COVID19 Outcome Prediction: Differences Between Raw Laboratory and Human Pre-Processed Data

Alessandro Principe (alessandro.principe@upf.edu)
Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Universitat Autònoma de Barcelona (UAB)
Silvia Gómez-Zorrilla
Infectious Diseases Department, Hospital del Mar, Infectious Pathology and Antimicrobials Research Group (IPAR)
Inmaculada López-Montesinos
Infectious Diseases Department, Hospital del Mar, Infectious Pathology and Antimicrobials Research Group (IPAR)
Eva Pérez-Almengor
Epilepsy Unit, Neurology Department, Hospital del Mar
Pau Pomés-Arnau
Bioenginery-Universitat Pompeu Fabra (UPF), Barcelona
María Sorli
Infectious Diseases Department, Hospital del Mar, Infectious Pathology and Antimicrobials Research Group (IPAR)
Robert Güerri-Fernández
Infectious Diseases Department, Hospital del Mar, Infectious Pathology and Antimicrobials Research Group (IPAR)
Santiago Grau
Pharmacy Department, Infectious Pathology and Antimicrobials Research Group (IPAR)
Albert Márquez
Information and Comunication System department, Hospital el Mar, Barcelona
Jordi Martínez-Roldán
Innovation and Digital Transformation Manager, Hospital del Mar, Barcelona
Juan Horcajada
Infectious Diseases Department, Hospital del Mar, Infectious Pathology and Antimicrobials Research Group (IPAR)

Research Article

Keywords: COVID19, machine learning, EHR

Posted Date: July 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-590537/v1

Abstract

Beyond health consequences, coronavirus disease 2019 (COVID19) has shown inadequacies and limitations of healthcare systems all over the world. Predictions of good and bad outcomes might improve lockdown and curfew policies and reduce the COVID19 health impact. We describe a multimodal machine learning (ML) approach through which we predict early hospital discharge and other key clinical outcomes as respiratory worsening, intensive care unit admission or the need of rescue therapy. Using this model, our hospital occupation might have decreased around 50% during the first wave of the pandemic. Our model used both, raw laboratory results (LR) and electronic health records (EHR), which allow a better anticipation of worsening through selective reporting of LRs. Thus, EHR models can be considered as a human preprocessing enhancing ML classification. Our approach may be a useful, customizable, and adaptable hospital resource management tool, especially to help interventions targeted at lowering COVID19 infection rates.

Introduction

Despite the positive results of vaccine trials[1], several months are needed to reach the critical amount of vaccinated in most of countries. Moreover, the assessment of long-term immunity will be crucial to predict whether the vaccination strategy will be able to eradicate, or at least curb the pandemic. For this reason, apart from strategies to reduce infection rates, it is of utmost importance the good management of hospitals, which could break the vicious cycle in which many economies have fallen into. Indeed, increasing early discharge (ED) and allocating the needed number of resources for in-room patients might cut hospital occupation and expenses. Many studies recurred to machine learning (ML) for coronavirus disease 19 (COVID19[2]), but the majority focused on diagnosis[3–5] and disease lethality[6–11]. Even though the latter is related to outcome, it cannot be translated into better resource allocation. Only a handful of studies tackled the problem in terms of time[12,13]. Here we describe a multimodal ML approach not only able to predict ED, but also other key clinical variables such as Intensive Care Unit admission (ICU), the intensification of oxygen support (OSi) or the need of rescue therapy. The clinical outcome prediction (COP) framework, COP through time (COP-tt), implements a ML algorithm already employed for mortality risk assessment[11] and clinical outcome prediction[13]. As in the study by Arvind & colleagues[13], COP-tt analyses time windows, but using a dynamic array of variables and more freedom in terms of missing data.

Briefly, for every subject, *states* are recorded. An electronic health record (EHR) state is a set of variables extracted from a clinical history entry, while a laboratory record (LR) state is a blood test result set. States are updated after one or more days and are defined when at least some variables or the outcome change. Each state was associated with an outcome (Fig. 1). To make predictions, models were constructed using the worst outcome. In other words, early stages of the severe acute respiratory syndrome (SARS) were associated with the worst stage reached by the patient, when the database was queried to train models. Models were constructed using a percentage of available subjects (sampling) and predicted the outcome of patients not considered at training. Sampling-prediction iterations were repeated to ensure homogeneous outcome prediction results (see methods for details). For a healthcare structure, each iteration could correspond to a week, after which a new set of models would be constructed for the upcoming one, with the advantage that in a real scenario sampling would not be necessary and new subjects would increase model precision.

