ± 7.27	9.75 ± 6.64		8.72 ± 1.73
<b>LDH (U/L)</b>	44.28	0.921	0.811	60–100	311.10 ± 197.47	408.1 ± 231.0		483.0 ± 119
<b>D-dimer (ug/mL)</b>	36.51	0.927	0.829	<0.5	1.53 ± 2.26	4.0 ± 7.0		
<b>Basophils (1000/uL)</b>	26.69	0.472	0.027	<0.3	0.02 ± 0.03			
<b>Monocyte %</b>	24.27	0.449	0.013	2–8	9.62 ± 4.32			
<b>AST (U/L)</b>	19.94	0.541	0.046	0–35	41.75 ± 40.00	38.2 ± 24.6		25.5 ± 17.0
<b>Albumin (g/dL)</b>	18.95	0.473	0.011	3.5–5.5	3.47 ± 0.65	3.44 ± 0.57		3.07 ± 0.27
<b>Hematocrit (%)</b>	16.52	0.357	0.000	38–48 (male)	38.73 ± 6.30			
<b>BNP (pg/mL)</b>	14.72	0.846	0.668	<100	379.48 ± 798.90			
<b>Platelets (1000/uL)</b>	14.65	0.353	0.001	150–450	213.29 ± 88.2	162.7 ± 45	168 ± 39	166.5 ± 26
<b>Alkaline phosphatase (U/L)</b>	14.29	0.512	0.012	36–92	82.04 ± 45.78			
<b>Eosinophil %</b>	13.85	0.458	0.016	<6	1.11 ± 1.67			
<b>Neutrophil: lymphocyte ratio</b>	13.39	0.451	0.017	0.88–4.0	5.52 ± 5.97			
<b>Total bilirubin (mg/dL)</b>	12.67	0.501	0.015	0.1–1.2	0.68 ± 0.43			
<b>Mean corpuscular hemoglobin (pg)</b>	12.19	0.361	0.000	28–32	29.34 ± 2.47			
<b>Monocytes (1000/uL)</b>	6.17	0.441	0.006	0.3–0.9	0.62 ± 0.33			

Variable importances range from 0 to 100 on a relative scale. Missingness in the COVID-19 tested cohort before and after the application of the any-15-of-20 completeness requirement are shown. All feature data are listed as mean ± standard deviation. Information for the last 3 columns are taken from the referenced publications.

Abbreviations: ARDS, acute respiratory distress syndrome; AST, aspartate aminotransferase; BNP, B-type natriuretic peptide; COVID-19, coronavirus disease 2019; CRP, C-reactive protein; LDH, lactate dehydrogenase.



**Figure 1.** Separation in score distributions for each feature. The histograms represent relative probability density. In case of multiple scores within the time window, the median value was used. Red indicates SARS-CoV-2 positive patients and blue SARS-CoV-2 negative patients. See Table 3 for the importance of these and other features. Abbreviations: AST, aspartate aminotransferase; BNP, B-type natriuretic peptide; COVID-19, coronavirus disease 2019; LDH, lactate dehydrogenase; MCH, mean corpuscular hemoglobin; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WBC, white blood cell.

following positives), compared to the 8% prevalence of tested positives within our cohort, is evidence for the sensitivity of our algorithm to factors independent of those measured in the standard tests.

The possible effect of comorbidities on diagnosis using this method of testing was another area of concern, especially considering that comorbidities are known to have an important bearing on COVID-19 prognosis. We investigated this question by producing the composite Charlson and Elixhauser comorbidity indices for all patients in our cohort. After generating these scores, we confirmed that the 2 scores correlated well with each other, with a correlation of 0.77. We also determined that, not unexpectedly, they correlated well with age as well (0.65 and 0.45, respectively). However, when we performed a correlation analysis with our COVID-19 machine learning predictions, we found negligible correlations of  $-0.02$  for both, and we also noted that introduction of the indices as features in our model resulted in degraded performance and so did not include them.

## DISCUSSION

Our results are consistent with other reports suggesting that SARS-CoV-2 infection may be identifiable using vital signs, certain laboratory test results, or imaging. However, each of these taken in isolation lacks sensitivity or specificity. For instance, airport screening for elevated temperatures in travelers was estimated to capture only 46% of SARS-CoV-2-positive patients [6]. One report (in preprint format) described a machine learning-based method of diagnosing SARS-CoV-2 using chest CT exam features with much better results [22–24].

Our analysis identified several features as having high importance in spite of relatively high missingness in our dataset. Four such features are serum ferritin, CRP, LDH, and D-dimers. Indeed, COVID-19 has been classified as a hyperferritinemic syndrome due to the frequent finding of high ferritin [25]. Similarly, CRP, LDH, and D-dimer have been noted to be high in patients [12, 26]. Also of interest was the finding that unusually low blood eosinophil and basophil counts may be relevant to COVID-19 pathophysiology: eosinophil counts of 0,

known as eosinopenia, were observed in a cohort of 50 New York-based COVID-19 patients [27]. Lymphopenia has also been noted to be common in COVID-19 patients, and we saw this in the importance of one of our features, the neutrophil-lymphocyte ratio, which was well above the normal range (5.52 vs 0.884) (Table 3) [7, 28, 29].

