Machine learning is the key to diagnose COVID-19: a proof of concept study.

Cedric Gangloff (cedric.ganglof@gmail.com)
LTSI laboratory, INSERM U1099, Université de Rennes 1, Rennes, France. https://orcid.org/0000-0002-0253-4888

Sonia Rafi
LTSI laboratory, INSERM U1099, Université de Rennes 1, Rennes, France.

Guillaume Bouzillé
LTSI laboratory, INSERM U1099, Université de Rennes 1, Rennes, France.

Louis Soulat
Department of Emergency Medicine, Pontchaillou University Hospital, 35033 Rennes, France

Marc Cuggia
LTSI laboratory, INSERM U1099, Université de Rennes 1, Rennes, France.

Research Article

Keywords: COVID-19, SARS-coV-2, diagnosis, machine learning, chest-CT, RT-PCR

Posted Date: October 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-100691/v1

Version of Record: A version of this preprint was published at Scientific Reports on March 30th, 2021. See the published version at https://doi.org/10.1038/s41598-021-86735-9.

Abstract

Background: Two tests are currently available to diagnose COVID-19: chest-CT and RT-PCR. These tests are sub-optimal when they are used independently but their combination on every suspected COVID-19 patient requires considerable resources. The potential contribution of machine learning in this situation has not been evaluated. The objective of this study was to develop and evaluate machine learning models to diagnose COVID-19 among post-emergency hospitalized patients.

Methods: All post-emergency hospitalized adults patients admitted in our academic hospital between 2020/03/20 and 2020/05/05 and explored for COVID-19 were included in the study. Three types of machine learning models were created: logistic regressions, random forests, and neural networks. Each type of model was trained to diagnose COVID-19 with different sets of variables. Area under the ROC curve was the primary outcome to evaluate model's performances.

Results: 536 patients were included in the study: 106 in the COVID group, 430 in the NOT-COVID group. AUC of chest-CT increased from 0.778 to 0.889 with the contribution of machine learning. Similarly, AUC of RT-PCR increased from 0.852 to 0.929 with machine learning.

Conclusions: After generalization, machine learning models will allow to increase chest-CT and RT-PCR performances to diagnose COVID-19.

Background

The severe acute respiratory syndrome coronavirus 2 (SARS-coV-2) outbreak started in December 2019 in Hubei province in China. The associated disease, coronavirus disease 2019 (COVID-19), has now spread to all continents. The World Health Organization currently reports more than 11 million confirmed cases and 500,000 deaths (1). Increased mortality rates and the collapse of healthcare systems have been reported in several regions (2). Such a collapse is characterized by high contamination rates among healthcare workers, thus turning hospitals into epidemic hotspots (3,4). To limit this effect, patients with COVID-19 infection are hospitalized in specific units after being emergency department (ED) triage (5). Therefore, it is essential to have a reliable and easy-to-use tool for diagnosing COVID-19. Two tests are currently available (6,7): SARS-coV-2 real-time reverse transcriptase–polymerase chain reaction (RT-PCR) and chest computed tomography (chest-CT). Unfortunately, the diagnostic performance of these tests is sub-optimal when used independently of each other (8,9). The gold standard for diagnosing COVID-19 is currently represented by the combination of RT-PCR and chest-CT (10). However, it is a challenge to perform these two tests on each suspected COVID-19 patient for many reasons, including reagent shortages (11), device unavailability, lack of human resources, and high costs. In addition, the time required to perform the two tests increases the risk of ED overcrowding due to numerous patients awaiting their results. Hence, caregivers are forced to adapt their diagnostic strategy in accordance with their resources (12). To our knowledge, the potential contribution of machine learning in this situation has not been evaluated. Machine learning is an inherited data-science approach that enables computers to

extract or classify patterns. It allows an investigator to predict whether or not a patient belongs to a predefined group based on explanatory variables. The recent increase in machine learning models in the healthcare field suggests that these methods could improve the diagnostic strategy for COVID-19 (13). The objective of this study was to develop and evaluate machine learning models to diagnose COVID-19 among post-emergency hospitalized patients.

Materials And Methods

All methods were carried out in accordance with relevant guidelines and regulations. This study was approved by the ethic committee of Rennes academic hospital (number of approval: 20.93). Informed consent was waived by the local ethics committee in accordance with the reference methology MR 004 edicted by the French national data privacy regulatory commission.

