OPEN ACCESS

Detection of patients with COVID-19 by the emergency medical services in Lombardy through an operator-based interview and machine learning models

Stefano Spina [1,2] Lorenzo Gianquintieri,[1,3] Francesco Marrazzo,[1,2] Maurizio Migliari,[1,2] Giuseppe Maria Sechi,[1] Maurizio Migliori,[1] Andrea Pagliosa,[1] Rodolfo Bonora,[1] Thomas Langer,[2,4] Enrico Gianluca Caiani,[3] Roberto Fumagalli [1,2,4] AREU 118 EMS Network Collaborators

Handling editor Shammi L Ramlakhan

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/emermed-2022-212853).

[1]SOREU, Agenzia Regionale Emergenza Urgenza (AREU), Milano, Italy
[2]Department of Anesthesia, Critical Care and Pain Medicine, ASST Grande Ospedale Metropolitano Niguarda, Milano, Italy
[3]Electronics, Information and Biomedical Engineering Department, Politecnico di Milano, Milano, Italy
[4]Department of Medicine and Surgery, University of Milan-Bicocca, Monza, Italy

Correspondence to
Professor Roberto Fumagalli, University of Milan-Bicocca Faculty of Medicine and Surgery, Monza, Lombardia, Italy; roberto.fumagalli@unimib.it

SS, LG and FM contributed equally.

Received 19 September 2022
Accepted 24 August 2023

Check for updates

To cite: Spina S, Gianquintieri L, Marrazzo F, *et al*. *Emerg Med J* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/emermed-2022-212853

ABSTRACT

Background The regional emergency medical service (EMS) in Lombardy (Italy) developed clinical algorithms based on operator-based interviews to detect patients with COVID-19 and refer them to the most appropriate hospitals. Machine learning (ML)-based models using additional clinical and geospatial epidemiological data may improve the identification of infected patients and guide EMS in detecting COVID-19 cases before confirmation with SARS-CoV-2 reverse transcriptase PCR (rtPCR).

Methods This was an observational, retrospective cohort study using data from October 2020 to July 2021 (training set) and October 2021 to December 2021 (validation set) from patients who underwent a SARS-CoV-2 rtPCR test within 7 days of an EMS call. The performance of an operator-based interview using close contact history and signs/symptoms of COVID-19 was assessed in the training set for its ability to determine which patients had an rtPCR in the 7 days before or after the call. The interview accuracy was compared with four supervised ML models to predict positivity for SARS-CoV-2 within 7 days using readily available prehospital data retrieved from both training and validation sets.

Results The training set includes 264 976 patients, median age 74 (IQR 55–84). Test characteristics for the detection of COVID-19-positive patients of the operator-based interview were: sensitivity 85.5%, specificity 58.7%, positive predictive value (PPV) 37.5% and negative predictive value (NPV) 93.3%. Contact history, fever and cough showed the highest association with SARS-CoV-2 infection. In the validation set (103 336 patients, median age 73 (IQR 50–84)), the best-performing ML model had an AUC of 0.85 (95% CI 0.84 to 0.86), sensitivity 91.4% (95 CI% 0.91 to 0.92), specificity 44.2% (95% CI 0.44 to 0.45) and accuracy 85% (95% CI 0.84 to 0.85). PPV and NPV were 13.3% (95% CI 0.13 to 0.14) and 98.2% (95% CI 0.98 to 0.98), respectively. Contact history, fever, call geographical distribution and cough were the most important variables in determining the outcome.

Conclusion ML-based models might help EMS identify patients with SARS-CoV-2 infection, and in guiding EMS allocation of hospital resources based on prespecified criteria.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ There have been several risk tools created to determine either the presence of SARS-CoV-2 or its likely course; however, there is little information about the identification of these patients in the prehospital phase of care.
⇒ The use of machine learning (ML) algorithms in the prehospital context has been limited to specific conditions, such as the recognition of out-of-hospital cardiac arrest and predicting the need for critical care resources.

WHAT THIS STUDY ADDS

⇒ Using retrospective data from operator-based telephone interviews by emergency medicine services, several variables were sensitive for identifying patients who later tested positive for SARS-CoV-2.
⇒ However, an ML model based on contact history, clinical parameters, geographical data and local epidemiology had greater sensitivity in detecting SARS-CoV-2 infection.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ ML models may guide emergency medical service in detecting COVID-19 cases before confirmation with SARS-CoV-2 rtPCR results and could be useful in other pandemic outbreaks to allow appropriate isolation and referral to dedicated hospital resources.

INTRODUCTION

The SARS-CoV-2 pandemic has been spreading worldwide over the last 4 years and the continued emergence of new viral variants has put a strain on public health systems.[1] As with other prehospital providers, the emergency medical service (EMS) in the Lombardy region (Italy) was challenged by a remarkable increase in calls directed to its public safety answering points (PSAP) since the first COVID-19 outbreak.[2] Detecting COVID-19 cases has been crucial to directing these patients to dedicated hospital resources while guaranteeing other routine EMS activity.

A strategic plan based on multiple clinical algorithms was implemented by the Agenzia Regionale Emergenza Urgenza (AREU) to manage the escalating volume of calls, control ambulance allocations and ultimately avoid EMS collapse.[3]

Both operators and healthcare professionals working in PSAP use these algorithms to identify these individuals based on signs and symptoms severity as well as identifying early indicators of new local viral outbreaks in the Lombardy region.[4 5]

However, it is still uncertain which signs and symptoms obtained at the prehospital level are most predictive of SARS-CoV-2 infection and no large-scale analysis has been done to assess the accuracy of EMS-collected clinical data in the determination of SARS-CoV-2 as confirmed by reverse transcriptase PCR (rtPCR). Furthermore, machine learning (ML) models have shown promise in predictive tasks during the COVID-19 outbreak. However, this has been limited to ED cohorts rather than in the prehospital setting.[6 7]

In this study, we assessed the performance of the clinical algorithms currently used in our PSAP (ie, operator-based interview) to identify patients that will test positive on SARS-CoV-2 rtPCR. We also aimed to develop an ML model to predict SARS-CoV-2 rtPCR positivity, based on clinical and geospatial data obtained during the PSAP call and provided by the ambulance report on the scene. We evaluated such models: (1) as a screening test to detect positive patients and (2) as a support decision tool to guide patient' allocation in a real-world scenario. We hypothesised that an ML model could achieve a better performance than clinical interviews in detecting cases of COVID-19 in the prehospital setting.

METHODS

Study design and participants

This observational, retrospective cohort study included all patients managed by the AREU EMS in Lombardy, Italy from October 2020 to July 2021 (training set) and from October 2021 to December 2021 (validation set). Patients who received assistance by the regional EMS were included if a result of the SARS-CoV-2 rtPCR was available in the timeframe of 7 days before or after ED admission, regardless of whether their index EMS call was for COVID-19-related symptoms or not.

Setting

The AREU is responsible for the EMS in the Lombardy region (Italy), covering a population of almost 10 million people in an area of about 24 000 km^2. A primary-level PSAP is the first recipient of 1-1-2 phone calls from citizens asking for police, fire or medical assistance (ie, the equivalent of 9-1-1 or 9-9-9 systems used in other countries). When medical assistance is required, callers are redirected to a secondary-level PSAP (PSAP-2), which manages all regional EMS resources.[3]

During pandemic surges, hospitals were designated by the regional public health authorities for COVID-19 treatment through a hub-and-spoke model based on severity.[8] Patients screened as non-COVID-19 were allocated elsewhere (such as trauma or stroke centres), however, each hospital had a COVID-19 and non-COVID-19 pathway in its ED, based on SARS-CoV-2 testing results.

Data sources

The following variables were retrieved and analysed:

- *General*: unique identifiers for ambulance mission and for individual patients, date, administrative area where the event occurred, caller's Global Positioning System coordinates, classification of the cause of the event requiring intervention, gender and age of the patient, admitting ED.
- *Operator-based interview*: binary answers to questions asked by operators at the PSAP-2: close contact with a person who tested positive for SARS-CoV-2, complaining of or audible shortness of breath, presence of fever, vomiting, diarrhoea, cough and/or other cold-like symptoms, ageusia/anosmia, asthenia and/or diffuse pain. The caller was considered a suspected case for COVID-19 if she/he reported one or more of these signs and symptoms.
- *Clinical parameters*, retrieved by nurses or physicians from on-scene ambulance reports: mental status ("Alert, Verbal, Pain, Unresponsive (AVPU) score), RR, oxygen saturation in room air (SpO_2), respiratory quality (normal or distress), HR, systolic and diastolic BP and temperature.
- *Daily report of SARS-CoV-2 positivity rate in the Lombardy region*, computed as an average over the previous 5 days (data source: Protezione Civile repository, https://github.com/pcm-dpc/COVID-19).
- *SARS-CoV-2 rtPCR testing:* positive or negative result ±7 days from the EMS call.[9]

Machine learning model development

The initial dataset was preprocessed by removing records with missing and outlier values (detected by the z-index method with a threshold set to five), and deleting the variable 'RR', as it was reported in only half of the records. Categorical variables were converted into dummy numerical values, and all variables were scaled in the 0–1 range.[7]

We implemented four supervised learning models to predict the positivity for SARS-CoV-2 on the rtPCR test (ie, the gold standard), the target variable for all models (figure 1). Results were evaluated with a 10-fold cross-validation protocol: the entire available dataset was divided into 10 subsets, and each of them was used once to validate a model trained on the other 9 subsets, with a final evaluation based on the distribution of the metrics across the different iterations.[7] For each model, we tested four different algorithms (ie, logistic regression, random forest classifier, support vector machine and Gaussian Naïve Bayes). The different explanatory variables were included in the models as follows:

- Model 1: age, gender and variables retrieved by the operator-based interview.
- Model 2: as for model 1, plus clinical parameters retrieved by healthcare professionals from the on-scene ambulance.
- Model 3: variables in model 1, plus the current SARS-CoV-2 epidemiology in the Lombardy region, and the geographical distribution of EMS calls for respiratory and infectious diseases in the previous 7 days.[4]
- Model 4: all variables used in models 1–3.

Additional information regarding each model development is reported in online supplemental methods and online supplemental figure 1.

A further ML model was developed to simulate a real-world application. Specifically, the model was implemented to support the EMS decision-making capability to allocate patients to the appropriate hospital, based on specific criteria such as the patient's clinical condition and her/his SARS-CoV-2 positivity. Here, an iterative procedure was implemented using historical data, repeating the whole process on every week of records for

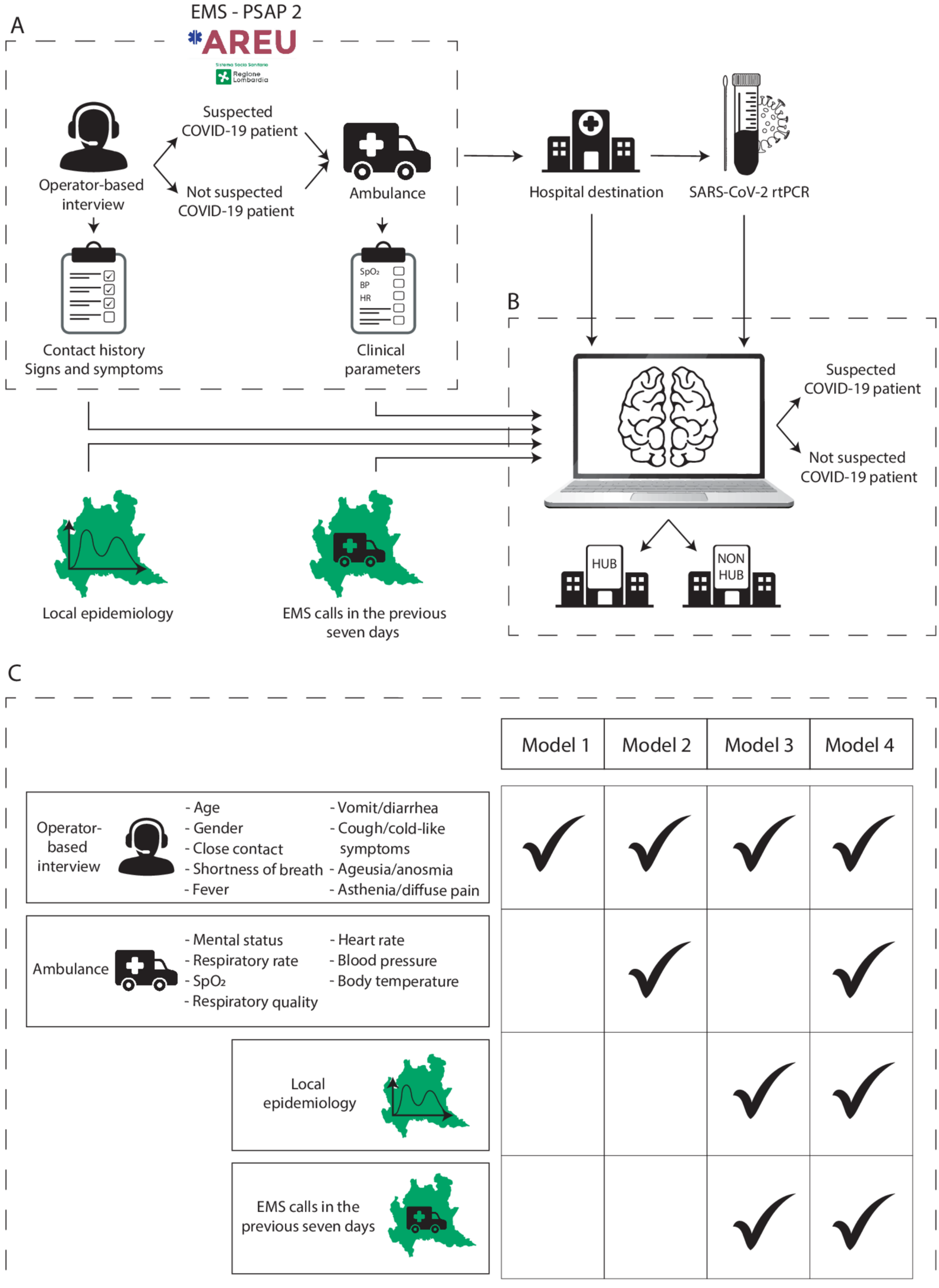


Figure 1 Model description. The performance of the operator-based interview in detecting patients with COVID-19 is evaluated by matching the results of the available SARS-CoV-2 rtPCR (box A). The machine learning models are implemented considering different combinations of the variables of the operator-based interview, the clinical parameters provided by on-scene ambulances, the local epidemiology and the distribution of EMS calls in the previous 7 days. The ultimate goal of the models is to detect cases of COVID-19. An additional model is also tested in two scenarios that could be used to guide the decision to refer patients to the proper hospital destination, based on prespecified criteria (box B). The explanatory variables included in each model are reported in the table (box C). rtPCR, reverse transcriptase PCR; AREU, Agenzia Regionale Emergenza Urgenza; EMS, emergency medical service; PSAP-2, secondary-level public safety answering point; SpO_2, pulse oximeter oxygen saturation.

a total of 38 cycles. Additional information regarding this model development is provided in online supplemental methods and online supplemental figure 2.

The first analysis included all patients in the low prevalence period (online supplemental table 1). Here, we assumed that positive patients should be allocated to hub hospitals and negative patients to non-hub hospitals. A second analysis included only patients presenting with severe features (ie, $SpO_2<94\%$ or RR >30) at EMS calls in the high prevalence period (online supplemental table 1). Therefore, in the latter, positive patients would be addressed to hub hospitals, whereas negative patients would be addressed to non-hub hospitals. Therefore, the two analyses assessed the model's capability to address patients to the appropriate hospital, which was the ultimate outcome of the model.

Statistical analysis

Continuous variables were expressed as median (IQR) and categorical variables as count (n) and percentage (%). Sensitivity, specificity, positive predictive value (PPV) and negative predicted value (NPV) were calculated to quantify the performance of the variables collected through the operator-based interview as compared with the rtPCR gold standard. In order to study the variability of the operator-based interview performance among the different phases of the pandemic, the dataset was divided into quartiles. The official daily number of positive patients in the Lombardy region was retrieved for the entire period and filtered with a 7-day window moving average, with each day assigned to one of the four quartiles accordingly. Four different datasets were thus obtained, each one reporting the records that occurred on all days belonging to the same quartile of SARS-CoV-2 prevalence in the territory (online supplemental table 1 and online supplemental figure 3).