There are several potentially useful applications for this model within healthcare settings. Given the need to protect inpatients admitted for non-SARS-CoV-2 medical issues, it could be used to monitor inpatients as results of commonly ordered laboratory tests become available. The high negative predictive value seen in our test (98.25%) could provide added assurance for patients and healthcare providers that these patients are and continue to be SARS-CoV-2 free and possibly conserve valuable personal protective equipment supplies. Healthcare providers could be alerted when potentially positive cases are identified to facilitate prompt and timely action for confirmatory SARS-CoV-2 PCR testing and/or implementing appropriate precautions. However, it should be noted that our patient cohort, having all received at least 1 SARS-CoV-2 test, were likely to have presented symptoms justifying the administration of the test and are therefore not representative of a general inpatient population; our algorithm would have to be retrained accordingly for use in such a population.

Another promising application is the provision of a method to cross-check the reference standard SARS-CoV-2 molecular tests such as the RT-PCR test performed on nasopharyngeal swabs. Although these tests are highly specific, as mentioned above their sensitivity has been questioned [30]. Of particular interest was the finding that our prediction model contradicted a negative molecular result 70% of the time when an individual later tested positive for SARS-CoV-2 within 3 days' time. Therefore, it is possible that our model can detect SARS-CoV-2 positive patients that received a false negative from the molecular test. If this conclusion is borne out in future analyses, our estimated true positive predictive value will likely be higher than what is reported here.

A third application is early detection of asymptomatic SARS-CoV-2 patients. The patient population in whom testing is performed will continue to evolve over time. Initially, a high percentage of patients being tested were symptomatic. As testing becomes more available, this percentage will likely fall significantly when more asymptomatic patients are tested for exposure or contact tracing. Our model's potential application to the identification of principally asymptomatic SARS-CoV-2 carriers is as yet unknown.

One potential limitation of our model is that it was trained on an older, mostly male patient population, although we noted that sex, age, and comorbidity indices are not critical predictors. Another limitation is the need for a number of patient laboratory data and vital signs for the most accurate predictions. However, the 15 required features are found in over 40% of

patients tested for SARS-CoV-2, and so it is not a significant concern. Although the required laboratory data and vital signs are not normally a heavy burden for a hospital, during a pandemic they may not always be readily available. However, in such cases, a reduced test panel may produce results with only a modest penalty to accuracy, as discussed above.

Molecular-based tests are at present the reference standard method of confirming a COVID-19 diagnosis; however, it is becoming well established that it is a limited standard, with many false negatives possibly due to patient status, sample type, and testing sensitivity and quality. The machine learning-based model described herein, utilizing vital signs and common laboratory test results, may provide an alternate, complementary method of accurately identifying COVID-19 patients. The generic character of our algorithm raises the possibility also noted by others, of using commonly performed hospital laboratory test results and vital signs to identify other diseases in addition to COVID-19 [31].

## Notes

**Author contributions.** V. B., S. P., R. R., P. E., F. S., and M. H. contributed to study conception, design, data analysis, and the writing of the manuscript. J. M., H. P., and C. L. contributed to data analysis. All authors have read, edited, and approved the final manuscript.

**Acknowledgments.** The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government. The authors thank Dr Jude Lopez for reviewing the manuscript. This project was approved by the Stanford University Institutional Review Board under the protocol entitled "Public Health Surveillance in the Department of Veterans Affairs." As the project was considered minimal risk, consent to participate was not required. Bitscopic is operating under a 10-year Research and Development agreement with the VA signed in 2019.

**Financial support.** This work was supported by Bitscopic's R&D budget and intramural funding from the Department of Veterans Affairs.

**Potential conflicts of interest.** The authors, with the exception of M. H., are all employees of Bitscopic, Inc. Otherwise they declare no conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

1. Guang C, Wu D, Guo W, et al. Clinical and immunologic features in severe and moderate forms of coronavirus disease 2019. *J Clin Invest* 2020; 130:2620–9.
2. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. *BMJ* 2020; 369:m1328.
3. Artesi M, Bontems S, Gobbels P, et al. Failure of the cobas® SARS-CoV-2 (Roche) E-gene assay is associated with a C-to-T transition at position 26340 of the SARS-CoV-2 genome. *J Clin Microbiol* 2020; JCM.01598-20. doi:10.1128/JCM.01598-20. Available at: <https://pubmed.ncbi.nlm.nih.gov/32690547/>.
4. Zhang L, Wang S, Ren Q, et al. Genome-wide variations of SARS-CoV-2 infer evolution relationship and transmission route. medRxiv. 2020. Available at: <https://www.medrxiv.org/content/10.1101/2020.04.27.20081349v2.full.pdf+html>.
5. Zhao W, Zhong Z, Xie X, Yu Q, Liu J. Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. *AJR Am J Roentgenol* 2020; 214:1072–7.
6. Quilty B, Clifford S, Flasche S, Eggo R, CMMID nCoV working group2. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). *Euro Surveill* 2020; 25:2000080.
7. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020; 382:1708–20.