Software. Data extractions, manipulations, statistical analyses, and modelizations were performed with “R-studio Server”, version 1.3.959, RStudio PBC, 2009-2020. Specialized packages and functions were used for specific analysis: “Dplyr”, version 1.0.0 was used for data manipulation, “Purrr”, version 0.3.4 for data simplification, and “missForest”, version 1.4 for missing data imputation. Random forests were built with “randomForest” version 4.6-14 and artificial neural networks with “neuralnet” version 1.44.2. “pROC” version 1.16.2 was used to generate the receiver operating characteristic (ROC) curves and calculate the area under the curve (AUC) for each model.

Setting. Data were collected retrospectively from patients admitted to the adult ED of Rennes Academic Hospital, France.

Patient selection. All post-emergency hospitalized patients ≥18 years old admitted between 2020/03/20 and 2020/05/05 and examined for COVID-19 with chest-CT and RT-PCR were included in the study. Patients opposed to the use of their data for research purposes were excluded.

Data collection. Data were automatically collected from “eHOP”, a local clinical data warehouse in which health data are integrated and de-identified in real time (14). Structured data, such as laboratory results, were directly collected from the data warehouse. Text fields were structured by using regular expressions (15).

Data pre-processing. In the raw data-frame, all values were associated with a unique identifier (ID) corresponding to each patient’s admission. This data-frame contained multiple lines per ID (figure 1, step 1). Variables collected more than once during the patient journey appeared as lists (figure 1, step 2). Lists were simplified according to the type of variable (figure 1, step 3).

Predicted variable. The predicted variable for each patient was the presence of COVID-19. This variable, denominated “COVID”, was coded “true” when RT-PCR and/or chest CT results were positive for COVID-19 and “false” otherwise. Chest-CT were coded “positive” when typical COVID-19 defects were identified by radiologists. Patients were allocated to “COVID” and “NOT-COVID” groups accordingly.

Predicting variables. All clinico-biological variables present in our local database were collected. Student's t- and chi-square tests were used to compare means between groups for numeric and binary variables, respectively. A p-value < 0.05 was considered statistically significant. Variables with a p-value < 0.2 were considered as variables of interest and selected to build machine learning models.

Data split. Data were randomly divided in two parts: the train data-frame, and the test data-frame. The train data-frame, corresponding to 80% of the whole data-frame, was used to build models. Models performances were evaluated on the test data-frame, corresponding to the remaining 20%.

Missing data imputation. Imputations of missing values were performed independently on each data-frame before the training process.

Model training. Three types of models were constructed: binary logistic regressions, random forests and artificial neural networks. Each type was trained with three sets of variables: clinico-biological variables, clinico-biological variables with chest-CT, and clinico-biological variables with RT-PCR.

Performance measurement. The area under ROC curves are commonly used to evaluate and compare classifiers in machine learning and biomedical and bioinformatics applications (16). In this study, the model's predictions were compared to the "COVID" variable in the test data-frame and ROC curves were constructed accordingly. AUC was the primary outcome used to evaluate the model's performances.

Results

Selected patients.

The patient selection flow chart is presented figure 2.

Selected variables.

Twenty-three clinico-biological variables were selected to build machine learning models (table 1).

	NOT-COVID (n=430)	COVID (n=106)	p
Clinicals and treatments			
Cough,%	83.0 (79.1–87.0	92.4 (86.5–98.2)	0.0563
Hyperthermia, %	66.7 (61.8–71.7)	77.2 (67.9–86.4)	0.0940
Myalgias, %	17.1 (13.2–21.1)	34.1 (23.7–44.6)	0.0012*
Asthenia, %	30.9 (26.1–35.8)	45.5 (34.5–56.5)	0.0187*
Diarrhea, %	22.9 (18.5–27.3)	32.9 (22.5–43.2)	0.0867
Confusion, %	21.7 (17.4–26.1)	7.5 (1.7–13.4)	0.0063*
Furosemid (usual treatment), %	16.0 (12.2–19.9)	6.3 (0.9–11.7)	0.0401*
Arterial Blood gas			
Base excess, mmol/L	3.0 (2.6–3.4)	2.7 (1.8–3.6)	0.0151*
Lactates, mmol/L	1.7 (1.5–1.9)	1.3 (1.1–1.5)	< 0.001*
Complete blood count			
Red blood cell count, Tera/L	4.2 (4.1–4.3)	4.5 (4.3–4.7)	< 0.001*
Mean platelet volume, fL	8.6 (8.4–8.8)	8.8 (8.5–9.1)	0.0269*
Leukocytes, G/L	10.2 (9.6–10.8)	7.7 (6.7–8.7)	0.0568
Neutrophils, G/L	7.9 (7.4–8.4)	6 (5.1–6.9)	0.1488
Platelet count	236.1 (225.8–246.4)	198.9 (182.1–215.7)	0.0482*
Eosinophils percentage	1.4 (1.1–1.7)	0.8 (0.4–1.2)	0.0873
Basophils percentage	0.6 (0.5–0.7)	0.4 (0.3–0.5)	<0.001*
Lymphocytes, G/L	1.3 (1.2–1.4)	1 (0.8–1.2)	<0.001*
Monocytes, G/L	0.8 (0.7–0.9)	0.6 (0.5–0.7)	<0.001*
Ionogram			
Potassium, mmol/L	4.1 (4–4.2)	4 (3.8–4.2)	0.0039*
Phosphor, mmol/L	1 (0.9–1.1)	1.1 (0.9–1.3)	<0.001*