To assess the importance of clinical variables, univariate and multivariate logistic regression models were implemented including predictor variables retrieved by the operator-based interview (alone) and with the on-scene ambulance report (ie, clinical parameters). The OR, 95% CI and C-statistics were calculated. A two-sided p value of <0.05 was considered statistically significant.

To assess the performance of the ML model in the training set, receiver operating characteristic curves were plotted and the area under the curve (AUC) was calculated. Sensitivity, specificity, PPV, NPV and accuracy were also calculated for the ML-based model at a fixed cut-off with a sensitivity target threshold of 90% (95% CIs were estimated with the Clopper-Pearson method, considering the median values across five cycles of 10-fold cross-validation).[10] In order to assess the contribution of each variable to our models, Shapley additive explanations (SHAP) were applied.[11] This method builds on the game-theory approach to explain the results of ML models.

The performance of each ML model was additionally tested on a validation dataset, independent of the training set. A detailed assessment of the real-world simulation model is described in online supplemental methods and online supplemental figure 2.

Data were first collected in regionally developed software for computer-aided dispatch (Emma, V.6.8.5, Beta80 Group, Milan, Italy) and exported using SAS Web Report Studio V.4.4 M4 (SAS Institute, Cary, North Carolina, USA). Data analysis and model implementation were performed with Python (V.3.9); the libraries used are provided in the online supplemental methods. Quartiles distribution was performed with MatLab (V.2018b). Call distribution in the Lombardy region was performed with QGIS (V.3.4.6). All model scripts used in the analysis are publicly available on GitHub (https://github.com/LGpolimi/Detection-of-patients-with-COVID-19/tree/master/env/COVID_DIAGNOSIS_MODEL).

RESULTS

Baseline characteristics

The AREU managed 684 481 ambulance dispatches from October 2020 to July 2021 (training set), of which 549 755 were transported to a regional hospital. Of these, 264 976 (48.2%) patients had SARS-CoV-2 rtPCR tests performed within 7 days prior to (n=40 731, 15.4%) or after (n=224 245, 84.6%) their EMS call and were included in the training set. Median age was 74 (IQR 55–84) years, 127 215 (48%) were female and 59 526 (22.5%) tested positive.

The validation set included 238 387 ambulance dispatches from October 2021 to December 2021, of whom 191 838 were transported to a regional hospital. A SARS-CoV-2 rtPCR test result was available in 103 336 patients, and 8253 (8%) were positive.

The population characteristics of training and validation sets are reported in table 1. The distribution of positive cases in the Lombardy region during the study periods is reported in online supplemental figure 4. Overall, the prevalence in the study period ranged from 73 cases/100 000 to 2528 cases/100 000 population.

Operator-based interview

The operator-based interview is based on binary answers to questions asked by receiver technicians at the PSAP-2, investigating signs and symptoms related to SARS-CoV-2 infection. The caller was considered a suspected case of COVID-19 by the PSAP-2 operator if they reported one or more of the signs and symptoms detailed in the 'Methods' section. The sensitivity and specificity of the interview in the whole training set were 85.5% and 58.7%, respectively. The PPV and NPV were 37.5% and 93.3% and accuracy 0.65 (table 2).

Importance of clinical variables retrieved by operators and EMS

To assess the importance of clinical variables, univariate and multivariate logistic regression models were implemented including predictor variables retrieved by the operator-based interview alone or variables provided by the on-scene ambulance report (ie, clinical parameters) combined with variables retrieved by the operator-based interview. Complete results are reported in table 3. When variables retrieved at the operator-based interview and clinical parameters obtained in the field by EMS were both included in the analysis, close contact, fever, cough and $SpO_2<94\%$ showed the highest association with SARS-CoV-2 infection. The C-index of the model based on the operator-based interview alone was 0.79. The logistic regression model that included all variables (ie, operator-based interview *plus* clinical parameters) had a C-index of 0.83.

Machine learning models

The best performing algorithm for all models was the random forest (table 4) in both training and validation sets. Complete metrics of the different ML algorithms that were tested in the training set are reported in the online supplemental table 2.

The performance of ML models was lower in the validation set, especially model 1. Model 4 had the highest AUC in training (0.94, 95% CI 0.93 to 0.95) and validation (0.85, 95% CI

Table 1 Population characteristics

	Training set (n=264 976)	Missing, n (%)	Validation set (n=103 336)	Missing, n (%)
Age, years	74 (55–84)	0 (0)	73 (50–84)	0 (0)
Male, n (%)	133 209 (50.3)	4552 (1.7)	49 955 (48.3)	1662 (1.6)
EMS call				
Accidents, n (%)	46 336 (17.5)		23 587 (22.8)	
Heart disease, n (%)	50 489 (19.0)		20 762 (20.1)	
Respiratory disease, n (%)	54 630 (20.6)		15 611 (15.1)	
Neurological disease, n (%)	23 816 (9.0)		10 033 (9.7)	
Other medical disease, n (%)	83 387 (31.5)		30 565 (29.6)	
Other/Unknown, n (%)		6318 (2.4)		2778 (2.7)
Operator-based interview				
Close contact, n (%)	28 343 (10.7)		3528 (3.4)	
Shortness of breath, n (%)	87 944 (33.2)		1037 (1)	
Fever, n (%)	51 549 (19.4)		11 393 (11)	
Vomit, n (%)	16 704 (6.3)		7793 (7.5)	
Diarrhoea, n (%)	7896 (3.0)		1976 (1.9)	
Asthenia/Diffuse pain, n (%)	31 573 (11.9)		8034 (7.8)	
Ageusia/Anosmia, n (%)	2424 (0.9)		239 (0.2)	
Cough, n (%)	27 515 (10.4)		6570 (6.4)	
Clinical parameters				
AVPU score				
Alert, n (%)	233 228 (88.1)		91 512 (93.7)	
Verbal, n (%)	8736 (3.2)		3059 (3.1)	
Pain, n (%)	3418 (1.3)		1180 (1.2)	
Unresponsive, n (%)	2730 (1.0)		1029 (1.1)	
Unknown, n (%)		16 864 (6.4)		897 (0.9)
RR, bpm	20 (17–22)	114 996 (43.4)	18 (16–20)	45 032 (43.6)
Room air SpO_2	97 (94–98)	35 273 (13.3)	97 (96–99)	13 144 (12.7)
Respiratory quality				
Normal, n (%)	191 327 (72.2)		79 419 (81.3)	
Altered, n (%)	51 173 (19.3)		15 225 (15.6)	
Absent, n (%)	961 (0.4)		406 (0.4)	
Non-specified		21 515 (8.1)		2627 (2.7)
HR, bpm	86 (75–100)	54 594 (9.9%)	86 (74–100)	10 089 (9.8%)
SBP, mm Hg	135 (120–150)	32 225 (12.2)	138 (120–155)	12 430 (12)
DBP, mm Hg	80 (70–90)	32 523 (12.3)	80 (70–90)	12 554 (12.1)
Body temperature, °C	36.5 (36.1–36.8)	31 700 (12)	36.4 (36–36.7)	13 667 (13.2)
NEWS				
Low, n (%)	182 314 (68.8)		76 466 (74)	
Low-medium, n (%)	32 891 (12.4)		9013 (8.7)	
Medium, n (%)	28 967 (10.9)		11 255 (10.9)	
High, n (%)	20 806 (7.9)		6602 (6.4)	

Data reported as numbers and percentages, n (%) or median (IQR).
.DBP, diastolic BP; EMS, emergency medical service; NEWS, National Early Warning Score (low: 0–4; low-medium: 3 in any parameter; medium: 5–6; high: 7 or more)[29]. SBP, systolic BP; SpO_2, pulse oximeter oxygen saturation.

0.84 to 0.86) sets, respectively (figure 2 left, and table 4). The importance of each explanatory variable in the model output is reported as SHAP value and graphically represented in figure 2 (right). Briefly, close contact and fever were the most relevant variables in determining the outcome in all four models. Other important variables were cough and age in model 1, and SpO_2 and cough in model 2. Caller geographical distribution was the third most important variable in both models 3 and 4.

A further ML model was developed to test the ability to refer patients to the hub hospitals based on arbitrary criteria.

In the first simulation (panel A, figure 3), in the low prevalence period (n=96 984), positive patients would be addressed to hub hospitals, and negative patients to non-hub hospitals. Based on actual data, the goal was achieved in 61.6% (n=59 724) patients, while it would be achieved in 81.8% (n=79 386) in an ML-based scenario.

In the second simulation (panel B, figure 3), in patients presenting with severe features in the high prevalence period (n=37 230), positive patients would be addressed to hub hospitals, while negative patients would be addressed to non-hub

Table 2 Operator-based interview performances

Time interval (prevalence range)	rtPCR+	rtPCR–		
Whole training set (73–1579 per 100 000 people)				
Suspected cases, n	50 980	84 897	**37.5%**	**PPV**
Not suspected cases, n	8618	120 553	**93.3%**	**NPV**
	85.5%*	**58.7%†**		
Q1 (73–290 per 100 000 people)				
Suspected cases, n	1302	17 461	**6.9%**	**PPV**
Not suspected cases, n	463	37 518	**98.8%**	**NPV**
	73.8%*	**68.2%†**		
Q2 (306–542 per 100 000 people)				
Suspected cases, n	9227	24 040	**27.7%**	**PPV**
Not suspected cases, n	1665	31 230	**94.9%**	**NPV**
	84.7%*	**56.5%†**		
Q3 (543–815 per 100 000 people)				
Suspected cases, n	14 432	23 355	**38.2%**	**PPV**
Not suspected cases, n	2528	29 249	**92.0%**	**NPV**
	85.1%*	**55.6%†**		
Q4 (823–1579 per 100 000 people)				
Suspected cases, n	28 667	21 488	**57.2%**	**PPV**
Not suspected cases, n	4175	25 438	**85.9%**	**NPV**
	87.3%*	**54.2%†**		

*Sensitivity.
†Specificity.
Q, quartile; rtPCR, reverse transcriptase PCR; PPV, positive predictive value; NPV, negative predictive value.

hospitals. Based on actual data, the goal was achieved in 50.6% (n=18 850) of cases and would be achieved in 74.4% (n=27 688) in an ML-based scenario. Complete data of both simulations are provided in online supplemental figures 5-8 and online supplemental table 3.

DISCUSSION

This cohort study, conducted in one of the most involved areas in Europe during the pandemic, investigated the association of prehospital demographic data and clinical features among patients with a rtRCR-confirmed infection who called EMS.[12 13]

Similar to other studies, close contact with a known case, cough and fever were most predictive of COVID-19.[14 15] The presence of altered consciousness, vomiting, diarrhoea and haemodynamic instability was associated with a reduced risk of infection, suggesting that aetiologies other than COVID-19 were responsible for the symptoms for which the patient was seeking care.[16 17] A clinical algorithm using the variables obtained by an EMS operator had a sensitivity of >80%, but low specificity, which is reasonable for a screening test in the prehospital setting.[18] However, a ML model using additional clinical and epidemiological data, which were available to EMS in the prehospital setting, showed superior performance in detecting cases with greater sensitivity and specificity.

The implementation of ML models to guide clinical decisions has gained interest recently, especially in hospital settings.[6] During the pandemic, studies focused on early COVID-19 detection and prediction of disease progression.[19–24] Canas *et al* estimated the probability of an individual being infected with SARS-CoV-2 based on self-reported symptoms. They found that a hierarchical Gaussian process model trained on 3 days of symptoms had an AUC of 0.80 (95% CI 0.80 to 0.81), which is comparable to our models.[24] Soltan *et al* developed a tool (CURIAL-Lab) to screen for SARS-CoV-2 infection in the ED, with an AUC

Table 3 Logistic regression analysis

	Operator-based interview alone			Operator-based interview plus clinical parameters		
	OR	95% CI	P value	OR	95% CI	P value
Operator-based interview						
Age >74 years	0.19	0.19 to 0.2	<0.001	0.17	0.17 to 0.18	<0.001
Male gender	0.24	0.24 to 0.25	<0.001	0.3	0.3 to 0.31	<0.001
Close contact	67.54	63.78 to 71.52	<0.001	55.26	51.84 to 58.91	<0.001
Shortness of breath	0.71	0.69 to .073	<0.001	0.78	0.75 to 0.81	<0.001
Fever	3.26	3.16 to 3.35	<0.001	3.69	3.57 to 3.82	<0.001
Diarrhoea	0.67	0.62 to 0.71	<0.001	0.71	0.67 to 0.77	<0.001
Cough	3.35	3.22 to 3.49	<0.001	3.2	3.06 to 3.34	<0.001
Ageusia/Anosmia	3.14	2.74 to 3.59	<0.001	3.02	2.62 to 3.48	<0.001
Asthenia/Diffuse pain	1.09	1.05 to 1.13	<0.001	1.09	1.05 to 1.14	<0.001
Vomit	0.28	0.27 to 0.3	<0.001	0.33	0.31 to 0.35	<0.001
Clinical parameters						
EMS call for respiratory disease	/	/	/	1.07	1.03 to 1.12	0.002
Non-alert	/	/	/	0.39	0.36 to 0.42	<0.001
Altered respiratory quality	/	/	/	0.99	0.95 to 1.03	0.595
Body temperature >38°C	/	/	/	1.22	1.17 to 1.28	<0.001
Room air SpO_2 <94%	/	/	/	2.25	2.17 to 2.33	<0.001
HR >100 bpm	/	/	/	0.3	0.29 to 0.31	<0.001
SBP <90 mm Hg	/	/	/	0.64	0.57 to 0.72	<0.001
DBP <60 mm Hg	/	/	/	0.52	0.48 to 0.55	<0.001

DBP, diastolic BP; EMS, emergency medical service; SBP, systolic BP; SpO_2, pulse oximeter oxygen saturation.

Table 4 Detailed metrics of different machine learning (random forest) models in the training and validation sets

Machine learning model	Metrics	Training set	Validation set
Model 1 ► Age ► Gender ► Variables retrieved by the operator-based interview	AUC	0.85 (0.84 to 0.87)	0.76 (0.75 to 0.77)
	SENS	0.96 (0.96 to 0.97)	0.92 (0.91 to 0.92)
	SPEC	0.18 (0.17 to 0.19)	0.17 (0.17 to 0.18)
	ACC	0.58 (0.57 to 0.59)	0.75 (0.74 to 0.75)
	PPV	0.35 (0.34 to 0.36)	0.09 (0.09 to 0.09)
	NPV	0.91 (0.90 to 0.93)	0.96 (0.96 to 0.96)
Model 2 ► Model 1 variables ► Clinical parameters retrieved on-scene ambulance	AUC	0.92 (0.91 to 0.94)	0.80 (0.79 to 0.81)
	SENS	0.90 (0.89 to 0.91)	0.92 (0.92 to 0.93)
	SPEC	0.73 (0.72 to 0.74)	0.23 (0.23 to 0.23)
	ACC	0.94 (0.93 to 0.95)	0.77 (0.77 to 0.78)
	PPV	0.71 (0.70 to 0.72)	0.1 (0.1 to 0.1)
	NPV	0.62 (0.60 to 0.63)	0.97 (0.97 to 0.97)
Model 3 ► Model 1 variables ► Current local SARS-CoV-2 epidemiology, and geographical distribution of EMS calls for respiratory and infectious diseases in the previous 7 days	AUC	0.92 (0.91 to 0.93)	0.82 (0.81 to 0.82)
	SENS	0.90 (0.89 to 0.91)	0.91 (0.90 to 0.91)
	SPEC	0.70 (0.69 to 0.71)	0.41 (0.41 to 0.41)
	ACC	0.71 (0.70 to 0.72)	0.85 (0.85 to 0.85)
	PPV	0.57 (0.56 to 0.58)	0.12 (0.12 to 0.12)
	NPV	0.94 (0.94 to 0.95)	0.98 (0.98 to 0.98)
Model 4 All variables included in models 1–3	AUC	0.94 (0.93 to 0.95)	0.85 (0.84 to 0.86)
	SENS	0.90 (0.89 to 0.91)	0.91 (0.91 to 0.92)
	SPEC	0.81 (0.80 to 0.82)	0.44 (0.44 to 0.45)
	ACC	0.72 (0.71 to 0.73)	0.85 (0.85 to 0.85)
	PPV	0.69 (0.68 to 0.71)	0.13 (0.13 to 0.14)
	NPV	0.95 (0.94 to 0.95)	0.98 (0.98 to 0.98)

Detailed metrics (with 95% CIs) of the random forest algorithm trained to predict the positivity to the rtPCR test. Metrics are the AUC of the ROC curve, the SENS, SPEC, ACC, PPV and NPV of a custom working point (chosen as the first point with SENS ≥90%).
ACC, accuracy; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; ROC, receiving operating characteristics; SENS, sensitivity; SPEC, specificity.

range of 0.84–0.85 (95% CI 0.81 to 0.89) in validation cohorts. However, their model is based on full blood count values, along with vital signs, and is not applicable prehospital.[22]

The use of ML algorithms in the EMS context has been limited to specific subjects, such as the recognition of cardiac arrest and the need for critical care resources.[25–27] We developed an ML model that showed promise in helping EMS to detect COVID-19 cases. The integration of contact history, signs and symptoms, clinical parameters collected by ambulance personnel, along with geographical call distribution and current number of positive cases in a specific area, led to a model that could more accurately predict COVID-19 positivity by considering clinical data and up-to-date viral distribution in a specific territory. We included different explanatory variables in our models integrating the different information in a gradual manner. The first two models (ie, model 1 and model 2) include variables commonly retrieved by worldwide PSAP and might be applicable to other settings. The other models also include variables retrieved from local epidemiology and analysis of the geospatial distribution of EMS calls, hence leveraging information sharing between EMS and local public health authorities. The study also highlights that the weight of each variable changes throughout the analyses performed. In fact, when focusing on only interviews and clinical variables, close contact, fever and cough showed the strongest association with patients' positivity. When the same variables were included in the ML models, close contact and fever still showed the strongest association. However, call geographical distribution and local epidemiology played a significant role as well as improving the model's ability to detect positive cases.

