MEDDOCS Open Access Publisher

ISSN: 2639-4391

Annals of Epidemiology & Public Health

Open Access | Research Article

Screening for Incidental Sars-Cov-2 Infection in a Neurocritical Care Unit: A Longitudinal Diagnostic Prediction Model

Jens Boss[1] †; Jan Willms[2]*†; Philipp Karl Bühler[2]; Christoph Ganter[2]; Sascha David[2]; Peter Steiger[2]; Giovanna Brandi[2]; Marko Seric[1]; Daniel Baumann[1]; Emanuela Keller[1]

[1]Neurocritical Care Unit, Department of Neurosurgery and Institute of Intensive Care Medicine, University Hospital Zurich, Switzerland.

[2]Institute of Intensive Care Medicine, University Hospital Zurich, Switzerland.

†Shared first authorship

***Corresponding Author(s): Jan Willms**

Institute of Intensive Care Medicine, University Hospital Zurich, Switzerland.

Email: janfolkard.willms@usz.ch

Received: Nov 17, 2022

Accepted: Dec 06, 2022

Published Online: Dec 08, 2022

Journal: Annals of Epidemiology and Public health

Publisher: MedDocs Publishers LLC

Online edition: http://meddocsonline.org/

Abstract

Background: Rapid diagnosis of SARS-CoV-2 infection in patients not primarily assigned with the diagnosis of COVID-19 is highly relevant to effectively rule out virus transmission among patients and medical staff.

The purpose is to develop a model for the prediction of the actual presence of a SARS-CoV-2 infection before a valid test result is available and to avoid unnecessary testing in Critical Care Units.

Methods: Datasets of laboratory and blood gas analysis tests were collected retrospectively for the development and subsequent validation of machine learning (ML) based models. The data set was composed of 1. 254 SARS-CoV-2 positive cases, collected in an ICU dedicated to patients with COVID-19 pneumonia, 2a. 914 SARS-CoV-2 negative patients treated in a Neurocritical Care Unit and 2b. 32 patients treated for severe influenza pneumonia in a Medical ICU at the same hospital. The models were subsequently validated on a dataset collected from the Neurocritical Care Unit that consisted of data from 7 positive and 42 negative patients. Models were adapted to newly available laboratory values throughout their ICU stay. Extremely Randomized Trees (ERT) and Random Forest (RF) models were evaluated. A baseline model comprising fully grown trees, an optimized model including optimal values for the maximum depth, and a simplified model that only uses the 6 most important features were trained.

Cite this article: Boss J, Willms J, Bühler Pk, Ganter C, David S, et al. Screening for Incidental Sars-Cov-2 Infection in a Neurocritical Care Unit: A Longitudinal Diagnostic Prediction Model. A Epidemiol Public Health. 2022; 5(2): 1094.

Results: The overall best model, evaluated via cross-validation on the development set, is an optimized ERT model with a ROC AUC value of 0.946. The model performance on the validation set is best for the simplified RF model achieving a ROC AUC value of 0.701. Gini feature and permutation importance for the simplified RF model revealed hemoglobin, procalcitonin, C-reactive protein, glomerular filtration rate based on CKD-EPI equation, creatinine, and urea as the most important input features. Using the simplified RF model and a threshold of 0.012 for the probability, a sensitivity above 80% with a specificity of 43% is achieved. Compared to a hypothetical daily testing regimen, using a threshold of 0.145, the simplified RF model detects all positive cases, and, with a false positive rate of 35%, daily tests might be reduced by two thirds.

Conclusions: The model developed may support the medical staff in the ICUs by enabling faster and more reliable recognition of COVID-19. Unnecessary serial test sampling might be reduced. To ensure the quality of the model before clinical use, it should be further validated in prospective patient cohorts.

Introduction

During the COVID-19 pandemic, the health care system is facing extraordinary challenges, as human resources for ICU professionals are restricted, and the restriction is becoming even more pronounced if medical staff is infected. To avoid transmission between patients and among health care workers, tests for SARS-CoV-2 based on Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) are routinely performed in patients at ICU entry and repetitively every few days thereafter. Serial testing is tedious per se and the effectiveness of such a protection scheme is limited by the varying sensitivity of the tests, which depends on the collection technique, the momentary viral load and the time passed since exposure to the virus [1]. Furthermore, in future pandemic waves or even a novel phase, daily testing for routine screening may be inefficient and low yield in the long term. This is particularly important for pulmonary asymptomatic patients with incidental SARS-CoV-2 infection, e.g. patients hospitalized for primary stroke with concomitant SARS-CoV-2 infections [2]. In these patients, who are not primarily assigned with pneumonia or being hospitalized for elective surgery, rapid diagnosis of SARS-CoV-2 infection is highly relevant to effectively rule out virus transmission among patients and medical staff.