Hemostasis and liver enzymes			
Alanine aminotransferase, mmol/L	64.5 (47.1–81.9)	46.2 (33.9–58.5)	0.1845
International normalized ratio	1.3 (1.2–1.4)	1.2 (1.1–1.3)	< 0.001*
D-Dimer, ng/ml	2200 (1600–2800)	2800 (1400–4200)	< 0.001*

Table 1. Clinico-biological variables selected to build machine learning models. Means and percentage between groups were compared with Student's t- and chi-square tests, respectively. Only variables with p < 0.02 were selected for model building and represented in this table. Values in parentheses represent 95% confidence interval. *=p < 0.005.

3.2. Chest-CT and RT-PCR performances.

AUCs of chest-CT and RT-PCR used to diagnose COVID-19 were equal to 0.778 (CI 95% = 0.682–0.873) and 0.852 (CI 95% = 0.764–0.940), respectively.

3.2. Models performances.

Areas under ROC curves for the three types of models trained with each set of variables are represented table 2.

	Clinico-biological	Clinico-biological +Chest-CT	Clinico-biological +RT-PCR
Binary logistic regression	0.764 (0.658–0.869)	0.889 (0.809–0.969)	0.929 (0.862–0.997)
Random forest	0.758 (0.647–0.869)	0.868 (0.772–0.965)	0.909 (0.820–0.997)
Artificial neural network	0.620 (0.516–0.725)	0.825 (0.716–0.934)	0.890 (0.801–0.978)

Table 2. Areas under the receiver operating characteristic (ROC) curves for each model. Three types of models were constructed: binary logistic regressions, random forests, and artificial neural networks. Each model was trained with three sets of variables: clinico-biological, clinico-biological with chest computed tomography (CT), and clinico-biological with real-time reverse transcriptase polymerase chain reaction (RT-PCR). Values in brackets represents 95% confidence interval.

ROC curves for binary logistic regression trained with chest-CT and RT-PCR are represented figure 3.

Discussion

Variables selected for model-building were consistent with clinico-biological signs of COVID-19. These variables belong to five categories: clinical signs, arterial blood gas, blood cell count, ionogram, Hemostasis and liver enzymes. **Clinical signs**: The COVID group had a statistically significant greater proportion of cough, hyperthermia, myalgia, asthenia, diarrhea, and confusion than the NOT-COVID group.

Such symptoms have previously been reported in numerous studies (17–21). Interestingly, anosmia was not selected as a variable of interest, suggesting a lack of relevance of this symptom in this setting (22,23). **Arterial blood gas**: In the NOT-COVID group, serum lactate concentration was higher, and base-excess was lower than in the COVID group, revealing the presence of patients suffering from circulatory failure, a frequently reported complication of bacteremia (24). Serum lactate concentration and base-excess are therefore relevant for differentiating between patients suffering from COVID-19 and bacterial infections. **Blood cell count:** Mean leukocytes, lymphocytes, and platelet counts were lower in the COVID group than in the NOT-COVID group. Previous authors have reported similar results. Indeed, a meta-analysis from Zhu *et al.* revealed that COVID-19 patients do not have hyperleukocytosis, except in the case of associated bacteremia (25). COVID-19-associated lymphopenia correlates with the severity of the disease and is related to an immune response deficiency (26). Similarly, thrombocytopenia was previously identified as a poor prognosis factor in this context (27). Indeed, a meta-analysis by Lippi et al. revealed that platelet count was significantly lower in patients with severe COVID-19 (28), suggesting an inappropriate activation of the coagulation process. **Ionogram:** Mean potassium concentration was lower in the COVID group. This could be due to hyperventilation, but further investigation must be conducted to confirm this hypothesis. **Hemostasis and liver enzymes**: Mean D-Dimer concentration was higher in the COVID group than in the NOT-COVID group. Elevated D-Dimers are associated with higher rates of thromboembolic events (29). These results are in line with the theory of an increased thromboembolic risk in case of COVID-19 (30–32). This finding could be associated with the presence of antiphospholipid antibodies, but the pathophysiology of this phenomenon is still under debate (33). Variables selected for model building were therefore consistent with previous studies that have reported clinico-biological signs of COVID-19.