Results

Subjects and sources

Since LRs and EHRs in our hospital are stored and retrieved disjointly, the two sources were treated separately. Indeed they are rather different, since variables in EHRs are recorded only if deemed necessary by the clinician. Overall, data of 4511 subjects with suspected COVID19 was retrieved, from March to August 2020. Of them, 2023 (45%) were confirmed to have been infected by SARS coronavirus 2 (SARS-CoV2), through reverse transcriptase-polymerase chain reaction rapid antigen detection or serology. However, only subjects with retrievable LR were considered for calculations, therefore 1685 were processed. To calculate ED, all subjects with suspected COVID19 were considered, since biological confirmations were not reliable at the beginning of the pandemic. For the rest of the results, only SARS-CoV2 positive subjects were considered to avoid biases. See methods for a detailed description of variables and their selection. Throughout iterations, outcomes of 2331 patients (51% of all subjects) were predicted.

Prediction of early discharge

LR models allowed better ED anticipation than EHR ones (Fig. 2A). For both sources the sensitivity was 100%, receiver-operator-characteristic areas-under-the-curve (ROC-AUCs) were 93.1% and 93.3%, but the anticipation was of 11 ± 33 days and 3 ± 11 (median ± interquartile-range) respectively. However, we believe that LR models overestimated ED anticipation, since they did not account for other causes of prolonged stay. Our hospital occupation might have been from 30% (EHR, Fig. 2B) to 45% (LR) lower on overall average, if ED was aided by COP-tt. More importantly however, when the pandemic was partially contained and only restrictions were in place, the hospital occupation might have lowered to around 50% (from minus 43–64%). The expected readmission rates would have been less than 10% on average (4–14%). ROC-AUCs rose to 98.0% (EHR) and 97.6% (LR), when only SARS-CoV2 subjects were considered, but the differences between anticipation rates were reduced to 3 ± 10 and 5 ± 20 days respectively, even when adjusted to current needs for starting a clinical trial (see supplementary results).

Prediction of worsening and rescue therapy

For OSi and ICU, sources behaved opposedly. EHR models anticipated OSi by 0 ± 2 days and ICU by 4 ± 9 days with ROC-AUCs of 80.6% and 91.2% respectively. LR models were less effective and accurate (ROC-AUCs: 73.6% and 82.1%). Interestingly, using EHRs, the need of rescue therapy (RT) could be anticipated by 2 ± 5 days. The scarcity of RT at the beginning of the pandemic was a considerable problem, therefore this kind of prediction could have been valuable.

Differences between EHR and LR

Further exploring the differences between EHR and LR models, intersection variables were compared (i.e. blood test variables recorded both in LRs and EHRs). As expected almost all variables were less reported in EHRs, since they were recorded only when necessary and missing numerical fields were automatically set to zero (Fig. 2C). However, the overestimated C reactive protein (CRP) confirmed that clinicians selectively reported higher determinations or did not report decreases. On the other hand, D-dimer and Interleukin-6 (IL6)

values were less or not divergent since they were necessary for RT indication and thus transcribed to EHRs when available.

EHR-LR subset results

The anticipation and ROC-AUCs of EHR models did not significantly change when results were calculated only on LR subjects, which rules out biases due to unbalanced sampling. For ED, sensitivity was 100%, ROC-AUC was 95%. For OSi, sensitivity was 89.5%, ROC-AUC was 77.8%. Finally, for ICU, sensitivity was 92.4%, ROC-AUC was 88.5%.

Variable change through time and models

A total of 414 RDF models were calculated for ED, 234 and 180 from EHR and LR respectively, while a total of 252 models were built for outcome worsening, 120 for intensification of OS (60 from EHR), 132 for ICU admission (60 from EHR). The most important variable across time and sources was *Age*, which appears on top all outcome worsening variables. The second most important clinical variable was *Days*, which represents the evolution of COVID19 in days. Among laboratory variables, CRP is of special interest, since it was the only overestimated variables of EHR and the best variable to predict ED from EHR. Expectedly, CRP is replaced by *PaO2/FiO2* for outcome worsening prediction (see Table 1). Conceivably, variables describing oxygenation and the acid base balance always ranked high in all models built from LR (see supplementary results). The importance of Interleukin-6, contrary to CRP, progressively rose in time and therefore was highest for ICU admission prediction from both sources, EHR and LR. On the other hand, the importance of D-dimer was high and rather constant among model kinds and through time. Unexpectedly, scores of early clinical deterioration (CURB65[14] and MEWS[15]) were never able to explain more than 3.4% of model prediction power, and only for ED. See supplementary results for a comparison between EHR and LR.

Table 1
Variable weights in EHR (electronic health record) models according to the different outcomes evaluated. Abbreviations: OS, oxygen support; ICU, Intensive Care Unit, paO2, partial arterial pressure of oxygen, FiO2, fraction of inspired oxygen; LDH, lactate dehydrogenase; WBC, white blood cell; MEWS, modified early warning score; ALT, alanine aminotransferase; AST, aspartate aminotransferase, BNP, brain natriuretic peptide; COPD, Chronic obstructive pulmonary disease.