The purpose of this study is to develop a model for the prediction of the actual presence of a SARS-CoV-2 infection before a valid test result is available and to avoid unnecessary testing in Critical Care Units.

Methods

The authors strictly adhere to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) reporting guidelines [3].

Data source

In this study, we analyzed data from a cohort of patients treated in different ICUs at the Institute of Intensive Care Medicine, University Hospital Zurich. Two separate datasets of laboratory and Blood Gas Analysis (BGA) tests were collected retrospectively for the development and subsequent validation of candidate models. The data set used during model development was composed of data collected from three different patient groups: 1. For the SARS-CoV-2 positive cases, we collected data from patients (n=254) treated in an ICU dedicated to patients with COVID-19 pneumonia from March 2020 to January 2021. COVID-19 diagnosis was confirmed by a positive RT-PCR test for SARS-CoV-2 in a throat swab. The negative reference cases comprised of two patient groups: 2a. Patients treated on the Neurocritical Care Unit from September 2016 to October 2021 (n=914), and 2b. Patients treated for severe influenza pneumonia at the Medical ICU between December 2017 and November 2020 (n=32). Influenza diagnosis was confirmed by a positive RT-PCR test for Influenza virus in a throat swab or in tracheal secretion. For patients of the negative reference groups 2a and 2b, it was ensured that they had no positive SARS-CoV-2 test result.

The models were validated on the validation dataset collected from the Neurocritical Care Unit and consisted of data from 7 positive and 42 negative patients from March 2020 to January 2022 and from November 2021 to January 2022, respectively. RT-PCR testing for SARS-CoV-2 was performed routinely at admission and on day 1, 3 and every 5 days thereafter. Notice that the Neurocritical Care Unit was the intended location of deployment of the algorithm.

Outcome

The primary outcome for the evaluation of the algorithm's performance was the contraction of a SARS-CoV-2 infection. However, detection models were trained to discriminate between all three labels SARS-CoV-2, influenza, or neither. Moreover, the goal of the detection model was not to establish the outcome just at admission of the patients but to adapt to newly available laboratory values throughout their ICU stay. So, even though the herein used datasets were curated in a way that a single patient only belonged either to the category of SARS-CoV-2 positive or negative patients, the detection algorithm that results from the training should also be able to detect patients infected during their hospital stay.

Participants

The study was approved by the local ethics committee. Written consent was given by legal representatives, as all patients were incapable of judgment.

Predictors

Only commonly available laboratory and BGA findings were considered as predictors for detection of a SARS-CoV-2 infection. Thus, after compiling laboratory and BGA results for the development dataset, the number of model input variables was narrowed down by proceeding only with laboratory and BGA results that were simultaneously available for all three patient groups (SARS-CoV-2, influenza, neither) and in more than 75% of the cases. This pre-selection process resulted in the using the following set of 22 laboratory results as predictors: Hemoglobin (HB), white blood cell count (LC), lymphocyte count (LYM), monocyte count (MON), neutrophil count (NEU), platelet count (TC), fibrinogen (FBG), International Normalized Ratio (KHINR), albumin (ALB), alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin (BIT), creatine kinase (CK), glomerular filtration rate based on CKD-EPI equation (CKDEPI), C-reactive protein (CRP), procalcitonin (PCT), creatinine (CREA), urea (UREA), sodium (Na+), kalium (K+), lactate (Lac), standard bicarbonate (SBC). As these predictors were measured repeat-

edly for each patient, this data was represented by multidimensional time series. From this data, feature vectors were sampled, *i.e.*, sets of input variables such that the feature vectors appropriately modelled the distribution of inputs if one measured them for a random patient at a random time. The latter being crucial as longitudinal models were developed. To meet the distributional requirements, we generated the feature vectors by resampling the multidimensional predictor time series using a constant interval of 4 hours and always recording the last measurement for each input and interval. This process resulted in a time-weighted distribution of feature vectors that were independent of the frequency of repeating the laboratory analyses. Only the feature vectors sampled from the first 48 hours after ICU admission were included in the training set for positive cases. For the validation dataset, we proceeded in the same way but used a sampling interval of 10 min to achieve a higher time resolution to facilitate analysis of the output time series.