The sub-optimal performances of single-tests to diagnose COVID-19 is confirmed. In our study, AUCs for RT-PCR and chest-CT were equal to 0.852 and 0.778, respectively. Multiple studies have reported similar results and conclude that sub-optimal performances of these tests when each one is used alone to diagnose COVID-19 occurs (9,10,34,35).

Machine learning models based on common clinico-biological signs will help to triage COVID-19 patients. RT-PCR and chest-CT are expensive, not available everywhere, and require qualified professionals to perform them. An increasing number of patients awaiting results of these tests can lead to ED overcrowding and increased mortality rates in an epidemic context (36,37). The logistic regression model presented in this study and trained only with clinico-biological variables had an AUC equal to 0.764. This model requires clinical examination and common biology assays to be performed: complete blood cell count, ionogram, standard hemostasis tests, liver enzymes, and arterial blood gas. Such tests are low-cost and can be realized worldwide on automatic devices. Triaging ED patients when RT-PCR and chest-CT are saturated or unavailable and is therefore possible by using machine learning.

Machine learning improves the performance of chest-CT and RT-PCR. Indeed, the AUC of chest-CT increased from 0.778 to 0.889 as a result of the contribution of machine learning (Table 2, logistic regression model). The AUC of RT-PCR increased from 0.852 to 0.929 in the same context. The

generalization of such models will allow high-quality diagnostics with just one of the two tests performed among RT-PCR or chest-CT to be provided. This type of system is an interesting solution for medium-sized hospitals since they would not have to acquire RT-PCR devices and reagents if they already have a chest-CT. On the other hand, priority access to RT-PCR must be given to hospitals that are not equipped with chest-CT. In larger structures with both CT-scanners and PCR machines, the choice of diagnostic strategy can then be made according to the availability in real-time of one or the other of the two tests. These structures can apply the logistic regression with chest-CT to certain patients and the logistic regression with RT-PCR to the remaining. This strategy will distribute the workload between the imaging and biology departments, reduce patient wait times, and minimize ED overcrowding, especially when facing large numbers of patients. In synthesis, the machine learning models described in this study should make it possible to decrease resources needed to diagnose COVID-19.

Limitations. The machine learning models developed in this article are not directly transferrable to other hospitals due to the monocentric design of this study. These models must be further developed and tested on a larger scale to be generalized. However, predictive variables selected in this study were similar to clinical and biological signs reported by previous authors, suggesting the absence of major obstacles for achieving this generalization.

Conclusions

This study demonstrates that machine learning models can be developed for diagnosing COVID-19 among post-ED hospitalized patients. The logistic regression model trained with clinico-biological variables is an acceptable solution to triage patients when RT-PCR and chest-CT are not available. Models based on chest-CT or RT-PCR will allow an increase in their performances and also performance of only one of these two tests to diagnose COVID-19 with a high level of confidence. After generalization, machine learning should play a key role in the management of the outbreak by reducing resources needed to diagnose COVID-19.

Declarations

Authors' contributions: C.G. designed the experiment, collected data, performed statistical analysis and build machine-learning models. C.G. and S.R. wrote the manuscript. G.B., L.S. and M.C read and approved the final manuscript.

Finantial support. None.

Competing Interest: Authors do not have any competing interest.

Ethics approval. This study was approved by the ethic committee of Rennes academic hospital (number of approval: 20.93)

Consent for publication. In accordance with the reference methology (MR 004) edicted by the French national data privacy regulatory commission, informed consent was waived by the local ethics committee. Patients opposed to the use of their data for research purposes were excluded from the study.

Availability of data and material. The dataset used in this study is not available.