Early Discharge		**Intensification of OS**		**ICU admission**	
variable	**weight**	**variable**	**weight**	**variable**	**weight**
C-reactive protein	12.8% ±0.8%	PaO2FiO2 ratio	10.7% ±0.7%	PaO2FiO2 ratio	13.9% ±1.0%
Age	8.9% ±0.3%	Age	6.7% ±0.8%	Age	6.5% ±0.3%
PaO_2FiO_2 ratio	7.6% ±0.5%	Procalcitonin	5.8% ±0.6%	LDH	5.4% ±0.4%
Days	5.8% ±0.3%	C-reactive protein	5.7% ±0.4%	Interleukin-6	5.1% ±0.5%
D-Dimer	5.6% ±0.4%	LDH	5.6% ±0.4%	D-Dimer	4.8% ±0.4%
LDH	5.3% ±0.5%	D-Dimer	4.6% ±0.2%	C-reactive protein	4.5% ±0.3%
Creatinine	4.7% ±0.1%	Interleukin-6	4.4% ±0.4%	Procalcitonin	4.5% ±0.5%
WBC	3.8% ±0.1%	Days	4.4% ±0.2%	Days	4.1% ±0.2%
Platelets	3.4% ±0.1%	Creatinine	4.0% ±0.5%	Bacterial co-infection	4.0% ±0.6%
MEWS score	3.4% ±0.3%	WBC	3.8% ±0.3%	WBC	3.9% ±0.3%
CURB-65 score	3.3% ±0.4%	Platelets (10^3/μL)	3.3% ±0.2%	Creatinine	3.6% ±0.2%
Interleukin-6	3.0% ±0.5%	Lymphocytes (10^3/μL)	3.1% ±0.1%	Platelets (10^3/μL)	3.0% ±0.2%
Procalcitonin	3.0% ±0.4%	ALT	3.0% ±0.2%	BNP	2.9% ±0.3%
Lymphocytes (10^3/μL)	2.9% ±0.2%	AST	2.9% ±0.1%	Lymphocytes (10^3/μL)	2.8% ±0.1%
Chronic heart disease	2.7% ±0.3%	MEWS score	2.8% ±0.2%	ALT	2.6% ±0.1%
ALT	2.3% ±0.1%	Bacterial co-infection	2.7% ±0.6%	AST	2.6% ±0.1%
BNP	2.3% ±0.2%	BNP	2.7% ±0.3%	MEWS score	2.4% ±0.2%

Early Discharge		Intensification of OS		ICU admission	
AST	2.2% ±0.1%	Platelets/Lymphocytes	2.5% ±0.3%	Ferritin	2.2% ±0.2%
Ferritin	2.1% ±0.2%	Troponin T	2.4% ±0.2%	Troponin T	2.2% ±0.2%
Platelets/Lymphocytes	1.9% ±0.3%	Ferritin	2.2% ±0.2%	Platelets/Lymphocytes	2.2% ±0.2%
Sex	1.9% ±0.2%	CURB-65 score	2.1% ±0.3%	CURB-65 score	2.0% ±0.3%
Bacterial co-infection	1.9% ±0.3%	Sex	1.9% ±0.2%	Chronic heart disease	2.0% ±0.3%
Troponin T	1.8% ±0.1%	Obesity	1.6% ±0.5%	Obesity	1.6% ±0.2%
Chronic renal disease	1.2% ±0.1%	Chronic heart disease	1.3% ±0.2%	Chronic renal disease	1.3% ±0.3%
Diabetes Mellitus	1.1% ±0.1%	Chronic renal disease	1.1% ±0.2%	Sex	1.1% ±0.1%
COPD	1.0% ±0.1%	Diabetes Mellitus	1.0% ±0.2%	Diabetes Mellitus	1.0% ±0.1%
Obesity	0.9% ±0.1%	Active cancer	0.8% ±0.1%	COPD	0.8% ±0.2%
Active cancer	0.8% ±0.0%	COPD	0.7% ±0.1%	Active cancer	0.7% ±0.1%
Asthma	0.6% ±0.0%	Asthma	0.6% ±0.1%	Autoimmune disease	0.5% ±0.2%
Autoimmune disease	0.5% ±0.0%	Autoimmune disease	0.5% ±0.1%	Asthma	0.5% ±0.1%
Lymphocytes T CD4	0.2% ±0.0%	Lymphocytes T CD4	0.1% ±0.0%	Lymphocytes T CD4	0.2% ±0.1%

Discussion

This is the first study that compares human preprocessed versus raw data using the same ML framework and subjects in a complex clinical setting. Indeed, for COVID19 outcome prediction the time dimension had to be added to a multimodal categorization mixing clinical and laboratory variables. To achieve this goal, a powerful ML tool working at its best for cathegorical classification was adjusted for a special time series analysis that considered patient states in time. For this purpose, a database was created and automated to extract target data from the digitally recorded clinical histories. Apart from that, also raw laboratory results were used as a source, therefore allowing us to analyze whether human interference could hamper or boost ML classification. We hypothesize that selective reporting performed by clinicians worked like a filter that increased the differences between patient states, thus facilitating ML classification, at least for complex

clinical conditions like global worsening and the necessity of admission to intesive care. In this sense, EHR models represent a human enhancement of machine learning analysis, a concept that might be applied to different clinical problems, both in diagnosis and follow-up.