Missing values

Forward filling for missing values was used by propagating the last known value for each predictor forward in time. To handle missing predictors, iterative imputation was used to fill missing values based on available values by the means of linear regression models.

Model development

Data processing was performed using Python 3.8, and in particular scikit-learn (version 1.0), which is an open-source Python module for machine learning (ML) model development and validation.

During the development phase, nested 5-fold Cross-Validation (CV) was used to train and test a set of different ML models, or, more precisely, ML pipelines. Subsequently, we used the trained models to make predictions on the validation dataset. Two types of tree ensemble models were evaluated, namely Extremely Randomized Trees (ERT) [4] and Random Forest (RF) [5] models. We restricted ourselves to tree ensembles to not rely on extensive pre-processing and normalization of the input values, which can be difficult for heavily skewed distributions of inputs, typical for laboratory findings, and which can also be a source of bias. To avoid feature leaking during CV, the data pre-processing as well as the actual ML model were specified as pipelines, which encapsulate all necessary steps to select and train a model and subsequently compute a model's output from a set of raw feature vectors. Moreover, for model training, we adjusted the sample weights such that the weighted number of samples was identical for all outputs categories/classes to effectively balance the dataset.

Model specification

For both above-mentioned model types, we trained a *baseline model* without hyper parameter tuning comprising of fully grown trees, *i.e.*, no restriction on the maximum depth of trees, an *optimized model* for which we performed a grid search to find optimal values for the maximum depth and the maximum number of considered features at each decision point in the tree, and a *simplified model* that only uses the 6 most important features. Exact model definitions can be found in the supplementary information.

Model performance metrics

Since the models were trained on all three labels (SARS-CoV-2, influenza, and neither), they also provide three probability estimates as output, namely the individual probabilities that a patient belongs to that exact category. However, to gauge model performance only the probability of contracting SARS-CoV-2 was considered relevant, i.e., turning the models into binary classifiers. Thus, the area under the receiver operating curve ROC AUC could be used as performance metric. Confidence intervals (CI) were computed for the ROC AUC values based on the (CV) predictions of the models using bootstrapping (random resampling with replacements of entire patients). To help with the interpretation of the achieved performance, Gini and permutation importance were computed for the development and validation datasets, respectively [5,6]. Additionally, a set of reference models that excluded the influenza dataset was also evaluated. Finally, the longitudinal behavior was accessed qualitatively by inspection of the predicted probabilities in timeseries charts and quantitatively by comparing the detection rate with the percentage of unnecessarily performed tests for a specific scenario with daily tests based on a threshold and the model output.

Results

Data sampling

The described methods of data sampling generated large development and validation datasets. Table 1 summarizes the number of patients stratified according to the three defined disease categories and resulting feature vectors for each category and the two datasets used for development and validation of the models.

Table 1: Number of patients and feature vectors (samples) per output class for the development and validation sets.

	Development set (sampling time 4 h)		Validation set (sampling time 10 min)	
	Number of patients (n 1200)	Number of samples	Number of patients (n 49)	Number of samples
SARS-CoV-2 positive	254	2'992	7	5'623
Influenza positive	32	3'296	--	--
Negative	914	44'687	42	40'077

Missing values

Restricting our analysis to commonly available laboratory and BGA findings, limited the number of missing values in the sampled feature vectors. Figure 1 shows a comparison of the availability of input values for the different output categories in the development set (availability of input values in the validation set is given in Figure A1 as supplemental material). Notice that the availabilities are time-weighted due to the used sampling scheme. If the availability falls below 75%, this means that this variable was often measured only later in a patient's stay on the ICU.

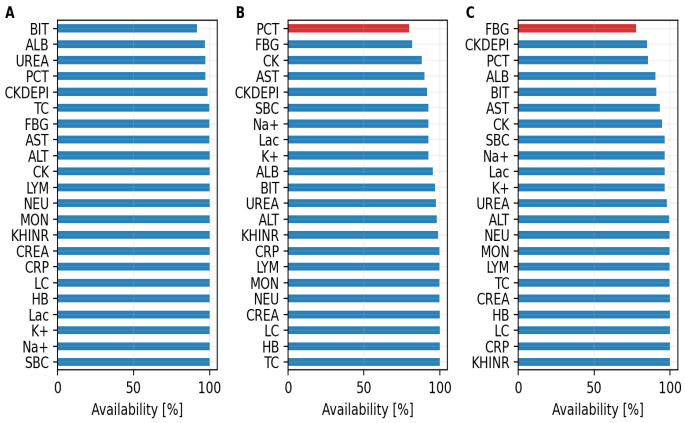

Figure 1: Overview of the availability of input data in the development set. Bar plots depicting the availability of the individual predictors in the feature vectors for the development set of the three output categories. The panels **A**, **B**, and **C** show the availability in the categories "negative", "influenza positive", and "SARS-CoV-2 positive", respectively. A red bar indicates an availability in less than 75% of the feature vectors.