Abbreviations

COVID-19: coronavirus disease 2019

RT-PCR: reverse transcriptase polymerase chain reaction

chest-CT: chest computed tomography

ED: emergency department

AUC: area under the curve

ROC: receiving operator characteristics

References

1. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 doi:10.1001/jama.2020.12839.
2. Coccolini, F. *et al.* COVID-19 the showdown for mass casualty preparedness and management: the Cassandra Syndrome. *World J. Emerg. Surg.* 15, 26 (2020).
3. Xiao, J., Fang, M., Chen, Q. & He, B. SARS, MERS and COVID-19 among healthcare workers: A narrative review. *J. Infect. Public Health* 13, 843–848 (2020).
4. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. *Zhonghua Liu Xing Bing Xue Za Zhi Zhonghua Liuxingbingxue Zazhi* 41, 145–151 (2020).
5. Maves, R. C. *et al.* Triage of Scarce Critical Care Resources in COVID-19 An Implementation Guide for Regional Allocation. *Chest* 158, 212–225 (2020).
6. Hanson, K. E. *et al.* Infectious Diseases Society of America Guidelines on the Diagnosis of COVID-19. *Clin. Infect. Dis.* ciaa760 (2020) doi:10.1093/cid/ciaa760.
7. Rubin, G. D. *et al.* The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. *Radiology* 296, 172–180 (2020).
8. Li, Y. *et al.* Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. *J. Med. Virol.* 92, 903–908 (2020).

9. Xiao, A. T., Tong, Y. X. & Zhang, S. False-negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence. *J. Med. Virol.* (2020) doi:10.1002/jmv.25855.
10. Ai, T. *et al.* Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* 200642 (2020) doi:10.1148/radiol.2020200642.
11. Beetz, C. *et al.* Rapid Large-Scale COVID-19 Testing during Shortages. *Diagnostics* 10, 464 (2020).
12. Lone, S. A. & Ahmad, A. COVID-19 pandemic – an African perspective. *Emerg. Microbes Infect.* 9, 1300–1308 (2020).
13. Furlow, B. Deep learning poised to revolutionise diagnostic imaging. *Lancet Respir. Med.* 5, 779 (2017).
14. Delamarre, D., Bouzille, G., Dalleau, K., Courtel, D. & Cuggia, M. Semantic integration of medication data into the EHOP Clinical Data Warehouse. *Stud. Health Technol. Inform.* 210, 702–706 (2015).
15. Tang, X., Zeng, Q., Cui, T. & Wu, Z. Regular expression-based reference metadata extraction from the web. in *2010 IEEE 2nd Symposium on Web Society* 346–350 (2010). doi:10.1109/SWS.2010.5607427.
16. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12, 77 (2011).
17. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet Lond. Engl.* 395, 497–506 (2020).
18. Jiang, F. *et al.* Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). *J. Gen. Intern. Med.* 35, 1545–1549 (2020).
19. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 395, 507–513 (2020).
20. Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* (2020) doi:10.1001/jama.2020.1585.
21. Goyal, P. *et al.* Clinical Characteristics of Covid-19 in New York City. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMc2010419.
22. Lechien, J. R. *et al.* Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur. Arch. Oto-Rhino-Laryngol. Off. J. Eur. Fed. Oto-Rhino-Laryngol. Soc. EUFOS Affil. Ger. Soc. Oto-Rhino-Laryngol. - Head Neck Surg.* 277, 2251–2261 (2020).
23. Beltrán-Corbellini, Á. *et al.* Acute-onset smell and taste disorders in the context of COVID-19: a pilot multicentre polymerase chain reaction based case–control study. *Eur. J. Neurol.* (2020) doi:10.1111/ene.14273.
24. Shankar-Hari, M. *et al.* Developing a New Definition and Assessing New Clinical Criteria for Septic Shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315, 775–787 (2016).