Although the impact of the pandemic is declining, other similar calamities might occur in the future. It is therefore conceivable that ML-based models might be adapted and applied in the EMS setting to other events. It may be crucial for public health authorities to estimate the extent and spread of a pandemic disease, especially in the early phases when the course is unpredictable. EMS has a role in managing calls and patients one step before hospital care. In that sense, if ML algorithms were integrated into the out-of-hospital data process, EMS might provide public health authorities with early clues of disease spread.[4] On the other hand, with the differences between health systems, it would be essential to have algorithms flexible enough to adapt to prespecified criteria, for instance, to allocate patients to different hospitals in a network. For this reason, we simulated the application of an ML model to test its utility in referring patients to hospital resources with different characteristics (ie, hub vs non-hub hospitals). We found that the algorithm could 'correct' the hospital destination for a significant proportion of patients. For instance, in a high prevalence scenario, it may be desirable to limit access to hub hospitals for positive patients with severe features, with >20% of patients correctly re-addressed by the ML algorithm. Therefore, although our models do not predict individual clinical severity and outcome, they might be potentially useful at a prehospital level for operational or public health reasons.

This study included a large number of patients managed by a regional EMS that links out-of-hospital clinical presentation with the result of the gold standard rtPCR test performed in a close time frame and retrieved from an official database directly provided by the regional public health authorities. Moreover, most variables included in the analysis are relatively simple, precise and commonly retrieved by other EMS. Thus, the information provided by our study could be relevant and applied to other services worldwide. The signs, symptoms and clinical parameters were screened and retrieved precisely and contemporaneously by trained personnel and using the same software. The dimension of the dataset allowed for consistent analysis, enabling the application of a 10-fold validation protocol. Finally, the ML models maintained a good performance (AUC >0.8) on validation on a large, independent dataset. This suggests a stable application of our models in the setting of different viral variants presenting with different clinical and epidemiological characteristics.

Our study has some limitations. First, we included in the analysis patients whose rtPCR test was done within 7 days of their EMS call. Therefore, patients whose tests were performed outside this time frame have been excluded. As most studies assume a median incubation period of up to 5–7 days, it is unlikely that this timeframe might significantly impact the performance of the models implemented in the study.[28] Second, the EMS in Lombardy is part of a two-level PSAP system, where the PSAP-2 dispatches ambulances in the regional territory and allocates them to different hospitals, which have different characteristics and resources. Thus, the applicability of our model might be challenged in areas with very different EMS and hospital systems. However, we tried to overcome such limitation by including in our models variables commonly retrieved by EMS worldwide. Third, our analysis does not consider the different viral variants that have been shown to impact viral

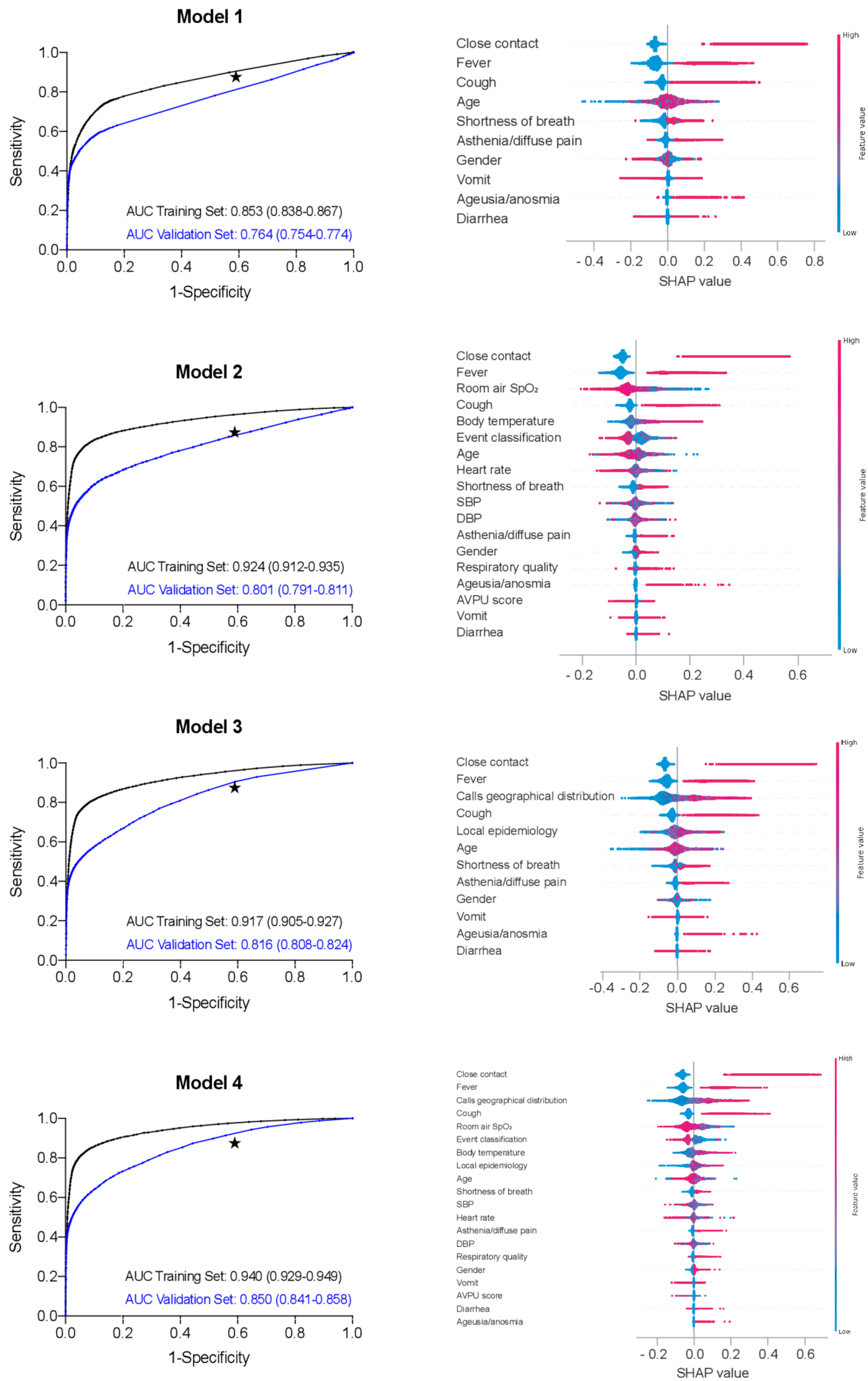


Figure 2 Machine learning models performance. *Left*: receiver operating characteristic (ROC) curves of the four models implemented, as compared with the performance of the operator-based interview (*). The box within each graph reports the AUC with the respective 95% CI, relevant to the random forest classifier. Black line, training set; blue line, validation set. *Right*: graphical representation of the contribution of each explanatory variable in predicting SARS-CoV-2 positivity within each model, according to SHAP. The impact on the model is reported as SHAP value. The lines represent the variables, whereas each dot represents a single record. The importance of the variable within the specific model decreases from the top to the bottom. The colour of a single dot is red if the value is high or blue if the value is low. AUC, area under the curve; DBP, diastolic blood pressure; SBP, systolic BP; SHAP, Shapley additive explanations; SpO_2, pulse oximeter oxygen saturation.

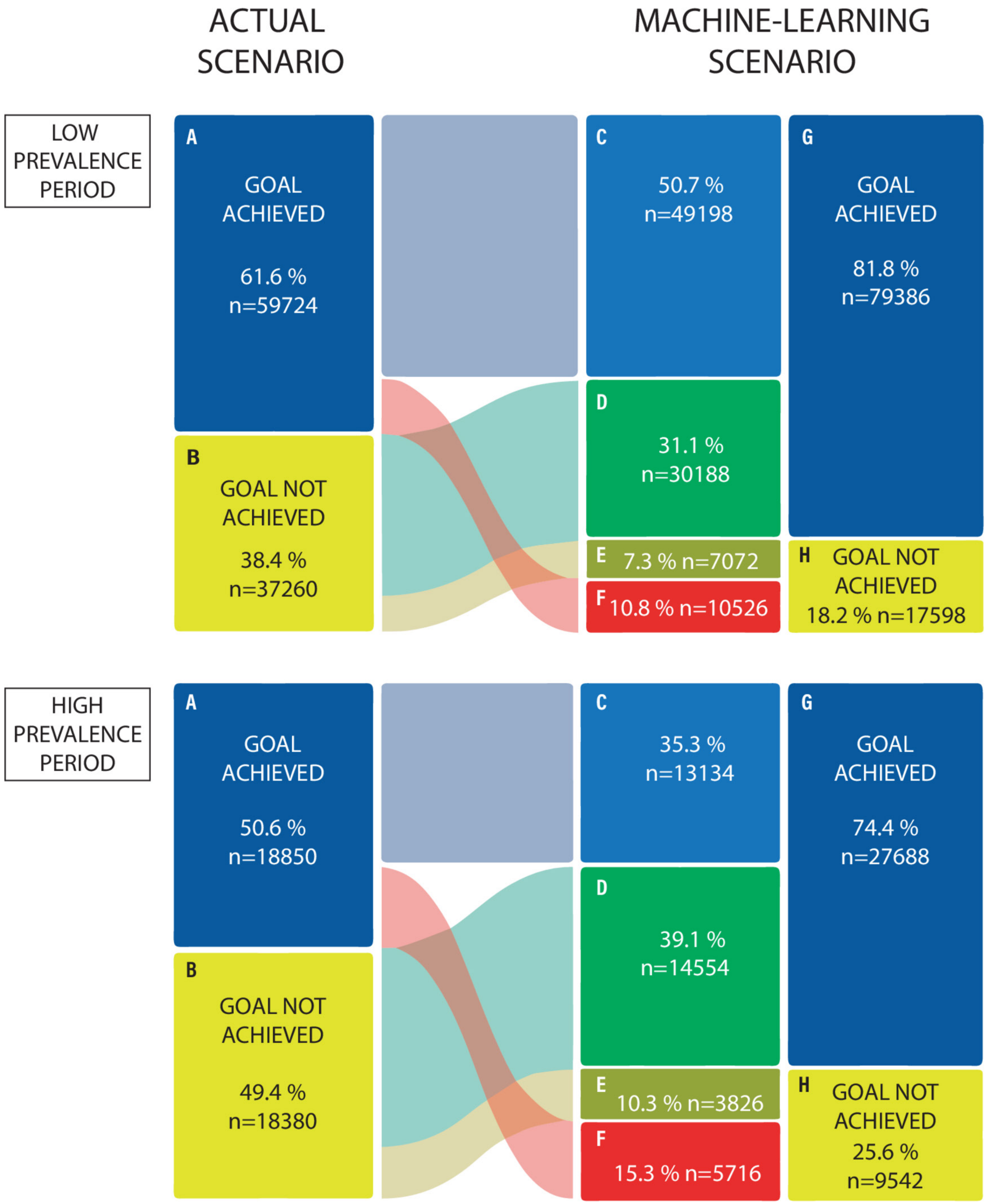


Figure 3 Real-world scenario simulation. *Upper panel*: a first simulation was performed in the low prevalence period (n=96 984). Here, positive patients would be addressed to hub hospitals, while negative patients to non-hub hospitals. Based on actual data (ie, actual scenario), the goal was achieved in 61.6% (n=59 724) patients, while it would be achieved in 81.8% (n=79 386) in a machine learning-based scenario. Box C, n=49 198 (50.7%). Box D, n=30 188 (31.1%). Box E, n=7072 (7.3%). Box F, n=10 526 (10.8%). *Lower panel*: a second simulation was performed in patients presenting with severe features at EMS calls in the high prevalence period (n=37 230). Here, positive patients presenting with severe features would be addressed to hub hospitals, while negative patients presenting with severe features would be addressed to non-hub hospitals. Here, the goal was achieved in 50.6% (n=18 850) in the actual scenario and it would be achieved in 74.4% (n=27 688) patients in a machine learning-based scenario. Box C, n=13 134 (35.3%). Box D, n=14 554 (39.1%). Box E, n=3826 (10.3%). Box F, n=5716 (15.3%).

shedding, contagiousness, transmissibility and clinical severity. Fourth, the analysis does not include the vaccination status of either single patients or the general population. However, as the training and the validation sets are temporally independent, it could be hypothesised that the patient profiles were different, especially with respect to different viral variants and vaccination status. Performance in the validation cohort was good, with an AUC >0.8 in most models. Fifth, we acknowledge that an rtPCR result was unavailable in about half of the subjects included in the study period. However, the risk of verification bias is low as all patients underwent an rtPCR test once admitted to the ED regardless of the reason for calling EMS. Moreover, RR was not included in model development due to the high proportion of missing data. Given that respiratory symptoms were a key feature of COVID-19, this may have impacted model performance. Finally, the estimated improvement in the achievement of hospital destination (hub vs non-hub) does not consider operational components of the real-world scenario, such as crowding level of different facilities and urgency of interventions, which could have affected decisions about actual hospital destination.

CONCLUSIONS

An operator-based interview that explores signs and symptoms most commonly associated with COVID-19 showed a sensitivity >80% for detecting patients with COVID-19. An ML model that integrates clinical variables, geographical information and current local epidemiology showed the best performance in detecting cases. When the ML model is tested in real-world scenarios, such as the determination of hospital destination, the model can guide EMS to refer a remarkable percentage of patients to the proper hospital resources, based on prespecified allocation criteria.

Twitter Stefano Spina @ste_spi

Acknowledgements The authors are deeply grateful to all the technicians, nurses and physicians working in the emergency medical system in Lombardy (Italy). The authors thank David Campeau for his critical revision of the manuscript and for his support in further improving the study.

Collaborators AREU 118 EMS Network Collaborators: Matteo Caresani, Rainiero Rizzini, Fabrizio Canevari, Alessandra Sforza, Dario Franchi, MD, Stefano Alberti, MD Antonella Brancaglione, MD, Paola Manzoni, MD, Fabio Sangalli, MD, Simone Redaelli, MD, Eleonora Brioschi.

Contributors SS, LG, FM, MM (Maurizio Migliari) and TL contributed to the study concept and design. AP, RB and MM (Maurizio Migliori) contributed to data acquisition. SS, LG and FM contributed to the initial drafting of the manuscript. All authors contributed to data interpretation and critical revision of the manuscript. RF, EGC and GMS contributed to study supervision. TL, MM (Maurizio Migliari), RF, EGC and GMS contributed to the discussion of the results and reviewed the data and the final manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. RF acts as guarantor for the study.