Model development

12 tree ensemble models have been trained in total, half of them were reference models excluding the influenza data. To make comparison easier, all models were based on 200 decision trees, and sample weights were always adjusted to balance the dataset. For the baseline models we kept the default parameters for all model parameters except from the number of base estimators, i.e., decision trees that were used.

For the optimized models we performed a grid search to find the optimal maximum tree depth of the decision trees and the optimal number of random features considered for choosing the decision criterion at the nodes of the trees. Table 2 summarizes properties of the best estimators. After cross-validation, we fitted a final model to the entire development set, which was subsequently used for validation.

Similarly, we trained the simplified models for which the feature importance was used to select the 6 most important features based on optimized models. Notably, during each step of the cross-validation, an optimized model was fitted to the training set, which internally used cross-validation to optimize itself. Selected features for different models are listed in Table 2. Again, final models were fitted after cross-validation.

Model performance

Performance was evaluated by cross-validation on the development set and making predictions on the validation set. For the validation, we re-trained each model on the entire development set. Thus, cross-validation results quantify the performance of the whole model building processes, whereas the validation measures the performance of the single final model. Table 3 summarizes the evaluation results including confidence intervals. It should be noted, however, that due to the small sample size of the validation set, the corresponding confidence intervals are expected to exhibit a limited accuracy. Model performance is generally high when tested on the development set, while optimization of the model parameters does not significantly improve performance. The overall best model with a ROC AUC value of 0.946 ± 0.022 is the baseline ERT model without the inclusion of the influenza data. The performance on the validation set is worse. In general, it can be asserted that the ERT and RF models perform similarly, with the ERT models performing slightly better on the validation set except for the simplified models. Moreover, only the simplified models benefit from the inclusion of the influenza data boosting their ROC AUC value on the validation set. Consequently, the simplified RF model (with influenza data) is the best performing model achieving a ROC AUC value of 0.701.

Table 2: Model selection results: The fitting of the optimized and simplified models involved hyper-parameter search as well as feature selection. The hyper-parameters of tuned models are summarized for the main and reference models including and excluding the influenza data, respectively.

	Max. depth		Max. features		Selected predictors	
Including influenza	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
ERT baseline model	-	-	4	4	all	all
ERT optimized model	32	32	8	4	all	all
ERT simplified model	128	8	2	2	CKDEPI, CRP, HB, UREA, CREA, MON	CKDEPI, CRP, HB, UREA, CREA, MON
RF baseline model	-	-	4	4	all	all
RF optimized model	256	256	4	4	all	all
RF simplified model	16	4	2	2	CKDEPI, CRP, HB, UREA, CREA, PCT	CKDEPI, CRP, HB, CREA, MON, PCT

Table 3: Model performances: Performance metrics of different models measured as the area under the receiver operating characteristic curve (ROC AUC). 5-fold cross-validation (CV) was used to compute the mean and standard deviation of the metrics on the development set. Fitted models were subsequently used to detect SARS-CoV-2 infection in the validation set. Best-performing models are highlighted in bold. Confidence intervals (CI) are shown in brackets. For the development set, we also state the average ROC AUC values as computed from CV predictions.

	ROC AUC for development set		ROC AUC for validation set	
Including influenza	*Yes*	*No*	*Yes*	*No*
ERT baseline model	0.940±0.007 (0.940; 95% CI: [0.936,0.956])	**0.946±0.022** (0.947; 95% CI: [0.942,0.961])	0.669 (95% CI: [0.613,0.825])	0.656 (95% CI: [0.603,0.817])
ERT optimized model	0.944±0.006 (0.944; 95% CI: [0.939,0.957])	0.944±0.023 (0.945; 95% CI: [0.939,0.961])	0.636 (95% CI: [0.584,0.802])	0.684 (95% CI: [0.630,0.847])