25. Zhu, J. *et al.* Clinicopathological characteristics of 8697 patients with COVID-19 in China: a meta-analysis. *Fam. Med. Community Health* 8, (2020).
26. Azkur, A. K. *et al.* Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19. *Allergy* 75, 1564–1581 (2020).
27. Hu, L. *et al.* Risk Factors Associated with Clinical Outcomes in 323 COVID-19 Hospitalized Patients in Wuhan, China. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* (2020) doi:10.1093/cid/ciaa539.
28. Lippi, G., Plebani, M. & Henry, B. M. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. *Clin. Chim. Acta Int. J. Clin. Chem.* 506, 145–148 (2020).
29. Cohen, A. *et al.* D-dimers in atrial fibrillation: a further step in risk stratification of thrombo-embolism? *Eur. Heart J.* 28, 2179–2180 (2007).
30. Zhang, L. *et al.* D-dimer levels on admission to predict in-hospital mortality in patients with Covid-19. *J. Thromb. Haemost. JTH* 18, 1324–1329 (2020).
31. Spiezia, L. *et al.* COVID-19-Related Severe Hypercoagulability in Patients Admitted to Intensive Care Unit for Acute Respiratory Failure. *Thromb. Haemost.* 120, 998–1000 (2020).
32. Oxley, T. J. *et al.* Large-Vessel Stroke as a Presenting Feature of Covid-19 in the Young. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMc2009787.
33. Zhang, Y. *et al.* Coagulopathy and Antiphospholipid Antibodies in Patients with Covid-19. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMc2007575.
34. Long, C. *et al.* Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *Eur. J. Radiol.* 126, 108961 (2020).
35. Liu, R. *et al.* Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clin. Chim. Acta Int. J. Clin. Chem.* 505, 172–175 (2020).
36. Toh, D., Thompson, C. H., Thomas, J. & Faunt, J. Emergency department overcrowding and mortality after the introduction of the 4-hour rule in Western Australia. *Med. J. Aust.* 196, 499–500; author reply 500 (2012).
37. Kim, J.-S. *et al.* Maximum emergency department overcrowding is correlated with occurrence of unexpected cardiac arrest. *Crit. Care Lond. Engl.* 24, 305 (2020).