Funding This study was funded by 'Development of geospatial-temporal predictive models for air quality and pandemic risk'—Department of Electronics, Information and Bioengineering, Politecnico di Milano (Milan, Italy) (ID 80724). 'Department of Anaesthesia, Critical Care, and Pain Medicine'—School of Medicine and Surgery, University of Milano-Bicocca (Milan, Italy) (ID 2016cont0078).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the institutional review board of Milano Area 2 (approval ID: 598_2021). Due to the retrospective nature of the study, the requirement for written informed consent was waived.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data relative to SARS-CoV-2 positivity rate in the territory of Lombardy region are available in a public, open-access repository (data source: Protezione Civile repository, https://github.com/pcm-dpc/COVID-19). Data relative to the machine learning models implemented in the study will be available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Stefano Spina http://orcid.org/0000-0003-3210-1714
Roberto Fumagalli http://orcid.org/0000-0003-0398-1329

REFERENCES

1 Arsenault C, Gage A, Kim MK, *et al*. COVID-19 and resilience of Healthcare systems in ten countries. *Nat Med* 2022;28:1314–24.
2 Spina S, Marrazzo F, Migliari M, *et al*. The response of Milan's emergency medical system to the COVID-19 outbreak in Italy. *The Lancet* 2020;395:e49–50.
3 Marrazzo F, Spina S, Pepe PE, *et al*. Rapid reorganization of the Milan metropolitan public safety answering point operations during the initial phase of the COVID-19 outbreak in Italy. *J Am Coll Emerg Physicians Open* 2020;1:1240–9.
4 Gianquintieri L, Brovelli MA, Pagliosa A, *et al*. Mapping Spatiotemporal diffusion of COVID-19 in Lombardy (Italy) on the base of emergency medical services activities. *IJGI* 2020;9:639.
5 Gianquintieri L, Brovelli MA, Pagliosa A, *et al*. Geospatial correlation analysis between air pollution indicators and estimated speed of COVID-19 diffusion in the Lombardy region (Italy). *Int J Environ Res Public Health* 2021;18:12154.
6 Ramlakhan S, Saatchi R, Sabir L, *et al*. Understanding and interpreting artificial intelligence, machine learning and deep learning in emergency medicine. *Emerg Med J* 2022;39:380–5.
7 Ramlakhan SL, Saatchi R, Sabir L, *et al*. Building artificial intelligence and machine learning models: a primer for emergency physicians. *Emerg Med J* 2022;39:e1.
8 World Heal Organ. COVID-19 WCM of. Living Guidance for clinical management of COVID-19, 2021. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-1
9 Zhang Z, Bi Q, Fang S, *et al*. Insight into the practical performance of RT-PCR testing for SARS-Cov-2 using serological data: a cohort study. *Lancet Microbe* 2021;2:e79–87.
10 Zieliński W. The shortest Clopper–Pearson confidence interval for binomial probability. *Communications in Statistics - Simulation and Computation* 2009;39:188–93.
11 Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;2017:4766–75.
12 Grasselli G, Zangrillo A, Zanella A, *et al*. Baseline characteristics and outcomes of 1591 patients infected with SARS-Cov-2 admitted to Icus of the Lombardy region, Italy. *JAMA* 2020;323:1574.
13 Grasselli G, Tonetti T, Protti A, *et al*. Pathophysiology of COVID-19-associated acute respiratory distress syndrome: a Multicentre prospective observational study. *Lancet Respir Med* 2020;8:1201–8.
14 Huang C, Wang Y, Li X, *et al*. Clinical features of patients infected with 2019 novel Coronavirus in Wuhan, China. *Lancet* 2020;395:497–506.
15 Mao B, Liu Y, Chai Y-H, *et al*. Assessing risk factors for SARS-Cov-2 infection in patients presenting with symptoms in Shanghai, China: a Multicentre, observational cohort study. *The Lancet Digital Health* 2020;2:e323–30.
16 Cummings MJ, Baldwin MR, Abrams D, *et al*. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York city: a prospective cohort study. *Lancet* 2020;395:1763–70.
17 Moeller AL, Mills EHA, Collatz Christensen H, *et al*. Symptom presentation of SARS-Cov-2-positive and negative patients: A nested case-control study among patients calling the emergency medical service and medical Helpline. *BMJ Open* 2021;11:e044208.
18 Trevethan R. Specificity, and predictive values: foundations, Pliabilities, and pitfalls in research and practice. *Front Public Health* 2017;5:307.
19 Wang M, Xia C, Huang L, *et al*. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. *The Lancet Digital Health* 2020;2:e506–15.

20 Langer T, Favarato M, Giudici R, *et al*. Development of machine learning models to predict RT-PCR results for severe acute respiratory syndrome Coronavirus 2 (SARS-Cov-2) in patients with influenza-like symptoms using only basic clinical data. *Scand J Trauma Resusc Emerg Med* 2020;28:113.
21 Jiao Z, Choi JW, Halsey K, *et al*. Prognostication of patients with COVID-19 using artificial intelligence based on chest X-rays and clinical data: a retrospective study. *The Lancet Digital Health* 2021;3:e286–94.
22 Soltan AAS, Yang J, Pattanshetty R, *et al*. Real-world evaluation of rapid and laboratory-free COVID-19 triage for emergency care: external validation and pilot deployment of artificial intelligence driven screening. *The Lancet Digital Health* 2022;4:e266–78.
23 Piekos SN, Roper RT, Hwang YM, *et al*. The effect of maternal SARS-Cov-2 infection timing on birth outcomes: a retrospective Multicentre cohort study. *The Lancet Digital Health* 2022;4:e95–104.
24 Canas LS, Sudre CH, Capdevila Pujol J, *et al*. Early detection of COVID-19 in the UK using self-reported symptoms: a large-scale, prospective, Epidemiological surveillance study. *The Lancet Digital Health* 2021;3:e587–98.
25 Kang D-Y, Cho K-J, Kwon O, *et al*. Artificial intelligence algorithm to predict the need for critical care in Prehospital emergency medical services. *Scand J Trauma Resusc Emerg Med* 2020;28:17.
26 Kwon J, Jeon K-H, Kim HM, *et al*. Deep-learning-based out-of-hospital cardiac arrest Prognostic system to predict clinical outcomes. *Resuscitation* 2019;139:84–91.
27 Blomberg SN, Christensen HC, Lippert F, *et al*. Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: A randomized clinical trial. *JAMA Netw Open* 2021;4:e2032320.
28 He X, Lau EHY, Wu P, *et al*. Temporal Dynamics in viral shedding and Transmissibility of COVID-19. *Nat Med* 2020;26:1491–3.
29 Bilben B, Grandal L, Søvik S. National early warning score (NEWS) as an emergency Department Predictor of disease severity and 90-day survival in the acutely Dyspneic patient - a prospective observational study. *Scand J Trauma Resusc Emerg Med* 2016;24:80.

Figure 1, graphical representation of the development and optimization process of a machine-learning algorithm to predict the positivity of patients rescued by EMS to SARS-CoV-2 in the pre-hospital emergency setting.

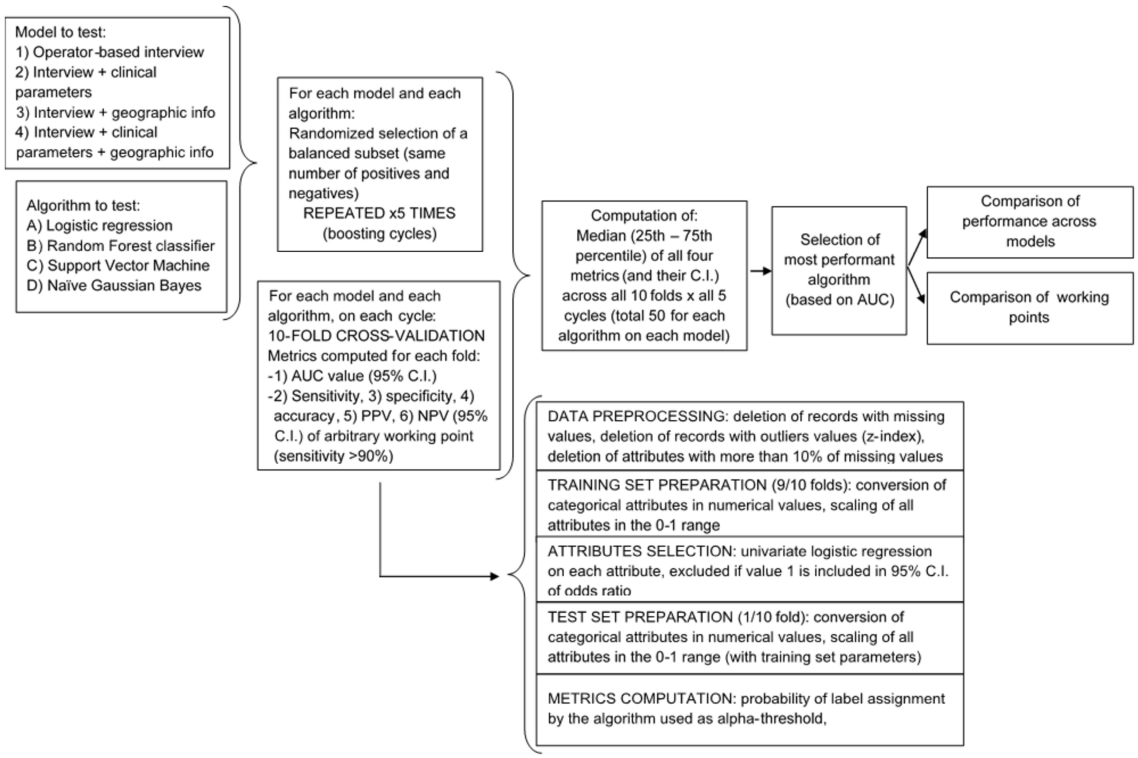


Figure 2, graphical representation of a machine-learning algorithm to simulate a real-time running decision making support for the destination of patients rescued by EMS towards COVID-19 hub or non-hub hospitals.

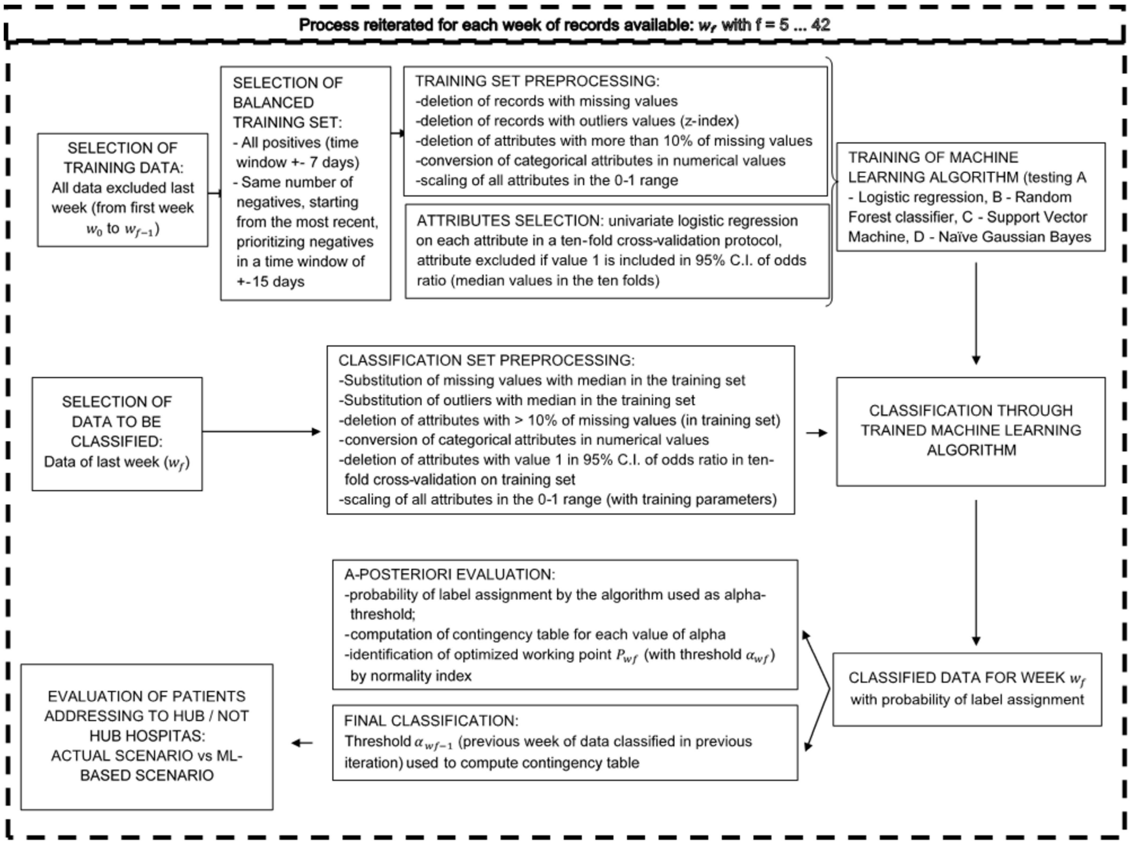
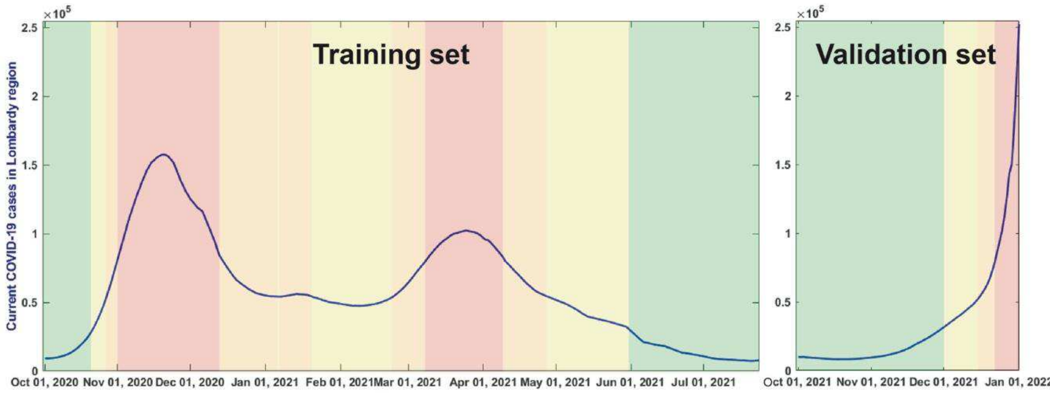
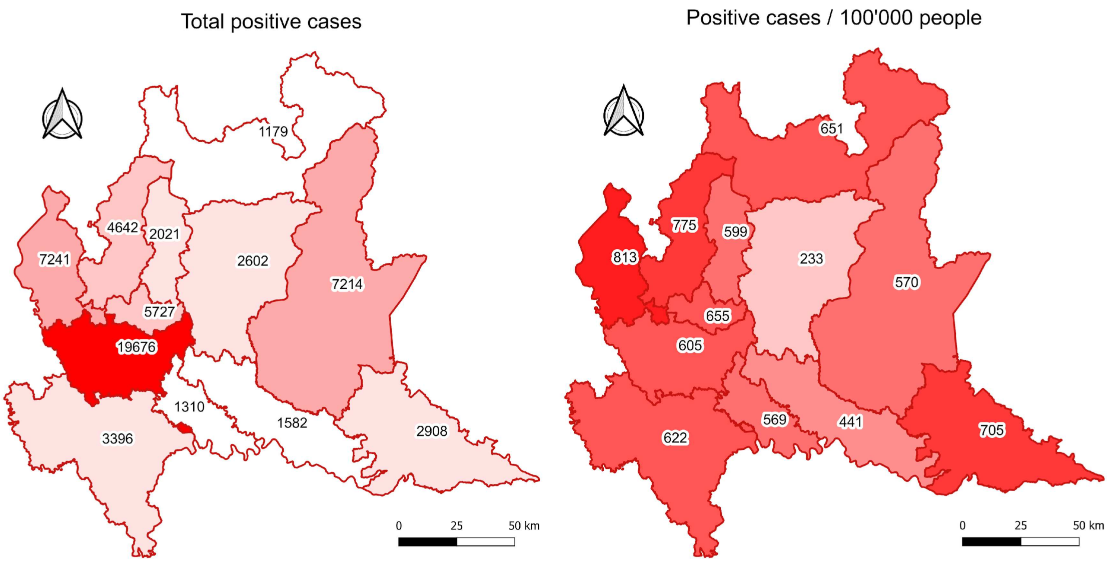


Figure 3, COVID-19 prevalence in the study period



Subdivision of days in the timeframe from Oct 1st, 2020, to Jul 23rd, 2021 (training set) and from Oct 1st, 2021, to Dec 31st, 2021, into four quartiles according to the number of patients currently positive for SARS-CoV-2 infection (filtered with a 7-day moving average) on the territory of Lombardy region, Italy. For each quartile of the training set, Table 1 reports the thresholds (minimum and maximum values), the range of values, and the timeframe assigned to it.