ERT simplified model	0.900±0.018 (0.899; 95% CI: [0.893,0.918])	0.906±0.031 (0.853; 95% CI: [0.844,0.880])	0.662 (95% CI: [0.611,0.821])	0.632 (95% CI: [0.582,0.788])
RF baseline model	0.939±0.010 (0.939; 95% CI: [0.934,0.954])	0.944±0.019 (0.944; 95% CI: [0.939,0.958])	0.648 (95% CI: [0.588,0.808])	0.643 (95% CI: [0.592,0.809])
RF optimized model	0.939±0.011 (0.939; 95% CI: [0.933,0.952])	0.942±0.020 (0.943; 95% CI: [0.938,0.956])	0.636 (95% CI: [0.577,0.812])	0.625 (95% CI: [0.568,0.773])
RF simplified model	0.887±0.024 (0.871; 95% CI: [0.862,0.893])	0.904±0.034 (0.904; 95% CI: [0.897,0.923])	**0.701** (95% CI: [0.658,0.832])	0.612 (95% CI: [0.559,0.782])

The ROC curves for the best performing simplified RF model and the analogous ERT model are shown in **Figure 2** for the development and validation datasets (ROC curves for all models are given in Figure A2 in the supplemental materials). The predicted probabilities for the samples in the development set were computed during cross-validation. We found that one could simultaneously achieve a sensitivity above 80% and specificity of 43% using the simplified RF and a threshold for the probability of 0.012 as an example.

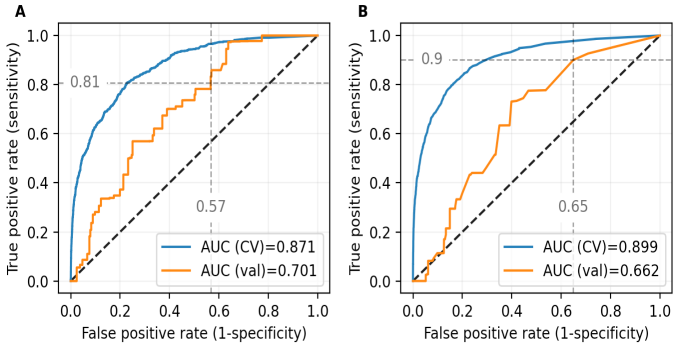

Figure 2: Comparison of ROC curves for the simplified RF and ERT models. Panel **A** shows the ROC curves of the simplified RF model computed based on the predicted probabilities for the development dataset (using cross-validation) and the validation set in blue and in orange, respectively. Panel **B** shows the corresponding plot for the simplified ERT model. The false positive rate is indicated in gray for the lowest true positive rate exceeding 0.8 on the validation set.

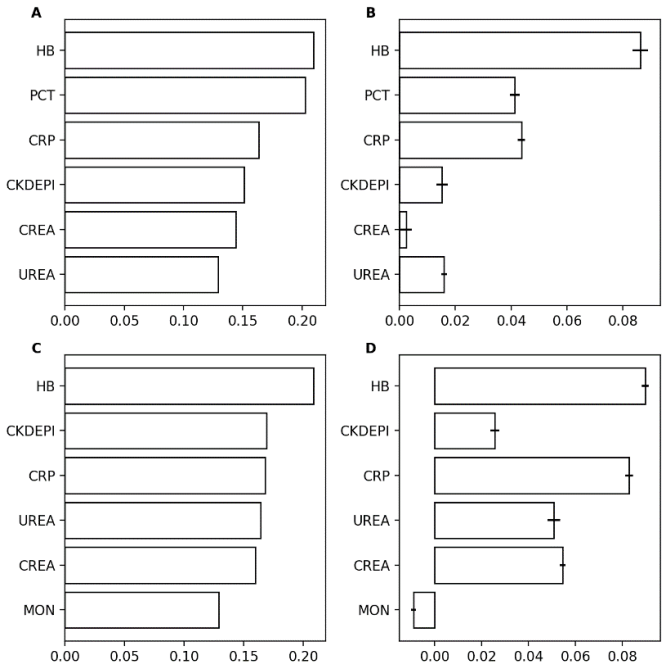

Figure 3: Gini feature importance and permutation importance for simplified models. Panels **A** and **C** depict the Gini feature importance for the simplified RF and ERT model, respectively. Panel **B** and **D** show the corresponding permutation importance for the RF and the ERT model, respectively. The computation of permutation importance was repeated 5 times, the vertical lines at the ends indicate the standard deviation between the repetitions. (Gini feature importance and permutation importance for all models are given in Figures A3 to A6 in the supplemental material.)