Figures

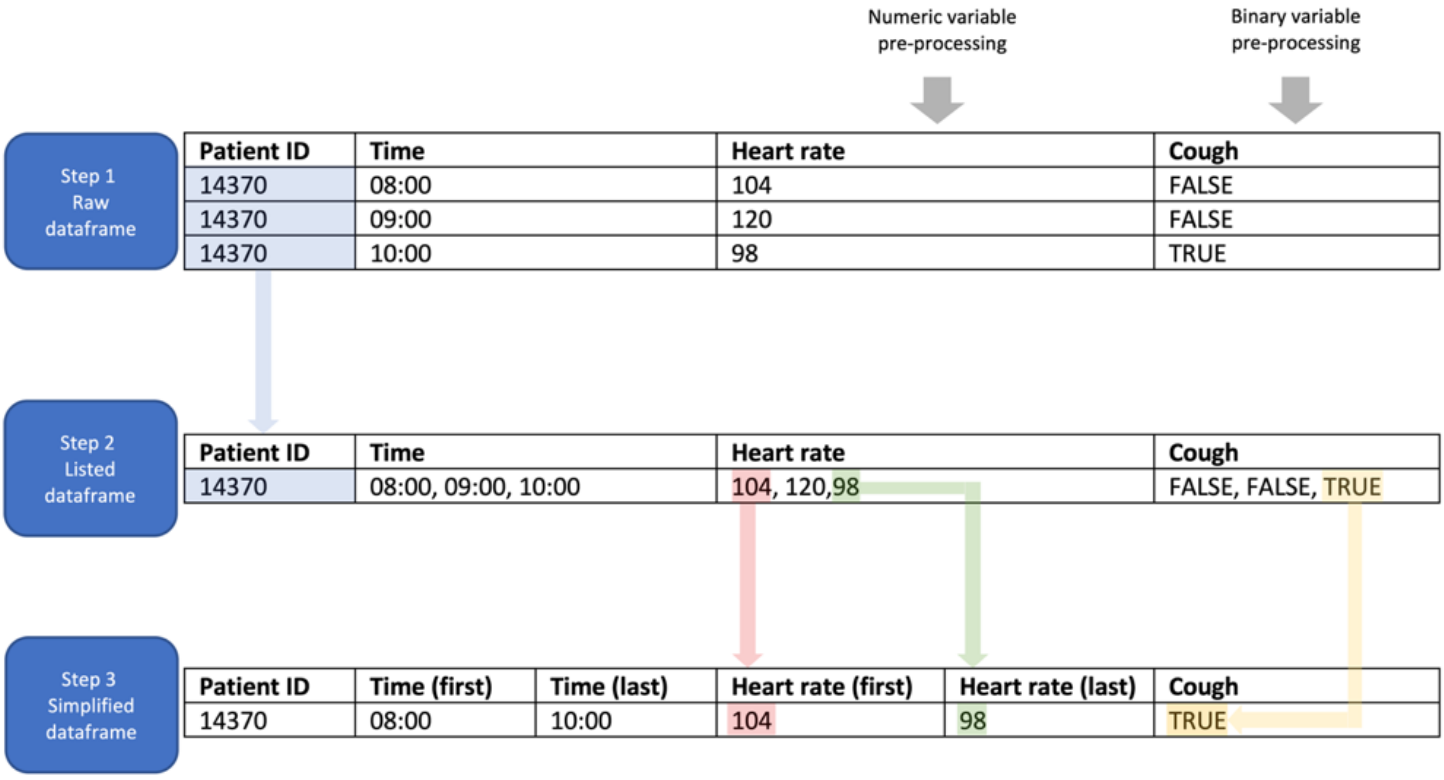

Figure 1

Data pre-processing. The first step corresponded to raw data, as they were initially stored in the database. Each ID was characterized by multiple rows. On the second step, data were listed in chronological order, with a single row per ID (blue arrow). In the third step, data were simplified. For numeric variables, only the first value was selected (red arrow). For binary variables, the value “true” was retained when it was present at least once in the list (yellow arrow).

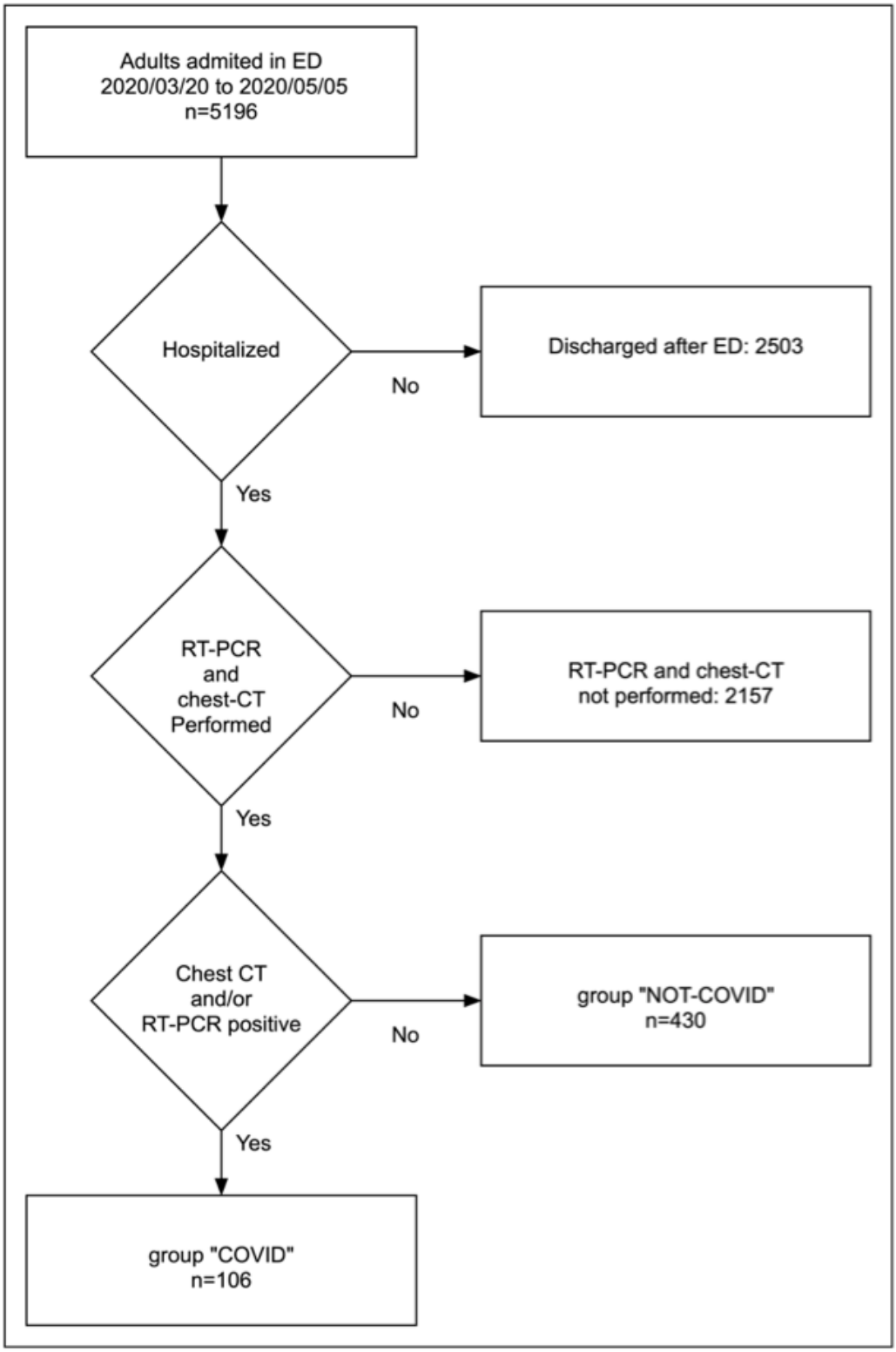

Figure 2

Flow chart of patient selection.