As for variable change through time and their weight in prediction, *Age* led the group across sources, as in other studies, followed by arterial oxygenation measurements. In addition to other studies though, the time analysis provided the possibility to see the shift of variable importance for different predictions, as shown by the proinflammatory CRP, useful for early discharge, and the *PaO2/FiO2*, more suitable for predicting clinical worsening and ICU admission.

By our knowledge, COP-tt is the first framework that might be successfully used for both early discharge and severe outcome prediction of COVID19. Throughout the pandemic the rise of new COVID19 cases have been related to higher hospital occupation and higher death rates. Despite the vaccination campaigns started, the number of new cases worldwide is still rising[16], there are reports of hospitalizations of vaccinated patients, as well as recorded deaths[17,18]. However, it is also vital to start reducing restrictions, since virtually all healthcare systems are tightly linked to the economic welfare of their related countries, therefore an economical downturn generated by the fall of several industry sectors would most certainly undermine the functioning of healthcare systems. For these reasons, the prediction of outcomes and thus a better allocation of healthcare resources could become even more crucial now than at the beginning of the COVID19 pandemic. By just using EHR models and selective restrictions, which are interventions aimed at reducing infection rates, our hospital occupation due to COVID19 might have dropped by almost 45% (Fig. 2B). Thus, we expect that COP-tt could increase the efficacy of other interventions like vaccination, which could be studied as a variable for outcome prediction.

The major study limitation is its monocentric nature. However, the number of analyzed patients matches and at times even overcomes that of multicentric studies[12]. The second major problem is the applicability to other health structures. Nevertheless, this study demonstrates that the framework can work using very different, if not opposite, sources, EHR and LR. This finding suggests that other variables or other kind of recordings might be used to achieve the same goal. In fact, as compared to huge multicentric models[19], which aim at generalizing as much as possible their prediction power, our study and framework is targeted at a local, customized usage, which could depend on the hospital infrastructure and laboratory possibilities. Indeed, the framework is available for clinicians and researchers in the GitHub repository mentioned in methods. Moreover, a new repository adapted for a clinical trial is also available and currently updated.

Summarizing, the presented ML framework is not a model but a model generator for diverse COVID19 clinical outcome predictions. It can work under very different circumstances, since it successfully processed 31 (EHR) and 120 (LR) variables reaching similar results, and it is very resistent to missing data. The framework actually transforms missing data into an advantage more than a limitation, as demonstrated by the highlighted differences between human reported laboratory results and the objective recordings. Since the framework is a model generator, we expect it to be adaptable to different realities and goals, like assessing the clinical impact of SARS-CoV2 variants[20] or emerging clinical variables[21].

Materials And Methods

Data sources and preprocessing

Hospital del Mar database separates electronic health records (EHRs), which are clinical history entries, from laboratory results (LRs). Since all records are virtually always in use, data retrieval errors may occur, especially during data usage peaks, which became longer and more frequent during the epidemic. Data from EHR were retrieved in a single query with the search word “COVID”, from March until August 2020, included. LRs could not be searched in the same way, therefore all data was retrieved using just temporal bounds. Identifier numbers of all patients were translated into new random identifiers before analysis, the translation index was kept by JPH and MLS, who were not directly involved in data processing and analysis. EHRs were further processed using keywords and proximity to retrieve numeric or binary variables. Presence of certain conditions was recorded as binary variables, e.g., cardiovascular diseases, while numerical variables were age and transcribed LRs. All EHR variables were located using synonym lists, proximity (distance from the located keyword) was used to parse numerical values. Outcomes were retrieved in the same fashion. To assign outcomes to LRs, EHR data was crossed with LR data. Confirmation of SARS-CoV2 infection was determined either by LR or EHR, giving preference to LRs.

Tables were constructed from raw EHR and LR data, where each line corresponded to a patient initialization or update. Therefore, for each patient one or more lines were compiled. Either initialization or update was located in time using the entry date. If some variable could not be retrieved from the entry in a specific time, default values were used: 0.0 for numerical values; “NO” for binary values. The special variable sex was treated as binary, referring uniquely to the biological phenotype. If no assessment of sex could be retrieved, the subject was discarded. Tables were saved as comma separated values (CSV) files, using “;” as separator.