To further investigate the performance degradation of the models on the validation set, Figure 3 shows the Gini feature importance obtained on the development set in direct comparison with the permutation importance calculated using the validation set. The Gini feature importance reflects the importance of features learnt from the development set by computing the mean decrease of the impurity between subsequent tree nodes. The permutation importance, on the other hand, reflects the feature importance when making prediction on the validation set. The Gini feature importance is distributed relatively even between all features for both models with HB, CKDEPI, and CRP being among the most important ones. A clear difference between the models is the use of PCT in the RF model, an input that has not been selected by the ERT model. Moreover, HB and CRP also remain important when making predictions on the validation set, whereas CKDEPI as well as UREA become less relevant, which could explain the performance degradation we have seen. Interestingly, PCT is also an important feature for making predictions on the validation set, which could explain why the RF model performs better than the ERT model, which does not consider it.

Figure 4 visualizes the outputs of the best performing model for all positive cases to qualitatively assess the model performance and stability. Stability means that the output of a model does not change radically between subsequent sets of laboratory results. Figure 4 shows that testing was conducted most of the time after the model indicated an infection.

To estimate potential benefits of implementing the model in clinical practice, a hypothetical scenario is analyzed. In this scenario, instead of regular daily testing, it is assumed that a SARS-CoV-2 test is performed as soon as the model output exceeds a certain threshold but at most once a day. Based on this scenario, the fraction of positive cases that would get tested at least once during their stay in the ICU as well as the percentage of patients that would get tested unnecessarily every day have been determined simultaneously. Further, assuming a low prevalence, the latter can be approximated by the percentage of negative patients that would get tested each day. Figure 5 shows the result of this analysis. Using a threshold of 0.145, all positive cases can still be detected, but daily tests would be reduced by two thirds.

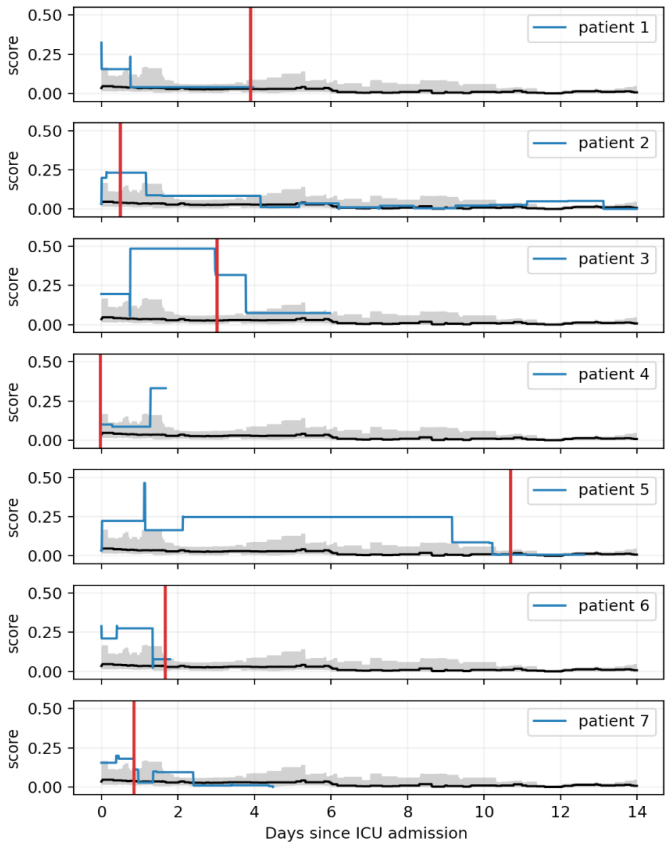

Figure 4: Time series of predicted scores for positive patients in the validation set. The different panels depict the output of the simplified RF in blue. The red vertical line indicates the time at which the SARS-CoV-2 infection was confirmed by a PCR test. The grey area and the black line indicate the interquartile range and the median of the model outputs for negative cases in the validation set.