Figure 4, Positive cases distribution in the Lombardy region in the study period (October 1st 2020 – July 23rd 2021)



Lombardy map, divided into provinces. Color intensity is directly proportional to the number of cases in each province. The map on the left shows the absolute number of total positive cases in each province. The map on the right shows the number of positive cases per 100000 inhabitants in each province.

Table 1, Quartiles of COVID-19 prevalence in the study period

	Min value	Max value	Range	Dates	
1st Quartile	7325	28983	21658	Oct 1, 2020 – Oct 20, 2020 May 31, 2021 – Jul 23, 2021	Low prevalence period
2nd Quartile	30584	54144	23560	Oct 21, 2020 – Oct 26, 2020 Jan 6, 2021 Jan 20, 2021– Feb 21, 2021 Apr 27, 2021 – May 30, 2021	
3rd Quartile	54226	81436	27210	Oct 27, 2020 – Oct 31, 2020 Dec 13, 2020 – Jan 19, 2021 (excluding Jan 6, 2021) Feb 22, 2021 – Mar 7, 2021 Apr 9, 2021 – Apr 26, 2021	High prevalence period
4th Quartile	82171	157614	75443	Nov 1, 2020 – Dec 12, 2020 Mar 8, 2021 – Apr 8, 2021	

Table 2, Detailed metrics of different machine learning algorithms in the training set.

Model	Logistic Regression	Random Forest	Support Vector Machine	Naïve Gaussian Bayes
Model 1	AUC: .853 (.838 - .867) SENS: .944 (.936 - .952) SPEC: .26 (.251 - .27) ACC: .605 - (.595 - .615) PPV: .361 (.351 - .371) NPV: .91 (.899 - .921)	AUC: .853 (.838 - .867) SENS: .963 (.956 - .969) SPEC: .178 (.17 - .187) ACC: .577 (.566 - .587) PPV: .347 (.338 - .356) NPV: .913 (.9 - .926)	AUC: .803 (.79 - .815) SENS: .996 (.994 - .998) SPEC: .009 (.007 - .012) ACC: .504 (.493 - .516) PPV: .309 (.301 - .317) NPV: .844 (.751 - .909)	AUC: .803 (.79 - .815) SENS: .996 (.994 - .998) SPEC: .009 (.007 - .012) ACC: .504 (.493 - .516) PPV: .309 (.301 - .317) NPV: .844 (.751 - .909)
Model 2	AUC: .867 (.852 - .882) SENS: .913 (.903 - .922) SPEC: .461 (.449 - .473) ACC: .654 (.644 - .664) PPV: .448 (.436 - .46) NPV: .919 (.909 - .927)	AUC: .917 (.905 - .927) SENS: .902 (.892 - .912) SPEC: .696 (.686 - .706) ACC: .712 (.704 - .719) PPV: .571 (.558 - .584) NPV: .941 (.935 - .947)	AUC: .836 (.821 - .851) SENS: .979 (.974 - .984) SPEC: .048 (.043 - .053) ACC: .519 (.507 - .531) PPV: .329 (.32 - .338) NPV: .827 (.787 - .863)	AUC: .84 (.826 - .854) SENS: 1 (.999 - 1) SPEC: 0 (0 - 0) ACC: .5 (.488 - .512) PPV: .323 (.314 - .333) NPV: 0 (0 - 0)
Model 3	AUC: .883 (.869 - .895) SENS: .905 (.895 - .914) SPEC: .576 (.565 - .586) ACC: .688 (.679 - .696) PPV: .486 (.474 - .498) NPV: .932 (.924 - .939)	AUC: .917 (.905 - .927) SENS: .902 (.892 - .912) SPEC: .696 (.686 - .706) ACC: .712 (.704 - .719) PPV: .571 (.558 - .584) NPV: .941 (.935 - .947)	AUC: .835 (.821 - .848) SENS: .98 (.975 - .984) SPEC: .048 (.044 - .053) ACC: .521 (.51 - .532) PPV: .315 (.306 - .323) NPV: .841 (.804 - .873)	AUC: .84 (.827 - .852) SENS: 1 (.999 - 1) SPEC: 0 (0 - 0) ACC: .5 (.488 - .512) PPV: .307 (.299 - .316) NPV: 0 (0 - 0)
Model 4	AUC: .89 (.876 - .903) SENS: .905 (.894 - .915) SPEC: .607 (.595 - .618) ACC: .686 (.677 - .695) PPV: .528 (.514 - .541) NPV: .931 (.923 - .938)	AUC: .94 (.929 - .949) SENS: .902 (.891 - .912) SPEC: .809 (.799 - .818) ACC: .722 (.714 - .731) PPV: .691 (.677 - .705) NPV: .946 (.939 - .951)	AUC: .864 (.849 - .877) SENS: .967 (.96 - .973) SPEC: .134 (.126 - .142) ACC: .554 (.543 - .566) PPV: .346 (.336 - .356) NPV: .892 (.871 - .911)	AUC: .85 (.836 - .863) SENS: 1 (.999 - 1) SPEC: 0 (0 - 0) ACC: .5 (.488 - .512) PPV: .323 (.314 - .332) NPV: 0 (0 - 0)

The four models refer to different subset of attributes (see main text for details), while the different columns report the results of four different machine learning algorithms. AUC = Area Under Curve, where the curve is a ROC obtained by variating the threshold on the output probability of the algorithm. All the other metrics are obtained in a specific working point of this curve, arbitrarily set to be the first point with a sensitivity >0.9. SENS = sensitivity, SPEC = specificity, ACC = accuracy, PPV = positive predictive value, NPV = negative predictive value.

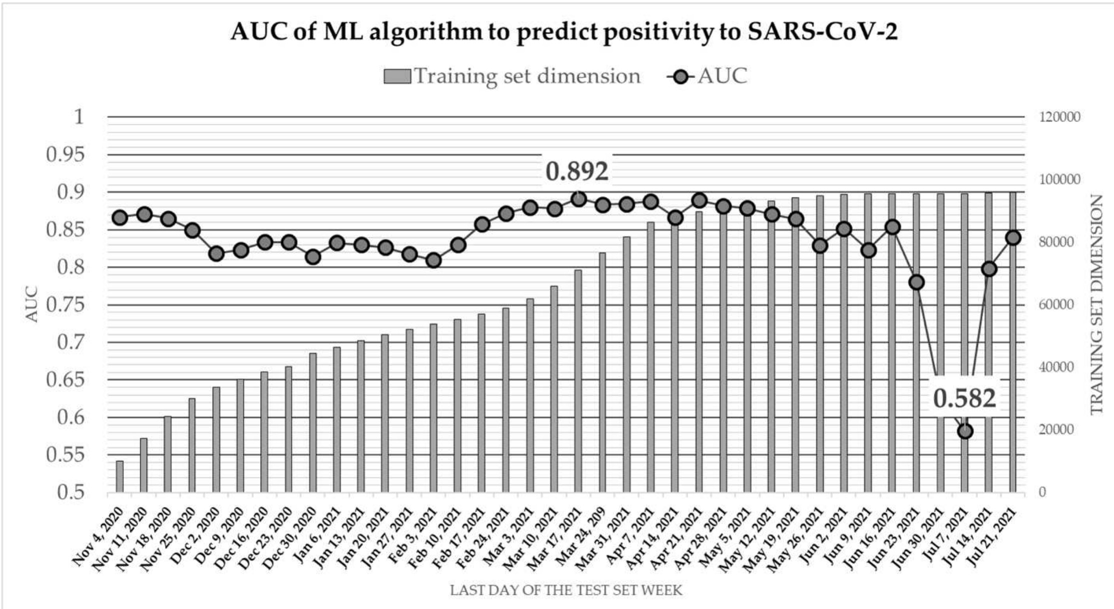
Table 3, Detailed results of a machine learning model (random forest) used to predict the positivity to SARS-CoV-2 infection.

Each week is progressively used as the test set, while previous records are used for the training. The different parameters computed are described in detail below Figure 2 ('perf.' stands for 'performance'). In the lower panel, TOTAL, MEDIAN, 25th Q(uartile), 75th Q(uartile) values are computed for the distribution of the different values across weeks.

	Training set	Test set (positives-negatives)	Attributes* removed (missing)	Attributes* removed (significance)	AUC	Cut-off	Baseline perf.	Model-based perf.	Perf. differential	Perf. differential - Positives	Perf. differential - Negatives	Error relative reduction (pos-neg)	Goal reached kept (%bv)	New goal achievement (%be)	Unidentified improvements (%be)	Induced mistakes (%bv)
Nov 04, 2020	10124	8198 (4287-3911)	12, 14	6, 10, 11	.867	.493	1303 (48.38%)	2063 (76.61%)	760 (28.22%)	1110(55.5%-129.5%)	-350 (-50.6%--78.5%)	-54.7% (-97.%|142.3%)	907 (69.61%)	1156 (83.17%)	234 (16.83%)	396 (30.39%)
Nov 11, 2020	17200	8188 (4459-3729)	12, 14	2, 6, 11, 16	.871	.501	1413 (46.77%)	2418 (80.04%)	1005 (33.27%)	1256 (55.1%-139.1%)	-251 (-33.8%--49.2%)	-62.5% (-91.3%|107.7%)	1037 (73.39%)	1381 (85.88%)	227 (14.12%)	376 (26.61%)
Nov 18, 2020	24356	7512 (3648-3864)	12, 14	6, 15, 16	.865	.502	1075 (43.42%)	1928 (77.87%)	853 (34.45%)	1087 (61.%-176.2%)	-234 (-33.7%--51.1%)	-60.9% (-93.3%|99.2%)	736 (68.47%)	1192 (85.08%)	209 (14.92%)	339 (31.53%)
Nov 25, 2020	30150	6525 (2368-4157)	12, 14	1, 6, 11, 16	.850	.505	881 (48.22%)	1291 (70.66%)	410 (22.44%)	630 (57.1%-150.4%)	-220 (-30.4%--47.6%)	-43.3% (-92.%|84.3%)	554 (62.88%)	737 (77.91%)	209 (22.09%)	327 (37.12%)
Dec 02, 2020	33752	6100 (1670-4430)	12, 14	6, 16	.819	.496	793 (52.97%)	927 (61.92%)	134 (8.95%)	364 (54.7%-140.%)	-230 (-27.7%--43.2%)	-19.% (-89.7%|77.2%)	437 (55.11%)	490 (69.6%)	214 (30.4%)	356 (44.89%)
Dec 09, 2020	36272	6144 (1469-4675)	12, 14	1, 2, 6, 11	.823	.489	831 (57.27%)	835 (57.55%)	4 (.28%)	292 (51.8%-128.6%)	-288 (-32.5%--47.7%)	-.6% (-86.6%|101.8%)	434 (52.23%)	401 (64.68%)	219 (35.32%)	397 (47.77%)
Dec 16, 2020	38490	5731 (1110-4621)	12, 14	2, 6, 16	.833	.506	689 (55.79%)	613 (49.64%)	-76 (-6.15%)	244 (56.9%-148.8%)	-320 (-39.7%--61.%)	13.9% (-92.1%|113.9%)	290 (42.09%)	323 (59.16%)	223 (40.84%)	399 (57.91%)
Dec 23, 2020	40232	5669 (1066-4603)	12, 14	1, 11, 16, 19	.834	.517	650 (55.41%)	648 (55.24%)	-2 (-.17%)	225 (56.8%-152.%)	-227 (-29.2%--45.2%)	.4% (-90.7%|82.5%)	327 (50.31%)	321 (61.38%)	202 (38.62%)	323 (49.69%)
Dec 30, 2020	44448	5961 (1094-4867)	12, 13, 14	1, 6, 17	.815	.492	794 (59.17%)	954 (71.09%)	160 (11.92%)	179 (41.3%-107.8%)	-19 (-2.1%--3.%)	-29.2% (-67.%|6.8%)	554 (69.77%)	400 (72.99%)	148 (27.01%)	240 (30.23%)
Jan 06, 2021	46314	6088 (1273-4815)	12, 13, 14	1, 6	.833	.518	809 (54.33%)	1160 (77.9%)	351 (23.57%)	210 (41.%-120.%)	141 (14.4%-22.2%)	-51.6% (-62.3%|-41.1%)	642 (79.36%)	518 (76.18%)	162 (23.82%)	167 (20.64%)
Jan 13, 2021	48496	6102 (1226-4876)	12, 13, 14	2, 6	.830	.481	1634 (57.88%)	2295 (81.3%)	661 (23.41%)	234 (31.6%-76.7%)	427 (20.5%-32.1%)	-55.6% (-53.8%|-56.6%)	1352 (82.74%)	943 (79.31%)	246 (20.69%)	282 (17.26%)
Jan 20, 2021	50558	5813 (949-4864)	12, 13, 14	1, 6	.827	.479	697 (57.32%)	983 (80.84%)	286 (23.52%)	107 (28.5%-67.3%)	179 (21.3%-33.3%)	-55.1% (-49.3%|-59.3%)	571 (81.92%)	412 (79.38%)	107 (20.62%)	126 (18.08%)
Jan 27, 2021	52152	5925 (966-4959)	12, 13, 14	2, 6	.818	.480	3072 (58.64%)	4328 (82.61%)	1256 (23.97%)	254 (28.8%-74.5%)	1002 (23.%-36.7%)	-58.% (-47.%|-61.6%)	2564 (83.46%)	1764 (81.4%)	403 (18.6%)	508 (16.54%)
Feb 03, 2021	53836	5869 (895-4974)	12, 13, 14	1, 2, 6	.810	.479	3503 (59.69%)	4760 (81.1%)	1257 (21.42%)	247 (27.6%-72.6%)	1010 (20.3%-31.9%)	-53.1% (-44.5%|-55.8%)	2861 (81.67%)	1899 (80.26%)	467 (19.74%)	642 (18.33%)
Feb 10, 2021	55358	6024 (993-5031)	12, 13, 14	1, 6	.830	.492	3603 (59.81%)	4927 (81.79%)	1324 (21.98%)	351 (35.3%-100.3%)	973 (19.3%-29.9%)	-54.7% (-54.6%|-54.7%)	2965 (82.29%)	1962 (81.04%)	459 (18.96%)	638 (17.71%)
Feb 17, 2021	57070	6013 (1162-4851)	12, 13, 14	1, 2, 6	.857	.487	3533 (58.76%)	5004 (83.22%)	1471 (24.46%)	463 (39.8%-114.3%)	1008 (20.8%-32.2%)	-59.3% (-61.2%|-58.5%)	2981 (84.38%)	2023 (81.57%)	457 (18.43%)	552 (15.62%)
Feb 24, 2021	59064	6466 (1581-4885)	12, 13, 14	1, 6	.872	.482	3277 (57.06%)	4748 (82.67%)	1471 (25.61%)	576 (39.9%-111.6%)	895 (20.8%-32.4%)	-59.7% (-62.%|-58.2%)	2717 (82.91%)	2031 (82.36%)	435 (17.64%)	560 (17.09%)
Mar 03, 2021	61862	7297 (2330-4967)	12, 13, 14	1, 2, 6	.880	.475	907 (51.16%)	1395 (78.68%)	488 (27.52%)	492 (50.3%-142.6%)	-4 (-.5%--.7%)	-56.4% (-77.6%|1.7%)	693 (76.41%)	702 (81.06%)	164 (18.94%)	214 (23.59%)
Mar 10, 2021	65908	7820 (2944-4876)	12, 13, 14	1, 6	.878	.489	986 (46.51%)	1717 (80.99%)	731 (34.48%)	713 (54.8%-157.7%)	18 (2.2%-3.4%)	-64.5% (-84.%|-6.3%)	768 (77.89%)	949 (83.69%)	185 (16.31%)	218 (22.11%)
Mar 17, 2021	71070	8109 (3153-4956)	12, 13, 14	1, 6	.892	.488	1101 (47.15%)	1846 (79.06%)	745 (31.91%)	831 (57.9%-168.2%)	-86 (-9.5%--14.2%)	-60.4% (-88.4%|29.3%)	823 (74.75%)	1023 (82.9%)	211 (17.1%)	278 (25.25%)
Mar 24, 209	76610	7818 (2964-4854)	12, 13, 14	1, 6, 17	.884	.495	1049 (47.42%)	1751 (79.16%)	702 (31.74%)	765 (55.3%-155.5%)	-63 (-7.6%--11.3%)	-60.4% (-85.9%|23.2%)	778 (74.17%)	973 (83.66%)	190 (16.34%)	271 (25.83%)
Mar 31, 2021	81810	7686 (2656-5030)	12, 13, 14	1, 6	.884	.499	994 (47.86%)	1658 (79.83%)	664 (31.97%)	602 (49.9%-143.7%)	62 (7.1%-10.8%)	-61.3% (-76.5%|-20.9%)	766 (77.06%)	892 (82.36%)	191 (17.64%)	228 (22.94%)
Apr 07, 2021	86500	7114 (2095-5019)	12, 13, 14	1, 6, 17	.888	.486	904 (50.73%)	1455 (81.65%)	551 (30.92%)	462 (50.8%-145.3%)	89 (10.2%-15.2%)	-62.8% (-78.%|-31.1%)	735 (81.31%)	720 (82.%)	158 (18.%)	169 (18.69%)
Apr 14, 2021	90248	6404 (1480-4924)	12, 13, 14	1, 6, 17	.867	.493	718 (50.42%)	1148 (80.62%)	430 (30.2%)	319 (52.%-161.1%)	111 (13.7%-21.3%)	-60.9% (-76.7%|-38.3%)	560 (77.99%)	588 (83.29%)	118 (16.71%)	158 (22.01%)