Data variables

During the COVID-19 pandemic, in our hospital, physicians from different specialties attended COVID-19 patients. In order to homogenize patient management, guidelines were elaborated by the Infectious Diseases Service for both data collection in EHR and therapeutic management. EHR variables were selected using the clinical history recording guide that was issued by the COVID19 task force of our structure, and by the clinical experience of the authors involved in COVID19 shifts, SG-Z, ILM and AP. Demographic, clinical, and epidemiological data were collected from EHR including the following: age, sex, and race, underlying diseases, clinical and outcome data of COVID-19, complementary tests (kidney and liver profile, electrolytes, myocardial enzymes, blood count, coagulation profile, including D-dimer, and inflammatory markers such as serum Interleukin-6, IL6, ferritin and C reactive protein). On the other hand, LR variables were selected among the 20% most requested blood tests during the aforementioned temporal range. For a complete list of EHR and LR variables, see variable weight tables in Results and Supplementary Results.

The need of rescue treatment was defined as the addition of IL-6 inhibitors such us *tocilizumab* or *sarilumab* to the standard therapy in patients with clinical deterioration (increasing of lung infiltrates or inflammatory markers and/or changes in need for oxygen or ventilatory support). Regarding respiratory requirements, ventilatory support was defined as the need for non-invasive mechanical ventilation (MV), high flow nasal cannula or invasive MV. The intensification of oxygen support was considered when during the hospital

admission was necessary to increase oxygen needs (need to increase FiO2 by nasal cannulas or mask and/or need to use ventilatory support).

COVID-19 was defined as a SARS-CoV-2 infection confirmed by quantitative reverse transcriptase–polymerase chain reaction, rapid diagnostic tests based on antigen detection in a nasopharyngeal sample or serology. Those patients with a negative test but who fulfilled clinical diagnostic criteria: respiratory symptoms (dyspnea, cough, sore throat, changes in taste/smell), chest X-Ray findings (uni- or bilateral interstitial infiltrates), and laboratory abnormalities (lymphopenia, the raise of C reactive protein, ferritin, IL-6 or D-dimer), which made COVID-19 the most likely diagnosis in the current epidemiological situation were considered as probable infection and processed for the anticipation of early discharge.

Clinical Outcome Prediction through time (COP-tt)

As first step, the framework uploads CSV tables to create an internal database, where subject initialization or updates are stored as states. In states, variables are either numerical or categorical, being the binary a subset of categorical. When a variable is categorical, COP-tt reads all possible text entries and automatically assigns them a numerical counterpart, which is used during analysis. Despite being categorical, the outcome variable was not automatically assigned but was defined as shown in Fig. 1 of the main text. Analysis consists in three steps: data sampling, model training and outcome prediction. For the latter, the outcome of each state is upgraded to the worst outcome reached by the subject. This upgrade can be disabled to create models for outcome detection instead of prediction.

During the first step, the database is split between training and testing set. The number of subjects in each set is defined by the training set ratio, which is arbitrary. The ratio in the current study was changed depending on the type of prediction and availability of patients (60 to 90%). In general terms, the bigger and more varied the training set, the more accurate the prediction.

Several random decision forest (RDF)[22] models are trained to simulate the k-fold validation[23], widely used in machine learning (ML). Before training a specific model, the training set is split into training and testing subsets, which are checked to ensure a balanced number of outcomes (50% good). An error threshold is employed to bypass unbalanced samplings, since it is not always possible to ensure well balanced sets. When the training set is unbalanced beyond the arbitrary error threshold, the sampling is repeated. In this study, six RDF models were trained before outcome prediction. The number of RDF trees can be set before analysis.

During the last step, each RDF model classifies all states of a testing subject. Since 0 is used for good and 1 for bad outcome, the bad outcome prediction is defined when the average model prediction overcomes or equals a threshold (0.5 in this study, see Fig. 1). For instance, if three models predict bad, while three good outcome, the final prediction is bad. The prediction is through time because it is calculated at each state, then the time before the target outcome is determined. For example, when calculating the anticipation of intensive care unit (ICU, target outcome), it is possible that the final prediction of bad outcome is achieved at a state preceding the real admission at ICU, which counts as real prediction; at the same state of admission, which corresponds to a detection; or after admission at ICU, which is failure to detect or anticipate. In this case, patients non admitted at ICU did not count for the final anticipation statistics. Target outcomes and therefore

outcome discriminators can be set before analysis. The bad versus good discriminators used in this study are shown in Fig. 1 of the main text.

The framework current version is written in Python (version 3.8.5), using the Miniconda data science platform (version 4.9.2). The Scikit-learn module (version 0.23.2) was used for the RDF implementation. COP-tt accepts CSV tables and text files as indexes of variable names and types. See code and instructions in the GitHub repository: joricomico/COVID19_cop: COP-tt test, application to COVID19 (github.com).

Variable comparison between EHR and LR

Because of technical retrieval problems, LRs represented a smaller and complete subset of ERHs. For this reason, to compare EHR and LR random samples of the intersection variables (variables present in both sets) were compared using either the Student's t test or the Mann-Whitney U test, depending on distributions, evaluated through the Shapiro-Wilk test. One-hundred samples of 1000 values were compared for each variable. Average p-values and the percentage of times $p < 0.05$ were used to determine statistical significance.