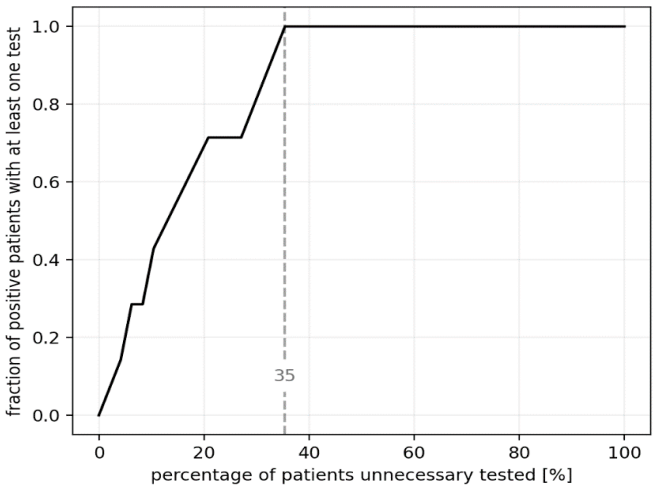

Figure 5: Comparison of detection of positive cases and unnecessary testing: The plot shows the fraction of patients with at least one test (which we assume has a positive result) as a function of the percentage of patients tested each day. In gray, we indicate the lowest percentage of patients that need to be tested to find all positive cases (threshold=0.145).

Discussion

With the present project, dynamic detection models based on daily collected laboratory values for SARS-CoV-2 infection have been developed. The overall best model, evaluated via cross-validation on the development set, is an optimized Extremely Randomized Trees model with a ROC AUC value of 0.946 ± 0.022. Models have been validated on a left-out-dataset with patients treated on the Neurocritical Care Unit, where the model is intended to be implemented in the future. The performance of all models is significantly worse on the validation data set compared to development set. The model performance on the validation set was best for the simplified Random Forest model achieving a ROC AUC value of 0.701. Using the simplified RF model and a threshold for the probability of 0.012, after all, a sensitivity above 80% with a specificity of 43% could be achieved, which qualifies the model as a screening tool in clinical practice. Due to the high sensitivity, only a few patients with COVID-19 might be missed, and due to a 43% false positive rate, unnecessary testing could be significantly reduced. Universal screening tests of patients admitted to healthcare facilities is performed in many emergency departments [7] and has been advocated especially in woman admitted for delivery [8]. A routine screening may be even more justifiable in patients admitted to a Neurocritical Care Unit, mostly suffering from impaired mental state or consciousness so that neither the vaccination status nor a history of COVID-19 like symptoms can be elicited. High costs or reduced testing capacities, however, may impair routine screening in the long term or render it inefficient with a low yield in a future endemic phase. Compared to a hypothetical daily testing regimen using a threshold of 0.145, the simplified RF model detects all positive cases, and, with a false positive rate of 35%, daily tests could be reduced by two thirds. A predictive model to identify patients with a high probability of SARS-CoV-2 infection – in the sense of an alarm system-may allow more specific testing, reduce the burden of repetitive testing for individual patients, medical personnel and laboratories, and save costs.

In a living systematic review by the COVID-PRECISE group, 33 multivariable models to distinguish between patients with and without COVID-19 have been identified [9] (https://www.covprecise.org/living-review/). However, all models resulted in a single time point prediction (snapshot situation) in contrast to our model assessing the specific patient risk daily on longitudinal data. The actual model developed is adapted to the newly available laboratory values throughout the ICU stay. This longitudinal application of the algorithm will make it especially useful in patients during the entire ICU stay. Furthermore, in contrast to other models, patients treated for severe influenza pneumonia were included to train the model to reduce the tendency of potential models to detect viral pneumonia in general instead of COVID-19. By comparison with reference models disregarding data of influenza patients during training [9], it could be shown that including this data significantly improves the performance of the simplified models on the validation set.

Gini feature and permutation importance for the simplified RF model revealed HB, PCT, CRP, CKEPI, CREA and UREA as the most important input features, which seems comprehensible. In a meta-analysis including data from 57,563 COVID-19 patients, hemoglobin levels were lower in patients admitted to intensive care units [10]. Quite specific for COVID-19, the SARS-CoV-2 virus attacks the heme group of hemoglobin delivering toxic iron leading to high ferritin values and anemia [11]. Inflammatory immune responses with elevated CRP and PCT are known to be strongly associated with severe COVID-19 [12,10]. As a result, patients with chronic kidney disease are disproportionally affected by COVID-19, while the course of illness in patients with severe COVID-19 is frequently complicated by acute kidney failure [13] which explains the values reflecting renal insufficiency.