Apr 21, 2021	87636	6532 (1251-5281)	12, 14	1, 2	.889	.493	772 (55.22%)	915 (65.45%)	143 (10.23%)	293 (56.7%-156.7%)	-150 (-17.%--25.6%)	-22.8% (-88.8%|50.7%)	461 (59.72%)	454 (72.52%)	172 (27.48%)	311 (40.28%)
Apr 28, 2021	89722	6385 (1128-5256)	12, 14	1, 2	.882	.492	757 (55.83%)	882 (65.04%)	125 (9.22%)	279 (58.%-161.3%)	-154 (-17.6%--26.4%)	-20.9% (-90.6%|52.9%)	445 (58.78%)	437 (72.95%)	162 (27.05%)	312 (41.22%)
May 05, 2021	91580	6408 (968-5440)	12, 14	/	.879	.502	3844 (59.99%)	5332 (83.21%)	1488 (23.22%)	422 (43.6%-127.1%)	1066 (19.6%-30.4%)	-58.% (-66.4%|-55.3%)	3202 (83.3%)	2130 (83.07%)	434 (16.93%)	642 (16.7%)
May 12, 2021	93158	6067 (599-5468)	12, 14	/	.871	.497	3787 (62.42%)	5123 (84.44%)	1336 (22.02%)	234 (39.1%-108.8%)	1102 (20.2%-30.9%)	-58.6% (-60.9%|-58.1%)	3167 (83.63%)	1956 (85.79%)	324 (14.21%)	620 (16.37%)
May 19, 2021	94172	5943 (467-5476)	12, 14	1	.864	.487	3706 (62.36%)	5077 (85.43%)	1371 (23.07%)	178 (38.1%-106.6%)	1193 (21.8%-33.7%)	-61.3% (-59.3%|-61.6%)	3188 (86.02%)	1889 (84.44%)	348 (15.56%)	518 (13.98%)
May 26, 2021	94962	5790 (322-5468)	12, 14	1	.830	.503	3618 (62.49%)	4842 (83.63%)	1224 (21.14%)	106 (32.9%-94.6%)	1118 (20.4%-31.9%)	-56.4% (-50.5%|-57.%)	3017 (83.39%)	1825 (84.02%)	347 (15.98%)	601 (16.61%)
Jun 02, 2021	95508	5910 (209-5701)	12, 14	1, 2	.852	.506	3718 (62.91%)	5028 (85.08%)	1310 (22.17%)	63 (30.1%-77.8%)	1247 (21.9%-34.3%)	-59.8% (-49.2%|-60.4%)	3164 (85.1%)	1864 (85.04%)	328 (14.96%)	554 (14.9%)
Jun 09, 2021	95866	5901 (180-5721)	12, 14	1	.823	.513	3787 (64.18%)	4988 (84.53%)	1201 (20.35%)	40 (22.2%-58.8%)	1161 (20.3%-31.2%)	-56.8% (-35.7%|-58.%)	3199 (84.47%)	1789 (84.63%)	325 (15.37%)	588 (15.53%)
Jun 16, 2021	95224	5743 (129-5614)	12, 14	2	.854	.493	3664 (63.8%)	4863 (84.68%)	1199 (20.88%)	31 (24.%-60.8%)	1168 (20.8%-32.3%)	-57.7% (-39.7%|-58.4%)	3062 (83.57%)	1801 (86.63%)	278 (13.37%)	602 (16.43%)
Jun 23, 2021	95448	5658 (83-5574)	12, 14	2	.780	.501	3598 (63.6%)	4753 (84.02%)	1155 (20.42%)	16 (19.3%-50.%)	1139 (20.4%-31.9%)	-56.1% (-31.4%|-56.7%)	2973 (82.63%)	1780 (86.45%)	279 (13.55%)	625 (17.37%)
Jun 30, 2021	95588	5438 (75-5363)	12, 14	/	.625	.491	3401 (62.54%)	5083 (93.47%)	1682 (30.93%)	-13 (-17.3%--35.1%)	1695 (31.6%-50.4%)	-82.6% (34.2%|-84.8%)	3182 (93.56%)	1901 (93.32%)	136 (6.68%)	219 (6.44%)
Jul 07, 2021	95476	5393 (64-5329)	12, 14	/	.582	.495	3439 (63.77%)	4990 (92.53%)	1551 (28.76%)	-3 (-4.7%--13.%)	1554 (29.2%-45.5%)	-79.4% (7.3%|-81.2%)	3170 (92.18%)	1820 (93.14%)	134 (6.86%)	269 (7.82%)
Jul 14, 2021	95584	5215 (81-5134)	12, 14	/	.799	.509	3284 (62.97%)	4880 (93.58%)	1596 (30.6%)	-7 (-8.6%--14.9%)	1603 (31.2%-49.5%)	-82.7% (20.6%|-84.5%)	3070 (93.48%)	1810 (93.73%)	121 (6.27%)	214 (6.52%)
Jul 21, 2021	95728	3141 (84-3057)	12, 14	/	.840	.507	1983 (63.13%)	2927 (93.19%)	944 (30.05%)	-2 (-2.4%--4.3%)	946 (30.9%-48.8%)	-81.5% (5.3%|-84.5%)	1836 (92.59%)	1091 (94.21%)	67 (5.79%)	147 (7.41%)
TOTAL	/	240’100 (53’478-186’620)	/	/	/	/	136336 (58.01 %)	199355 (84.82 %)	63019 (26.81 %)	20087 (40.8%-115.6%)	42932 (23.1%-36.1%)	-63.8% (-63.%|-64.3%)	116230 (85.25%)	83125 (84.22%)	15580 (15.78%)	20106 (14.75%)
MEDIAN	46859	5827 (500-4824)	/	/	.824	.488	3611 (58.73 %)	5179 (84.26 %)	1522 (25.62 %)	368 (34.7%-93.6%)	1100 (22.4%-34.5%)	-63.3% (-55.7%|-63.6%)	3122 (84.7%)	2033 (83.62%)	437 (16.38%)	559 (15.3%)
25TH Q	93918	6530 (1989-5317)	/	/	.872	.502	3472 (55.89 %)	5026 (82.07 %)	1467 (24.74 %)	121 (27.9%-71.7%)	955 (19.5%-30.6%)	-67.3% (-66.%|-70.2%)	2978 (82.61%)	1913 (81.75%)	278 (12.56%)	439 (12.04%)
75TH Q	10124	8198 (4287-3911)	/	/	.867	.493	3787 (62.47 %)	5393 (87.74 %)	1820 (29.27 %)	641 (43.2%-125.4%)	1313 (25.3%-39.3%)	-59.1% (-45.8%|-53.5%)	3315 (87.96%)	2303 (87.44%)	542 (18.25%)	687 (17.39%)

Figure 5



Results relevant to a machine learning algorithm (random forest) used to predict the positivity for SARS-CoV-2 infection in patients rescued by an ambulance on the territory of Lombardy region, Italy, between October 1st 2020 and July 21st 2021. The performance is measured through the Area Under Curve (AUC), plotted with dots, of a Receiver Operating Characteristic (ROC) curve computed on each available week of data (dates reported on X axis), using preceding records as the training set (dimension represented by columns). Details about the implementation of this iterative system are reported in Figure 2.

Supplementary Online Content

Detection of patients with COVID-19 by the emergency medical services in Lombardy through an operator-based interview and machine learning models

Stefano Spina[1,2]*, MD, Lorenzo Gianquintieri[3]*, PhD, Francesco Marrazzo[1,2]*, MD, Maurizio Migliari[1], MD, Giuseppe Sechi[1], MD, Maurizio Migliori[1], MD, Andrea Pagliosa[1], Rodolfo Bonora[1], Thomas Langer[2], MD, Enrico Gianluca Caiani[3], PhD, Prof., Roberto Fumagalli[1,2], MD, Prof.; for the AREU 118 EMS Network[#]

Table of contents

Spina S, *et al. Emerg Med J* 2023;0:1–11. doi: 10.1136/emermed-2022-212853

Machine learning models development

POSITIVE CASES DETECTION

The aim of the first model implemented is to measure the capability to predict the positivity for SARS-CoV-2 at the rtPCR test with four different sets of attributes, representing different levels of information potentially available in the pre-hospital emergency context. To this scope, a supervised learning approach was implemented, testing four different machine learning algorithms: A) Logistic regression, B) Random Forest classifier, C) Support Vector Machine, D) Naïve Gaussian Bayes.

The different algorithms across different attributes sets were tested with a ten-fold validation protocol with 5 boosting cycles:

I) The initial dataset is extracted and pre-processed, by deleting records with missing values, deleting records with outliers' values (detected by z-index method with a threshold set to 5), deleting attributes with more than 15% of missing values.

II_i) The obtained pre-processed dataset is randomly sampled in order to obtain a balanced subset, hence composed by the same number of 'positive' and 'negative' cases.

III_i) The extracted subset is divided into ten randomly selected groups, each one composed by the same number of 'positive' and 'negative' cases.

IV_{ij}) One group is selected as the 'validation' set, the other 9 groups are selected as the 'training' set.

V_{ij}) The training set is pre-processed, by converting categorical attributes in numerical values, and scaling all attributes in the 0-1 range.

VI_{ij}) Relevant attributes are selected by computing a univariate logistic regression on each of them and excluding those attributes for which the value 1 is included in the 95% confidence interval of the odds ratio.

VII_{ij}) The machine learning algorithm is trained with the data in the training set.

$VIII_{ij}$) The validation set is pre-processed by converting categorical attributes in numerical values and scaling all attributes in the 0-1 range with the same scaling parameters used for the training set.

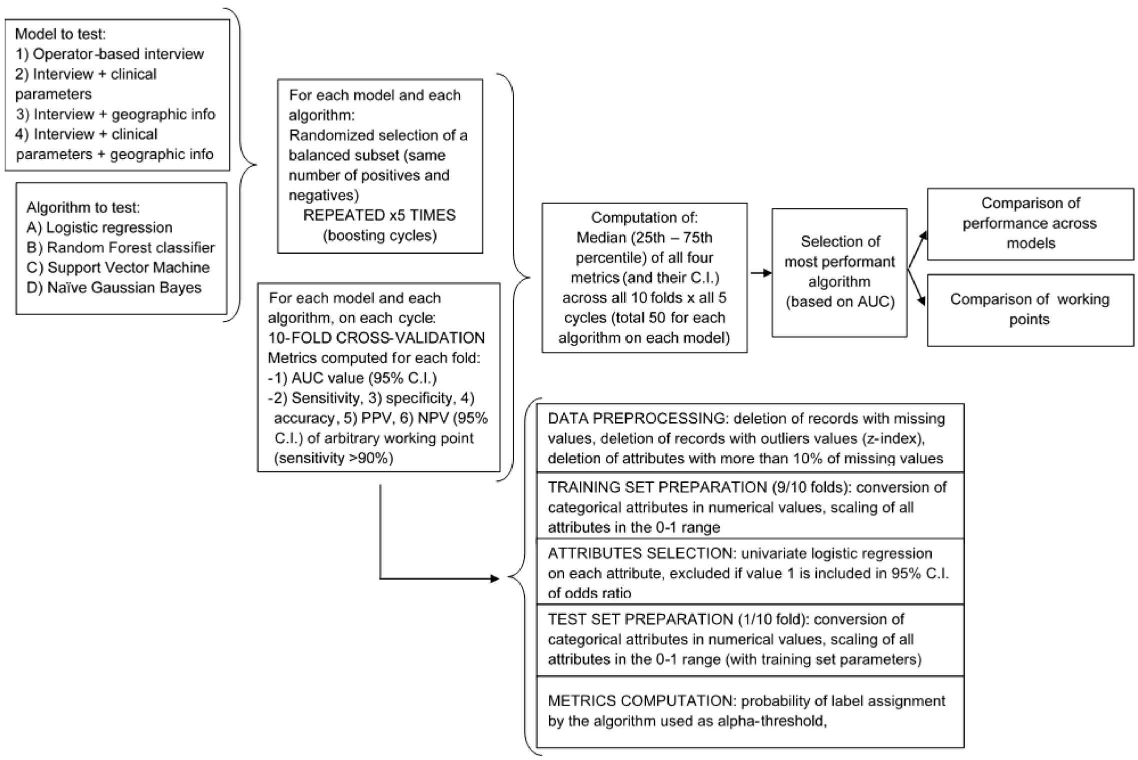
IX_{ij}) The validation set is classified with the trained machine learning algorithm.

X_{ij}) The probability of classification, provided as output by the machine learning algorithm, is used as an 'alpha' threshold to variate, thus allowing to compute a contingency table for each threshold value (with discrete steps of 1% probability), deriving as metrics the sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and accuracy (ACC), also inferring the 95% confidence intervals of each value with the Clopper-Pearson method; finally, the area under curve (AUC) of the receiving operating characteristic (ROC) curve (along with its 95% confidence interval) is computed with the sensitivity and 1-specificity values (and 95% confidence intervals) for each alpha threshold. Additionally, a custom working point is defined as the first point with a sensitivity >90% (in order to use the model as a screening test, thus optimizing the sensitivity), and all the metrics (SE, SP, PPV, NPV, ACC) relevant to this point are stored, along with their 95% confidence intervals.

- Steps IV_{ij} to X_{ij} are repeated ten times (with j = 1...10), using each one of the ten groups as the validation set once (and the other 9 as the training set). Therefore, ten values are obtained for each one of the metrics computed in step X_{ij}.
- Steps II_i to X_{ij} are repeated an additional 5 times (boosting cycles, with i = 1...5), using a different random sampling of the balanced subset and a different random subdivision in the ten groups. Therefore, five series of the ten values for each one of the metrics computed in step X_{ij} are obtained, for a total of 50 values.

The final metrics used to evaluate the different algorithms (and the custom working points) are computed as the median of the distributions of 50 values. A graphical representation of the procedure is reported in figure 1, while detailed results for the different algorithms applied to the different models are reported in Table 3 (main manuscript).

Figure 1, graphical representation of the development and optimization process of a machine-learning algorithm to predict the positivity of patients rescued by EMS to SARS-CoV-2 in the pre-hospital emergency setting.