8. 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**; 46:846–8.
9. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* **2020**; 8:475–81.
10. Han X, Cao Y, Jiang N, et al. Novel coronavirus disease 2019 (COVID-19) pneumonia progression course in 17 discharged patients: comparison of clinical and thin-section computed tomography features during recovery. *Clin Infect Dis* **2020**; 71:723–31.
11. Lagunas-Rangel F. Neutrophil-to-lymphocyte ratio and lymphocyte-to-C-reactive protein ratio in patients with severe coronavirus disease 2019 (COVID-19): a meta-analysis. *J Med Virol* **2020**; 10:1–2.
12. Ling W. C-reactive protein levels in the early stage of COVID-19. *Med Mal Infect* **2020**; 50:332–4.
13. Hussey PS, Jeanne SR, Sangeeta CA, et al. Farmer, resources and capabilities of the department of veterans affairs to provide timely and accessible care to veterans. Santa Monica, CA: RAND Corporation, 2015. Available at: [https://www.rand.org/pubs/research\\_reports/RR1165z2.html](https://www.rand.org/pubs/research_reports/RR1165z2.html).
14. Holodniy M, Winston C, Lucero-Obusan C, et al. Evaluation of Praedico™, a next generation big data biosurveillance application. *Online J Public Health Inform* **2015**; 7:e133.
15. Meng X, Wei G, Chang Q, et al. The platelet-to-lymphocyte ratio, superior to the neutrophil-to-lymphocyte ratio, correlates with hepatitis C virus infection. *Int J Infect Dis* **2016**; 45:72–7.
16. Oh J, Kim SH, Park KN, et al. High-sensitivity C-reactive protein/albumin ratio as a predictor of in-hospital mortality in older adults admitted to the emergency department. *Clin Exp Emerg Med* **2017**; 4:19–24.
17. Chu YT, Ng YY, Wu SC. Comparison of different comorbidity measures for use with administrative data in predicting short- and long-term mortality. *BMC Health Serv Res* **2010**; 10:140.
18. Van Buuren S. Flexible imputation of missing data. 2nd ed. Boca Raton, FL: Chapman & Hall, **2018**.
19. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Med Inform Decis Mak* **2019**; 19:211.
20. Chambers J. Software for data analysis: programming with R. New York: Springer, **2008**.
21. Kucirka L, Lauer S, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med* **2020**; M20–1495.
22. Yang Y, Yang M, Shen C, Wang F, Yuan J. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. *MedRxiv*, **2020**. Available at: <https://www.medrxiv.org/content/10.1101/2020.02.11.20021493v2>.
23. Ma X, Conrad T, Alchikh M, Reiche J, Schweiger B, Rath B. Can we distinguish respiratory viral infections based on clinical features? A prospective pediatric cohort compared to systematic literature review. *Rev Med Virol* **2018**; 28:e1997.
24. Tang X, Du R, Wang R, et al. Comparison of hospitalized patients with ARDS caused by COVID-19 and H1N1. *Chest* **2020**; 158:195–205.
25. Perricone C, Bartoloni E, Bursi R, et al. COVID-19 as part of the hyperferritinemic syndromes: the role of iron depletion therapy. *Immunol Res* **2020**; 68:213–24.
26. Chen L, Liu H, Liu W, et al. [Analysis of clinical features of 29 patients with 2019 novel coronavirus pneumonia]. *Zhonghua Jie He He Hu Xi Za Zhi* **2020**; 43:E005.
27. Tanni F, Akker E, Zaman MM, Figueroa N, Tharian B, Hupart KH. Eosinopenia and COVID-19. *J Am Osteopath Assoc* **2020**. doi:10.7556/jaoa.2020.091
28. Dang JZ, Zhu GY, Yang YJ, Zheng F. Clinical characteristics of coronavirus disease 2019 in patients aged 80 years and older. *J Integr Med* **2020**; S2095-4964(20)30072-8.
29. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **2020**; 395:507–13.
30. West C, Montori V, Sampathkumar P. COVID-19 testing: the threat of false-negative results. *Mayo Clin Proc* **2020**; 6196:30365–7.
31. Gunčar G, Kukar M, Notar M, et al. An application of machine learning to haematological diagnosis. *Sci Rep* **2018**; 8:411.