The study has several limitations concerning the data set, model development and its performance assessment. The number of included patients for the different datasets was limited by the availability of the data, i.e., the training of the models could benefit from a larger and more representative sample of input variables. The small data set of SARS-CoV-2 positive patients, furthermore, makes the model susceptible to over fitting. As for other diagnostic models, a case-control sampling was used and the characteristics in the development population extracted from the Medical ICU may be different from the target population at the Neurocritical Care Unit. Patients at the Medical ICU may not be representative of the model's target population at the Neurocritical Care Unit, which implies a relevant risk of bias. It is crucial that the training set is representative for the clinical context where the algorithm is intended to be implemented [14]. Thus, over time, model performance could greatly benefit from regular re-training the models on new data of positive and negative cases collected at the Neurocritical Care Unit as the target ICU for implementation. Neither epidemiologic data nor patient characteristics or clinical findings indicative for COVID-19 were included for model training. The inclusion of features as age, vaccination status, comorbidities, flu-like symptoms, together with physiological features like fever, respiratory rate, blood pressure and heart rate etc. may most likely improve the predictive performance. As the prevalence of COVID-19 in the training dataset does not represent a real word scenario, we could only study true positive and false negative rates when assessing model performance or work with assumptions as we have when analyzing the practical scenario. Due to the large number of negative reference cases relative to the other two categories, the development set is unbalanced. However, to include this large amount of negative reference cases was intentional to establish a strong baseline from which the models needed to learn to distinguish from. Moreover, the dataset was balanced by adjusting the samples weight to avoid producing biased models. Finally, generalizability is not given as the model has not been validated in external datasets by independent investigators.

To conclude, the model developed might support the medical staff in the ICUs by faster recognition of COVID-19. Unnecessary serial test sampling which burdens patients, medical personnel, and laboratories may be reduced. The model might be helpful especially in regions with limited test capacities and in triaging patients when allocating hospital resources. To ensure the quality of the model before clinical use, it should be further validated in prospective patient cohorts. Changing courses of the SARS-CoV-2 pandemic require adaptations of the model by re-calibrations of the baseline risk at close time intervals.

Declarations

Ethical Approval

We confirm that this study was approved by the local ethics committee (Kantonale Ethikkomission, Kanton Zürich, Stampfenbachstrasse 121, CH- 8090 Zürich; BASEC-Nr. 2016–01101). Written consent was given by the patients or by their legal representatives.

Competing interests: The authors declare no competing interests.

Authors' contributions

JB., JW. and EK. Conceived the study and wrote the first draft; JB., JW, GB., EK., PKB., CG., SD. and PS. selected the population and revised the text for intellectual content; JB., EK., MS., JW. and DB. collected the data and and conducted the statistical analysis. All authors read and approved the final manuscript.

Funding

The study was supported by the Swiss Innovation Agency, grant number 52441.1 IP-LS

Availability of data and materials

All data generated or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Kucirka LM, Lauer SA, 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; 173: 262-267.

2. Oxley TJ, Mocco J, Majidi S, Kellner CP, Shoirah H, et al. Large-Vessel Stroke as a Presenting Feature of Covid-19 in the Young. N Engl J Med. 2020.

3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg. 2015; 102: 148-158.

4. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006: 3-42.

5. Breiman L. Random Forests. Machine Learning. 2001; 45: 5-32.

6. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: A corrected feature importance measure. Bioinformatics. 2010; 26: 1340-1347.

7. Sheele JM, Lalljie AV, Fletcher S, Heckman M, Hochwald A, et al. Ability of emergency medicine clinicians to predict COVID-19 in their patients. Am J Emerg Med. 2021.

8. Sutton D, Fuchs K, D'Alton M, Goffman D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. N Engl J Med. 2020; 382: 2163-2164.

9. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328.

10. Taneri PE, Gomez-Ochoa SA, Llanaj E, Raguindin PF, Rojas LZ, et al. Anemia and iron metabolism in COVID-19: a systematic review and meta-analysis. Eur J Epidemiol. 2020; 35: 763-773.

11. Liu W, Hualan L. COVID-19:attacks the 1-Beta chain of hemoglobin and captures the porphyrin to inhibit human Heme metabolism.

12. Lee EE, Song KH, Hwang W, Ham SY, Jeong H, et al. Pattern of inflammatory immune response determines the clinical course and outcome of COVID-19: unbiased clustering analysis. Sci Rep. 2021; 11: 8080.

13. Bruchfeld A. The COVID-19 pandemic: consequences for nephrology. Nat Rev Nephrol. 2021; 17: 81-82.

14. Yu M, Tang A, Brown K, Bouchakri R, St-Onge P, et al. Integrating artificial intelligence in bedside care for covid-19 and future pandemics. BMJ. 2021; 375: e068197.