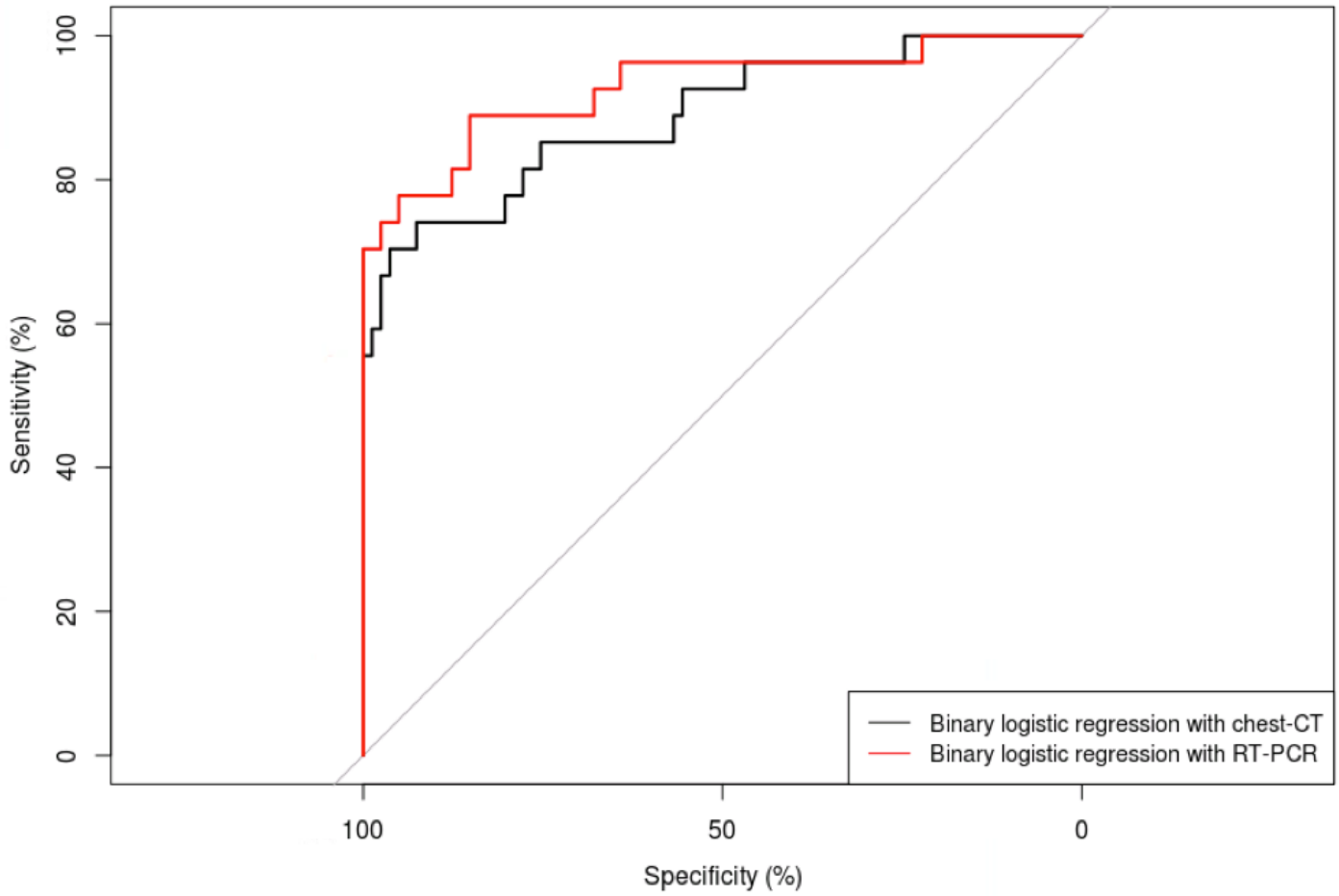

Figure 3

Receiver operating characteristics (ROC) curves for the two best performing models based on chest-computed tomography (CT) and reverse-transcriptase polymerase chain reaction (RT-PCR). “Binary logistic regression with chest-CT” (AUC] = 0.889) and “Binary logistic regression with RT-PCR” (AUC=0.929), appear in black and red, respectively. No statistical difference between these two AUCs was noted (p = 0.506).