SIMULATION OF DECISION SUPPORT FOR PATIENTS' MANAGEMENT

An iterative procedure was implemented and applied on historical data, repeating the whole process on every week of records, starting from the fifth week in the dataset (October 29th, 2020 – November 4th, 2020), as the first four weeks are considered as an offset to have a large enough training set, and ending with the 42nd week of available records

(15th July 2021 – 21st July 2021), for a total of 38 cycles. Each cycle (relevant to a target week $\underline{w_f}$) is composed by the following procedure:

- Subdivision in training and test set:

The last week of data (collected during $\underline{w_f}$) is used as the test set (data on which a previously trained ML algorithm is applied), thus simulating a real-time running application. The training set (data on which the ML algorithm is trained) is instead composed by records that precede the beginning of the target week $\underline{w_f}$ ($w_0 \leq w_{f-1}$). Specifically, it is composed by all the available records relevant to patients that resulted positive to an rtPCR test in a time window of +-7 days, plus the same number of records (starting from the most recent, hence starting in w_{f-1} and adding backward) relevant to patients resulted negative to all the rtPCR tests performed in a time window of +-15 days. With this approach, the largest possible balanced dataset is obtained, while the not-random selection of negative records allows to avoid doubt cases, such as patients resulting negative in a time window of +-7 days but resulting positive in larger time windows. Before balancing the training set, all the outliers in the dataset are identified through the z-index method (threshold set to z-index = 5) and all records with one outlier value are removed. Finally, the training set is shuffled to guarantee randomization.

- Scaling and transformation, filling of test set:

All numerical variables are scaled in the 0-1 range, while all categorical variables are transformed in numerical dummy values (again in the 0-1 range). All variables with more than 15% of missing values are removed from the model. Scaling parameters and transformations are computed on the training set, and also applied to the test set. All missing values in the test set are filled with the median value across all records in the training set.

- Attributes selection:

In order to identify relevant attributes, a univariate logistic regression model is implemented on the training set with a ten-fold cross-validation protocol, and attributes with value '1' included in the 95% confidence interval of the odds ratio are discarded.

- Training and test of ML algorithm:

A ML algorithm is trained and tested on the previously described datasets; specifically, four different algorithms were tested: logistic regression, random forest, support vector machine and Bayesian naïve Gaussian. The considered output, rather than the assigned label, is the probability of classification.

- Performance evaluation: diagnostic capabilities

The predictive skill of the algorithm is evaluated through the area under curve (AUC) of the receiving operating characteristic (ROC) curve computed on the test set, considering the classification probability as the ranging 'alpha' threshold.

- Performance evaluation: identification of the cut-off threshold

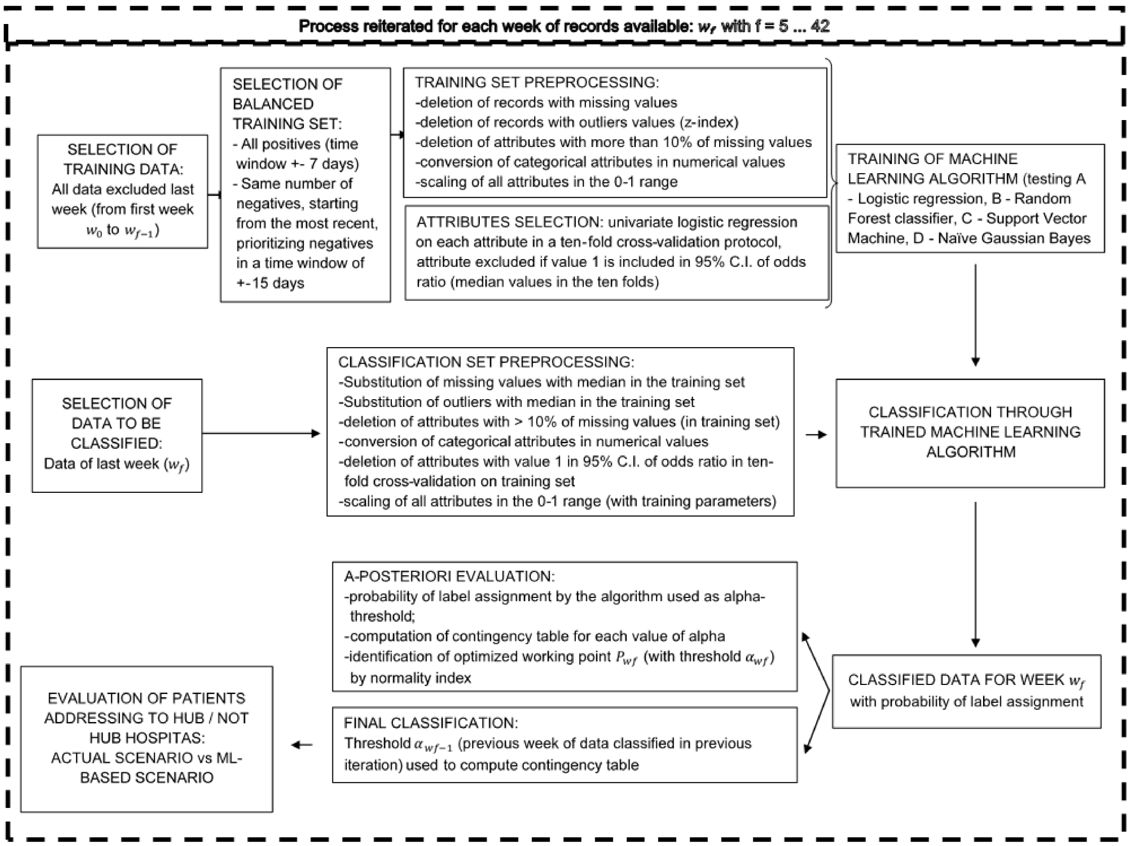
The working point (cut-off threshold) of the algorithm is identified by evaluating the Index of Union (IU) of a ten-fold cross-validation protocol performed on the training set.[1]

- Performance evaluation: patients management

The applicability of this model to support the decision about the destination of patients towards either 'hub' or 'non hub' hospitals is evaluated through the definition of the parameters described in the following.

A graphical representation of the procedure is reported in figure 2.

Figure 2, graphical representation of a machine-learning algorithm to simulate a real-time running decision making support for the destination of patients rescued by EMS towards COVID-19 hub or non-hub hospitals.



With:

Arbitrary criterion in low prevalence times (days with prevalence below median): positive patients addressed to hub hospitals, negative patients to non-hub hospitals.

Arbitrary criterion in high prevalence times (days with prevalence above median): positive patients presenting with severe features (i.e., SpO_2 <94% or RR >30) addressed to hub hospitals, negative patients presenting with severe features addressed to non-hub hospitals.

- **Bυ** baseline performance: patients for which the arbitrary criterion was met in the actual scenario; **Bυ** = $B\upsilon_p$ + $B\upsilon_n$, where $B\upsilon_p$ = baseline performance for positives and $B\upsilon_n$= baseline performance for negatives.
- **Bε**: baseline error: patients for which the arbitrary criterion was not reached in the actual scenario; **Bε** = $B\varepsilon_p$ + $B\varepsilon_n$, where $B\varepsilon_p$ = baseline error for positives and $B\varepsilon_n$= baseline error for negatives.

- **Mυ** model-based performance: patients for which the arbitrary criterion was met in the ML-based scenario, corresponding to model's true positives and true negatives; **Mυ** = $M\upsilon_p + M\upsilon_n$, where $M\upsilon_p$ = model-based performance for positives (true positives TP) and $M\upsilon_n$= model-based performance for negatives (true negatives TN); subdivided into:
  - Goal already reached, kept by the model (**Kt**): patients for which the arbitrary criterion was met in the actual scenario, whose destination would be the same on the base of the algorithm, can be subdivided in true positives (Kt_p) and true negatives (Kt_n).
  - New goal achievements (**Cr**): patients for which the arbitrary criterion was not reached in the actual scenario, but it is corrected in the ML-based scenario; again can be distinguished in true positives (Cr_p) and true negatives (Cr_n).

Therefore, **Mυ** = $Kt + Cr = (Kt_p + Kt_n) + (Cr_p + Cr_n) = M\upsilon_p + M\upsilon_n = (Kt_p + Cr_p) + (Kt_n + Cr_n)$

- **Mε** model-based error: patients for which the arbitrary criterion was not reached in the ML-based scenario, corresponding to model's false positives and false negatives; **Mε** = $M\varepsilon_p + M\varepsilon_n$, where $M\varepsilon_p$ = model-based error for positives (false positives FP) and $M\varepsilon_n$= model-based error for negatives (false negatives FN); subdivided into:
  - Unidentified possible improvements (**Ud**): patients for which the arbitrary criterion was not reached both in the actual scenario and in the ML-based scenario; subdivided into false negatives (Ud_p) and false positives (Ud_n).
  - Induced mistakes (**Im**): patients for which the arbitrary criterion was met in the actual scenario but is not in the ML-based scenario (added errors); again distinguished in false negatives (Im_p) and false positives (Im_n).

Therefore, **Mε** = $Ud + Im = (Ud_p + Ud_n) + (Im_p + Im_n) = M\varepsilon_p + M\varepsilon_n = (Ud_p + Im_p) + (Ud_n + Im_n)$

- **Δυ** performance differential: computed as **Mυ – Bυ,** also corresponding to **Kt + Cr – Bυ** and **Mυ – (Bε – Ud – Im**); derived measures are:
  - $\Delta\upsilon\%_{abs}$ percentage absolute performance differential, computed as (**Mυ – Bυ) / T** (total records)
  - $\Delta\upsilon\%_{rel}$ percentage relative performance differential, computed as (**Mυ – Bυ) / Bυ**

All parameters can be separately computed for positive and negative patients.

- **Δε** error differential: computed as **Mε – Bε,** also corresponding to **Ud + Im – Bε** and **Bυ – Kt – Cr** (hence corresponding to – Δυ); derived measures are:
  - $\Delta\varepsilon\%_{abs}$ percentage absolute error differential, computed as (**Mυ – Bυ) / T** (total records), corresponding to $-\Delta\upsilon\%_{abs}$
  - $\Delta\varepsilon\%_{rel}$ percentage relative error differential, computed as (**Mε – Bε) / Bε**

All parameters are also separately computed for positive and negative patients.

See below for complete results for the most performant algorithm (i.e., random forest classifier). Results for the other tested algorithms are available on request.

Spina S, *et al. Emerg Med J* 2023;0:1–11. doi: 10.1136/emermed-2022-212853

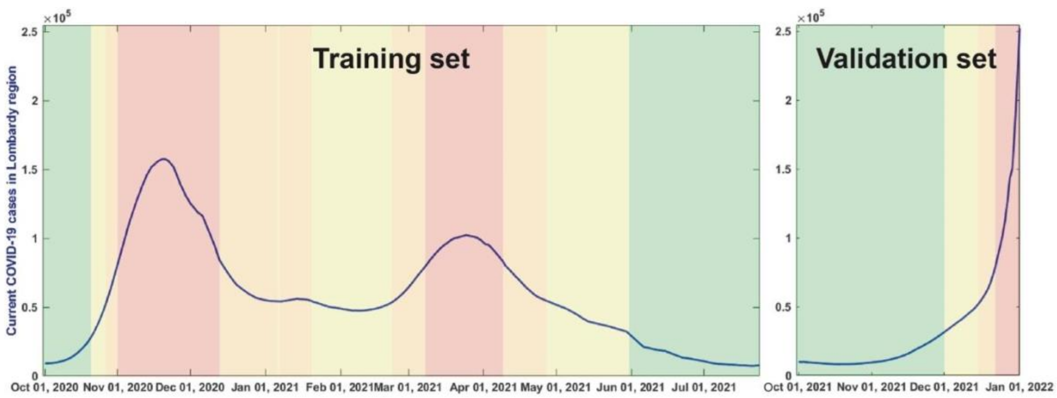
Python libraries used

“pandas”, “math”, “statistics”, “numpy”, “os”, “scipy”, “random”, “patsy”, “statsmodels”, “datetime”, “kaleido”, “sklearn” , “shap”, “matplotlib”

Table 1, Quartiles of COVID-19 prevalence in the study period

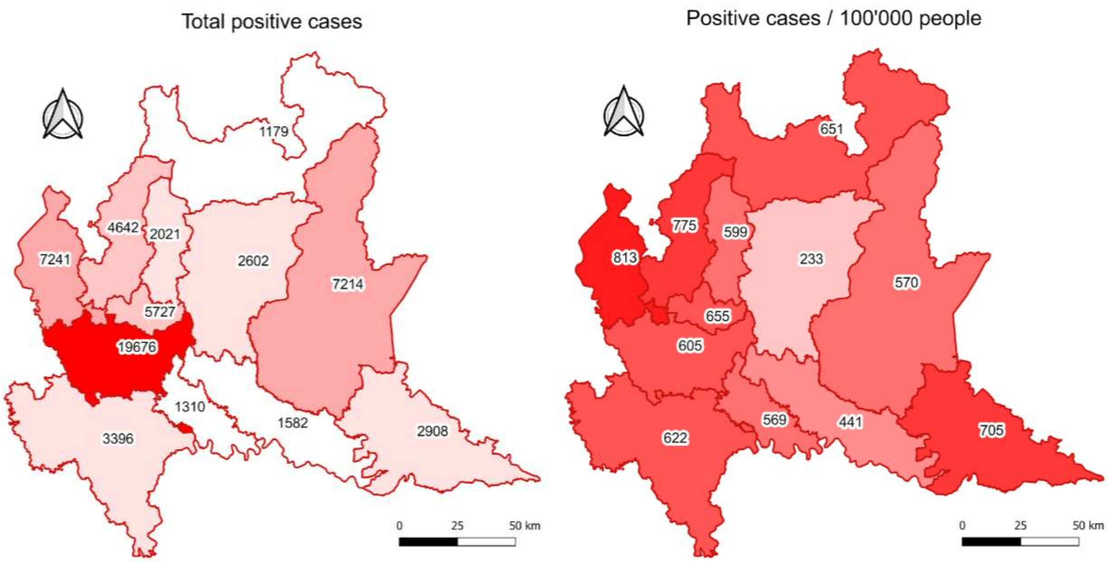
	Min value	Max value	Range	Dates	
1st Quartile	7325	28983	21658	Oct 1, 2020 – Oct 20, 2020 May 31, 2021 – Jul 23, 2021	Low prevalence period
2nd Quartile	30584	54144	23560	Oct 21, 2020 – Oct 26, 2020 Jan 6, 2021 Jan 20, 2021– Feb 21, 2021 Apr 27, 2021 – May 30, 2021	
3rd Quartile	54226	81436	27210	Oct 27, 2020 – Oct 31, 2020 Dec 13, 2020 – Jan 19, 2021 (excluding Jan 6, 2021) Feb 22, 2021 – Mar 7, 2021 Apr 9, 2021 – Apr 26, 2021	High prevalence period
4th Quartile	82171	157614	75443	Nov 1, 2020 – Dec 12, 2020 Mar 8, 2021 – Apr 8, 2021	

Figure 3, COVID-19 prevalence in the study period



Subdivision of days in the timeframe from Oct 1st, 2020, to Jul 23rd, 2021 (training set) and from Oct 1st, 2021, to Dec 31st, 2021, into four quartiles according to the number of patients currently positive for SARS-CoV-2 infection (filtered with a 7-day moving average) on the territory of Lombardy region, Italy. For each quartile of the training set, Table 1 reports the thresholds (minimum and maximum values), the range of values, and the timeframe assigned to it.

Figure 4, Positive cases distribution in the Lombardy region in the study period (October 1st 2020 – July 23rd 2021)



Lombardy map, divided into provinces. Color intensity is directly proportional to the number of cases in each province. The map on the left shows the absolute number of total positive cases in each province. The map on the right shows the number of positive cases per 100000 inhabitants in each province

Table 2, Detailed metrics of different machine learning algorithms in the training set.