Variable ponderation

RDF models are ensembles of decision trees, where variables are randomly split and linked to outcomes. A variable split (a threshold or threshold vector) in a decision tree defines how much a variable can discriminate between the two or more outcome groups. Through the discrimination power of each sub-model (a decision tree), variable weights can be assessed as the average discrimination power of each variable to classify target outcomes. To assess global variable ponderations, variable weights were calculated for each RDF model, then average weights were calculated across all models.

Ethics

Data collection, statistical and machine learning data treatments have been performed in accordance with the Declaration of Helsinki. The study was approved by the Clinical Research Ethics Committee of the Parc de Salut Mar (register no 2020/9329/I). Data anonymization was carried out in order to protect the privacy of patients. The need for written informed consent was waived due to the observational nature of the study and retrospective analysis.

Declarations

Acknowledgments: The authors thank the Spanish Society of Infectious Diseases and Clinical Microbiology (SEIMC) which supports Dr. Silvia Gómez-Zorrilla research.

Author contributions:

Conceptualization: AP, SG-Z, ILM, JPH, MLS

Methodology: AP, PP

Investigation: AP, SG-Z, ILM, PP, EP

Visualization: AP

Project administration: JMR, AM, JPH, MLS

Supervision: SG, JPH, MLS

Writing – original draft: AP, SG-Z, ILM

Writing – review & editing: EP, PP, MLS, RGF, SG, AM, JMR, JPH

Competing interests: SG has received fees as speaker for Pfizer, Angellini, Kern and MSD and research grants from Astellas Pharma and Pfizer. JPH reports being a speaker with honoraria in advisory boards for MSD, Pfizer, Angelini, and Menarini, and having a research grant from MSD. The authors report no conflicts of interest in this work.

Data and materials availability: All code is available in GitHub (see Supplementary Materials). AI results are equally available in the same repository.

References

1. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med*, **383**, 2603–2615 (2020).
2. Lalmuanawma, S., Hussain, J. & Chhakchhuak, L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons and Fractals*, **139**, 110059 (2020).
3. Albahli, S. Efficient gan-based chest radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia. *Int. J. Med. Sci*, **17**, 1439–1448 (2020).
4. Wang, S. *et al.* A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis.*Eur. Respir. J.***56**, (2020).
5. Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R. & Rouf, N. & Mohi Ud Din, M. Machine learning based approaches for detecting COVID-19 using clinical text data. *Int. J. Inf. Technol*, **12**, 731–739 (2020).
6. Wu, G. *et al.* A prediction model of outcome of SARS-CoV-2 pneumonia based on laboratory findings. *Sci. Rep*, **10**, 1–9 (2020).
7. Li, X. *et al.* Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. *PeerJ*, **8**, 1–19 (2020).
8. Cheng, F. Y. *et al.* Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. *J. Clin. Med*, **9**, 1668 (2020).
9. Wu, G. *et al.* Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: An international multicentre study.*Eur. Respir. J.***56**, (2020).
10. Assaf, D. *et al.* Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern. Emerg. Med*, **15**, 1435–1443 (2020).

11. Das, A. K., Mishra, S. & Gopalan, S. S. Predicting CoVID-19 community mortality risk using machine learning and development of an online prognostic tool. *PeerJ*, **8**, 1–12 (2020).
12. Liang, W. *et al.* Early triage of critically ill COVID-19 patients using deep learning. *Nat. Commun*, **11**, 1–7 (2020).
13. Arvind, V., Kim, J. S., Cho, B. H., Geng, E. & Cho, S. K. Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19. *J. Crit. Care*, **62**, 25–30 (2021).
14. Lim, W. S. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study., **58**, 377–382 (2003).
15. Subbe, C. P. Validation of a modified Early Warning Score in medical admissions. *QJM*, **94**, 521–526 (2001).
16. Data as received by WHO from national authorities. *Weekly epidemiological update on COVID-19–4 May 2021*. *WHO Epidemilogical Update* vol. 4 May https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—4-may-2021 (2021).
17. Keehner, J. *et al.* SARS-CoV-2 Infection after Vaccination in Health Care Workers in California. *N. Engl. J. Med*, **384**, 1774–1775 (2021).
18. Haas, E. J. *et al.* Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data., **397**, 1819–1829 (2021).
19. Schwab, P. *et al.* Real-time prediction of COVID-19 related mortality using electronic health records.*Nat. Commun.***12**, (2021).
20. Tse, H. *et al.* Emergence of a Severe Acute Respiratory Syndrome Coronavirus 2 virus variant with novel genomic architecture in Hong Kong. *Clin. Infect. Dis*, https://doi.org/10.1093/cid/ciab198 (2021).
21. Marfia, G. *et al.* Decreased serum level of sphingosine-1-phosphate: a novel predictor of clinical severity in COVID-19.*EMBO Mol. Med.***13**, (2021).
22. Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* vol. 1 278–282 (IEEE Comput. Soc. Press).
23. Yan-Shi, D. Ke-Song Han. A comparison of several ensemble methods for text categorization. in IEEE International Conference onServices Computing, 2004. (SCC 2004). Proceedings. 2004 419–422 (IEEE). doi:10.1109/SCC.2004.1358033.

Figures

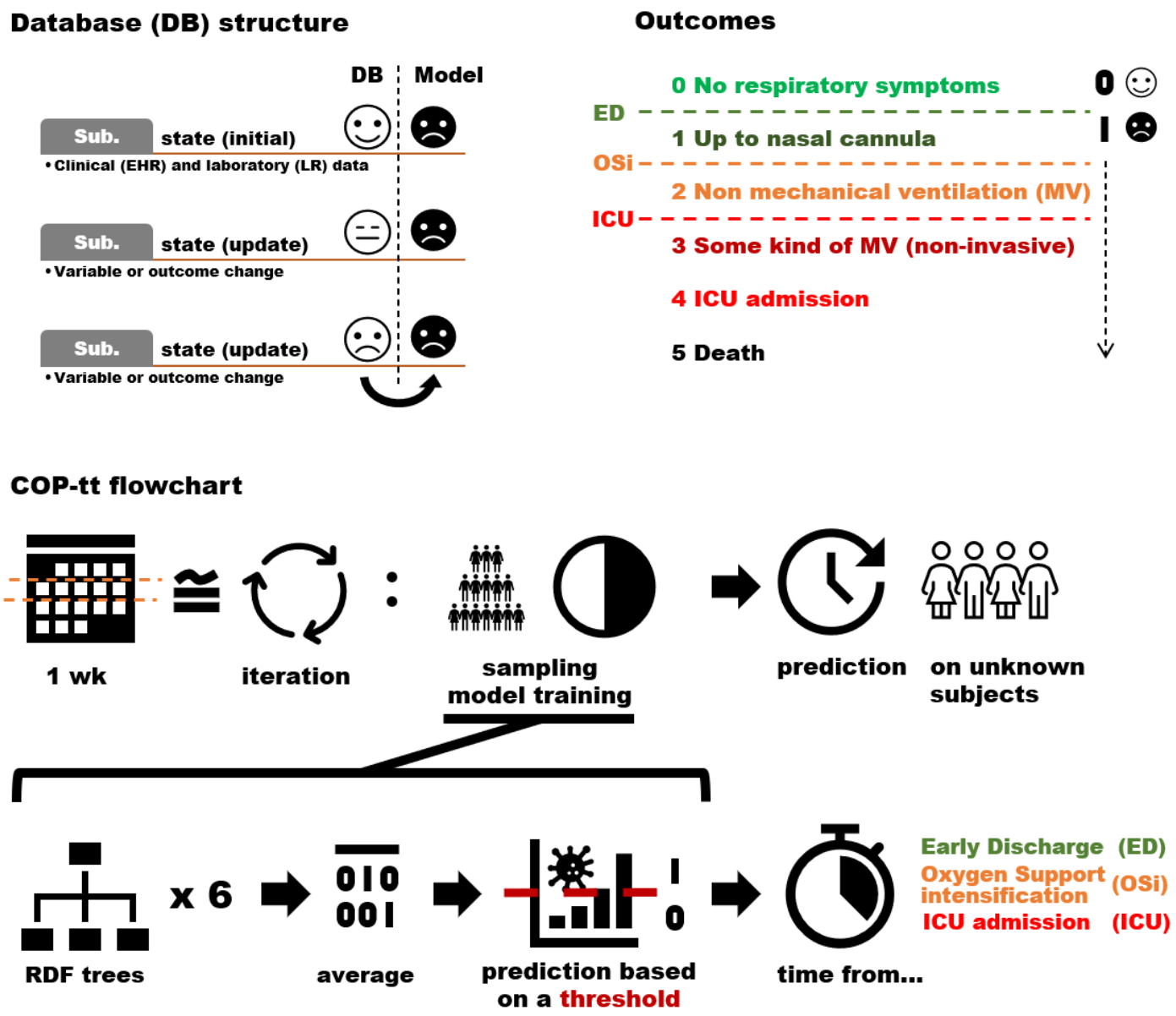

Figure 1

COP-tt database structure, outcomes and flowchart. Database structure: each subject is represented by a sequence of states that mirror disease progression; states are set to their worst outcome to make predictions. Outcomes range from no respiratory symptoms to death: since predictions were made using a binary classifier, the binary discriminator between good (0) and bad (1) outcomes was lowered depending on clinical prediction goals, from early discharge to admission to ICU. COP-tt flowchart: iterations could be set within any temporal constraints, here we imagine weekly iterations of sampling and predictions, which in a prospective study would become learning from all past subjects and prediction of new subjects. Models are assemblies of random decision forest (RDF) trees, six models were used to lower the risk of overfitting. RDF results were averaged, and the unknown subject's outcome was determined using a threshold. See methods for a detailed description.

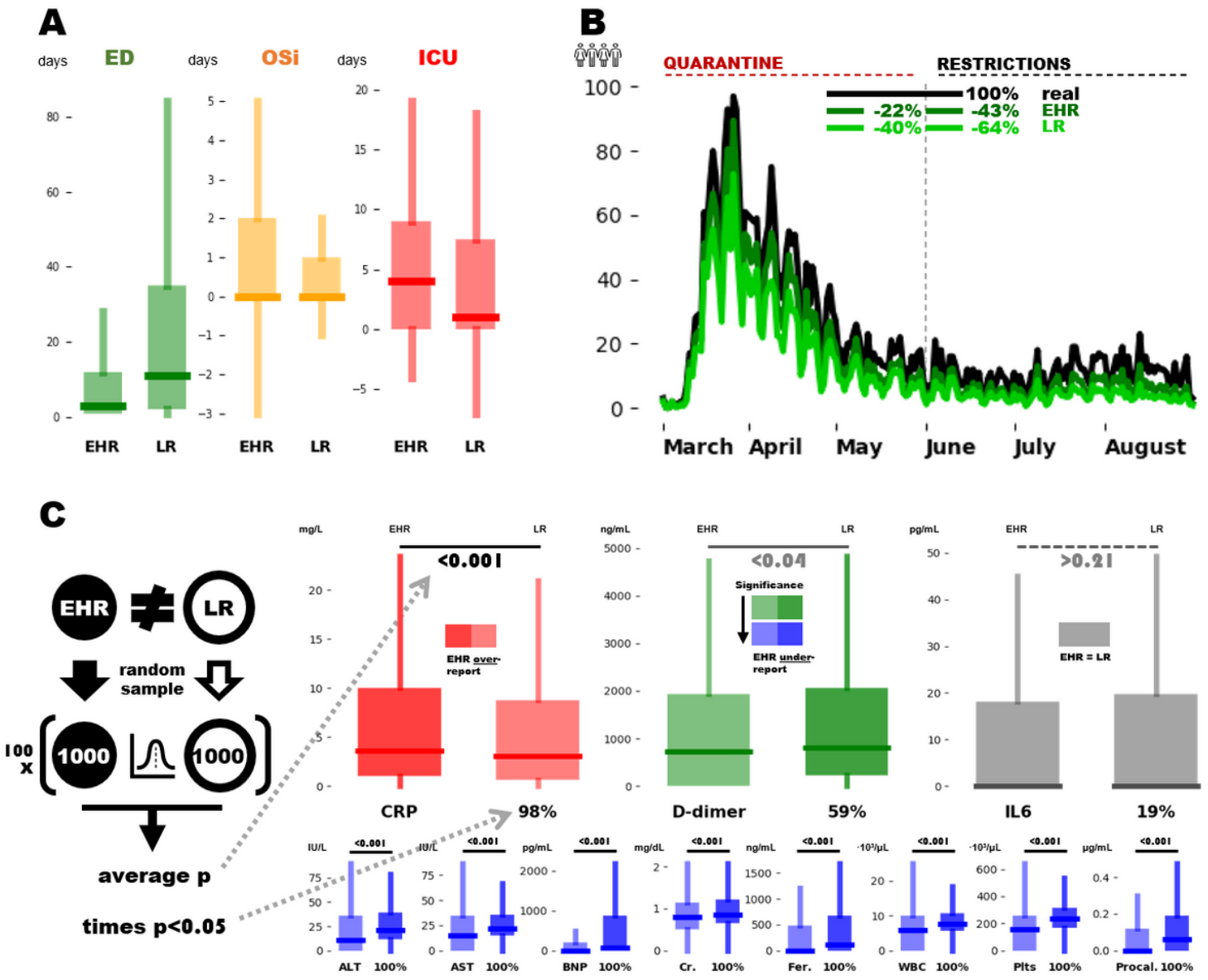

Figure 2

Anticipation, hospital occupation and model comparison. A. Anticipation in days of early discharge (ED), oxygen support intensification (OSi) and Intensive Care Unit admission (ICU). B. Occupation curve adjusted with electronic health records (EHR) or laboratory results (LR) models. C. Comparison between EHR and LR: since sets were not fully overlapping, random sampling was used for determining statistical differences. Alanine (ALT) and Aspartate aminotransferases (AST), Brain Natriuretic Peptide (BNP), Creatinine (Cr.), Ferritin (Fer.), Procalcitonin (Procal.), White Blood Cells (WBC) and Platelets (Plts) were underreported, instead the proinflammatory CRP was overreported in EHR.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- COVIDopSciRepv2Suppfinal.docx