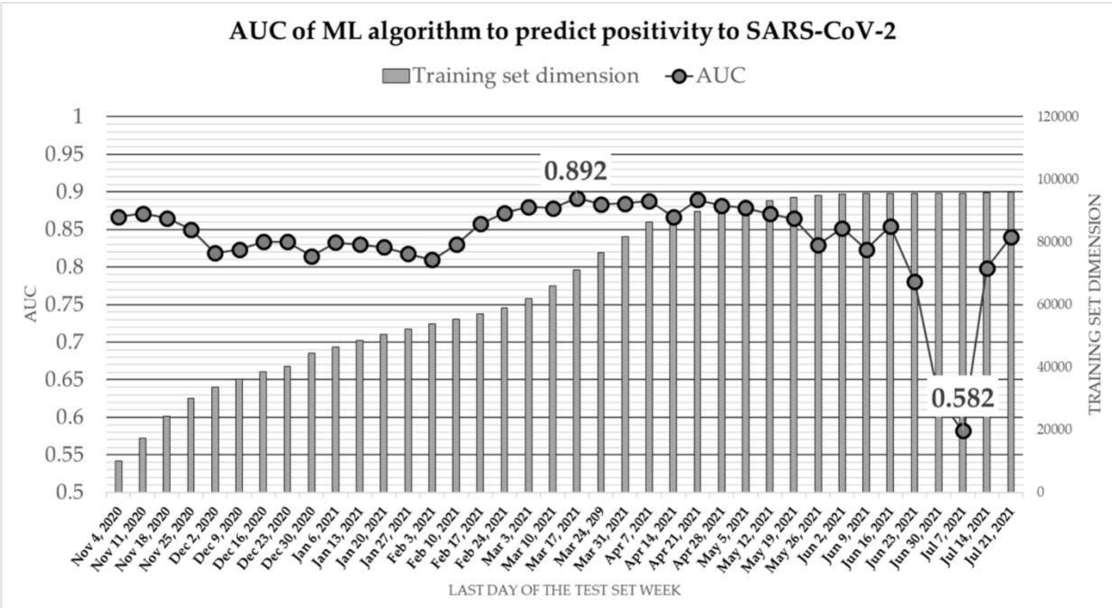
Model	Logistic Regression	Random Forest	Support Vector Machine	Naïve Gaussian Bayes
Model 1	AUC: .853 (.838 - .867) SENS: .944 (.936 - .952) SPEC: .26 (.251 - .27) ACC: .605 - (.595 - .615) PPV: .361 (.351 - .371) NPV: .91 (.899 - .921)	AUC: .853 (.838 - .867) SENS: .963 (.956 - .969) SPEC: .178 (.17 - .187) ACC: .577 (.566 - .587) PPV: .347 (.338 - .356) NPV: .913 (.9 - .926)	AUC: .803 (.79 - .815) SENS: .996 (.994 - .998) SPEC: .009 (.007 - .012) ACC: .504 (.493 - .516) PPV: .309 (.301 - .317) NPV: .844 (.751 - .909)	AUC: .803 (.79 - .815) SENS: .996 (.994 - .998) SPEC: .009 (.007 - .012) ACC: .504 (.493 - .516) PPV: .309 (.301 - .317) NPV: .844 (.751 - .909)
Model 2	AUC: .867 (.852 - .882) SENS: .913 (.903 - .922) SPEC: .461 (.449 - .473) ACC: .654 (.644 - .664) PPV: .448 (.436 - .46) NPV: .919 (.909 - .927)	AUC: .917 (.905 - .927) SENS: .902 (.892 - .912) SPEC: .696 (.686 - .706) ACC: .712 (.704 - .719) PPV: .571 (.558 - .584) NPV: .941 (.935 - .947)	AUC: .836 (.821 - .851) SENS: .979 (.974 - .984) SPEC: .048 (.043 - .053) ACC: .519 (.507 - .531) PPV: .329 (.32 - .338) NPV: .827 (.787 - .863)	AUC: .84 (.826 - .854) SENS: 1 (.999 - 1) SPEC: 0 (0 - 0) ACC: .5 (.488 - .512) PPV: .323 (.314 - .333) NPV: 0 (0 - 0)
Model 3	AUC: .883 (.869 - .895) SENS: .905 (.895 - .914) SPEC: .576 (.565 - .586) ACC: .688 (.679 - .696) PPV: .486 (.474 - .498) NPV: .932 (.924 - .939)	AUC: .917 (.905 - .927) SENS: .902 (.892 - .912) SPEC: .696 (.686 - .706) ACC: .712 (.704 - .719) PPV: .571 (.558 - .584) NPV: .941 (.935 - .947)	AUC: .835 (.821 - .848) SENS: .98 (.975 - .984) SPEC: .048 (.044 - .053) ACC: .521 (.51 - .532) PPV: .315 (.306 - .323) NPV: .841 (.804 - .873)	AUC: .84 (.827 - .852) SENS: 1 (.999 - 1) SPEC: 0 (0 - 0) ACC: .5 (.488 - .512) PPV: .307 (.299 - .316) NPV: 0 (0 - 0)
Model 4	AUC: .89 (.876 - .903) SENS: .905 (.894 - .915) SPEC: .607 (.595 - .618) ACC: .686 (.677 - .695) PPV: .528 (.514 - .541) NPV: .931 (.923 - .938)	AUC: .94 (.929 - .949) SENS: .902 (.891 - .912) SPEC: .809 (.799 - .818) ACC: .722 (.714 - .731) PPV: .691 (.677 - .705) NPV: .946 (.939 - .951)	AUC: .864 (.849 - .877) SENS: .967 (.96 - .973) SPEC: .134 (.126 - .142) ACC: .554 (.543 - .566) PPV: .346 (.336 - .356) NPV: .892 (.871 - .911)	AUC: .85 (.836 - .863) SENS: 1 (.999 - 1) SPEC: 0 (0 - 0) ACC: .5 (.488 - .512) PPV: .323 (.314 - .332) NPV: 0 (0 - 0)

The four models refer to different subset of attributes (see main text for details), while the different columns report the results of four different machine learning algorithms. AUC = Area Under Curve, where the curve is a ROC obtained by variating the threshold on the output probability of the algorithm. All the other metrics are obtained in a specific working point of this curve, arbitrarily set to be the first point with a sensitivity >0.9. SENS = sensitivity, SPEC = specificity, ACC = accuracy, PPV = positive predictive value, NPV = negative predictive value.

Spina S, *et al. Emerg Med J* 2023;0:1–11. doi: 10.1136/emermed-2022-212853

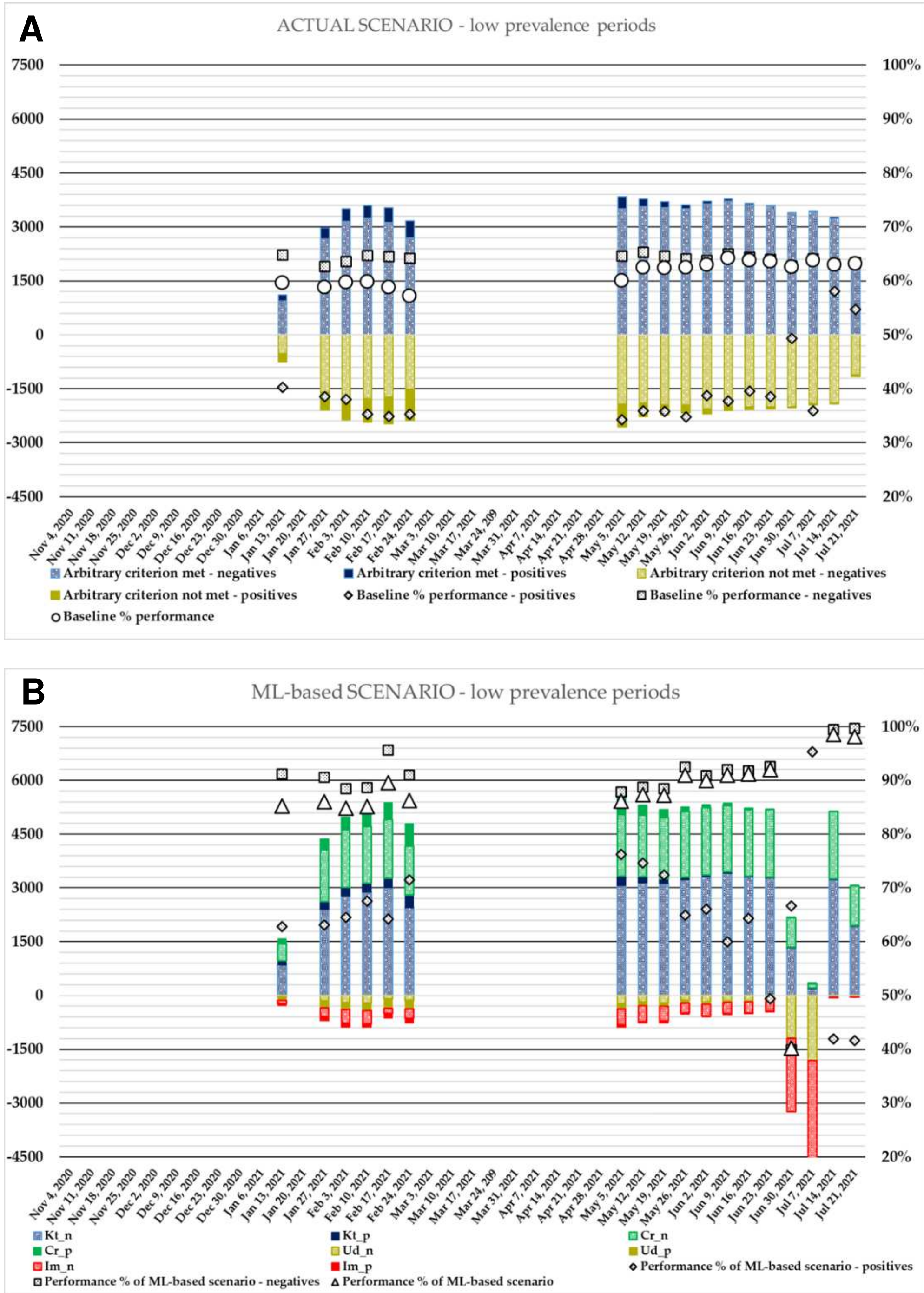
Real-world scenario simulation, details

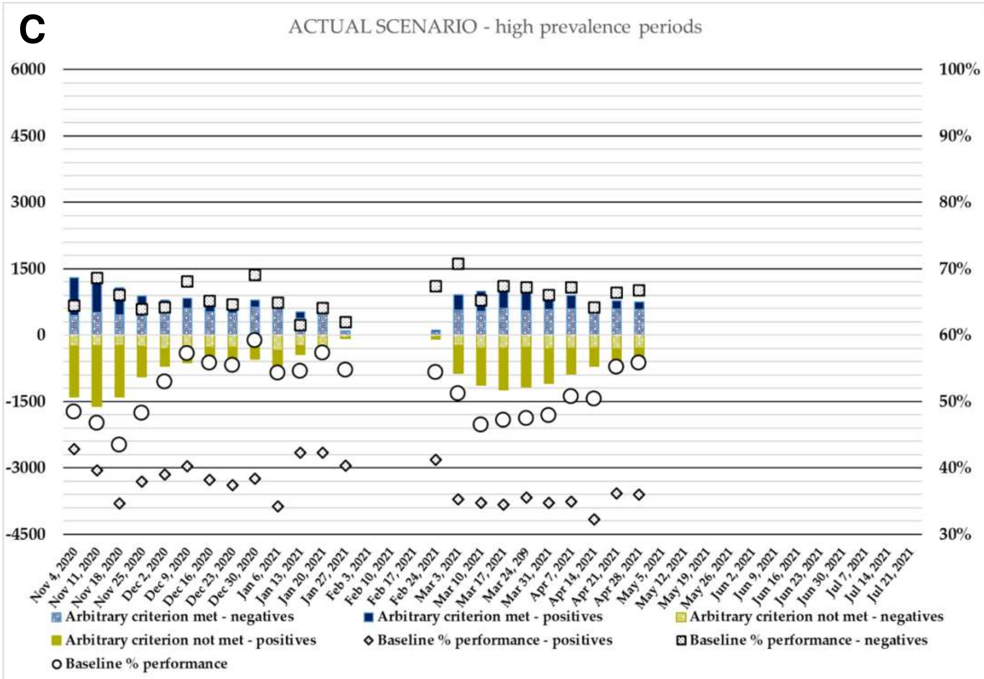
Figure 5

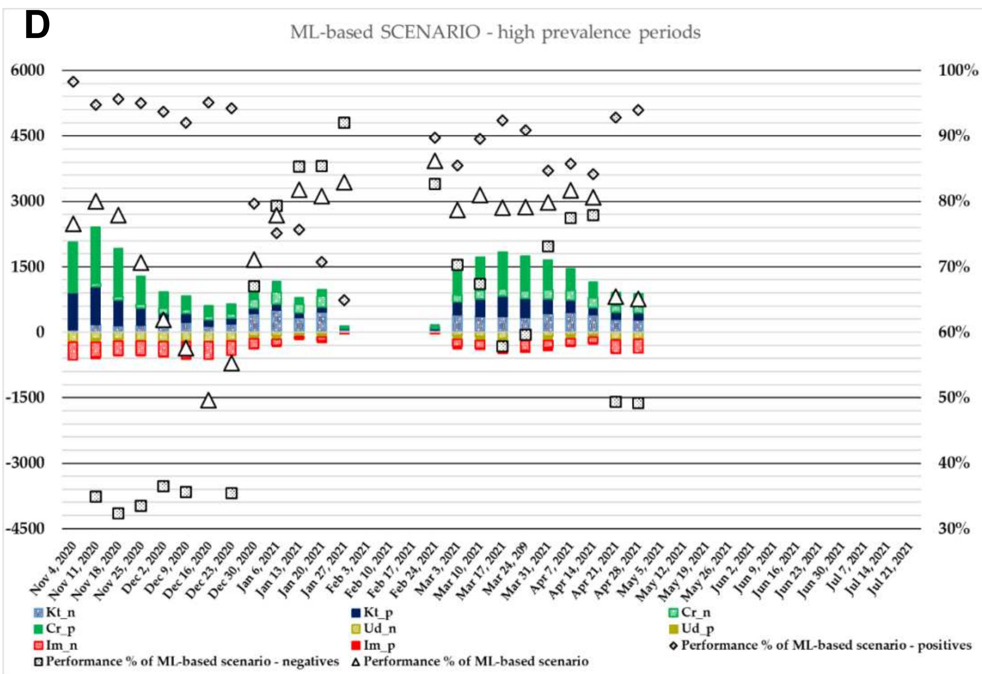


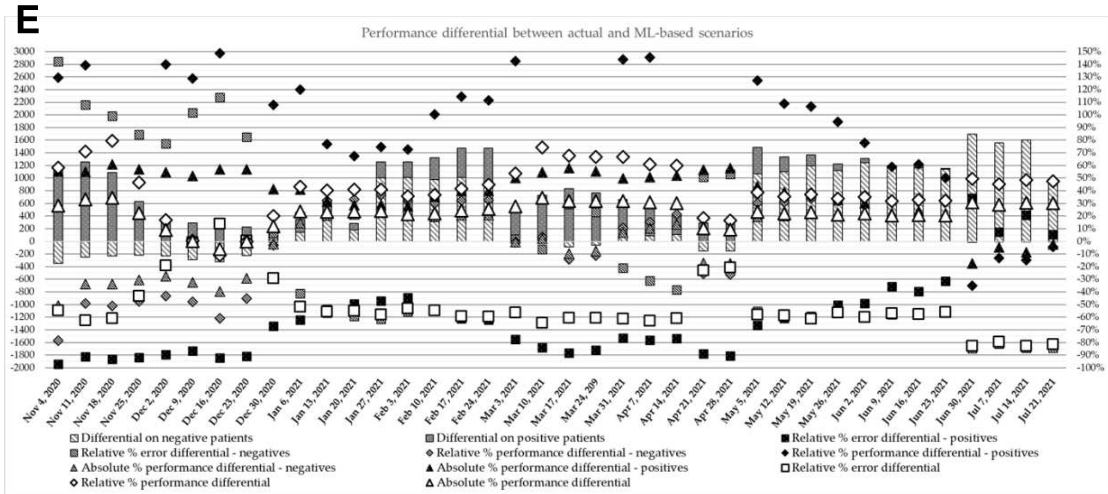
Results relevant to a machine learning algorithm (random forest) used to predict the positivity for SARS-CoV-2 infection in patients rescued by an ambulance on the territory of Lombardy region, Italy, between October 1st 2020 and July 21st 2021. The performance is measured through the Area Under Curve (AUC), plotted with dots, of a Receiver Operating Characteristic (ROC) curve computed on each available week of data (dates reported on X axis), using preceding records as the training set (dimension represented by columns). Details about the implementation of this iterative system are reported in Figure 2.

Figure 6


A
ACTUAL SCENARIO - low prevalence periods
Arbitrary criterion met - negatives
Arbitrary criterion met - positives
Arbitrary criterion not met - negatives
Arbitrary criterion not met - positives
Baseline % performance - positives
Baseline % performance - negatives
Baseline % performance
B
ML-based SCENARIO - low prevalence periods
Kt_n
Kt_p
Cr_n
Cr_p
Ud_n
Ud_p
Im_n
Im_p
Performance % of ML-based scenario - positives
Performance % of ML-based scenario - negatives
Performance % of ML-based scenario

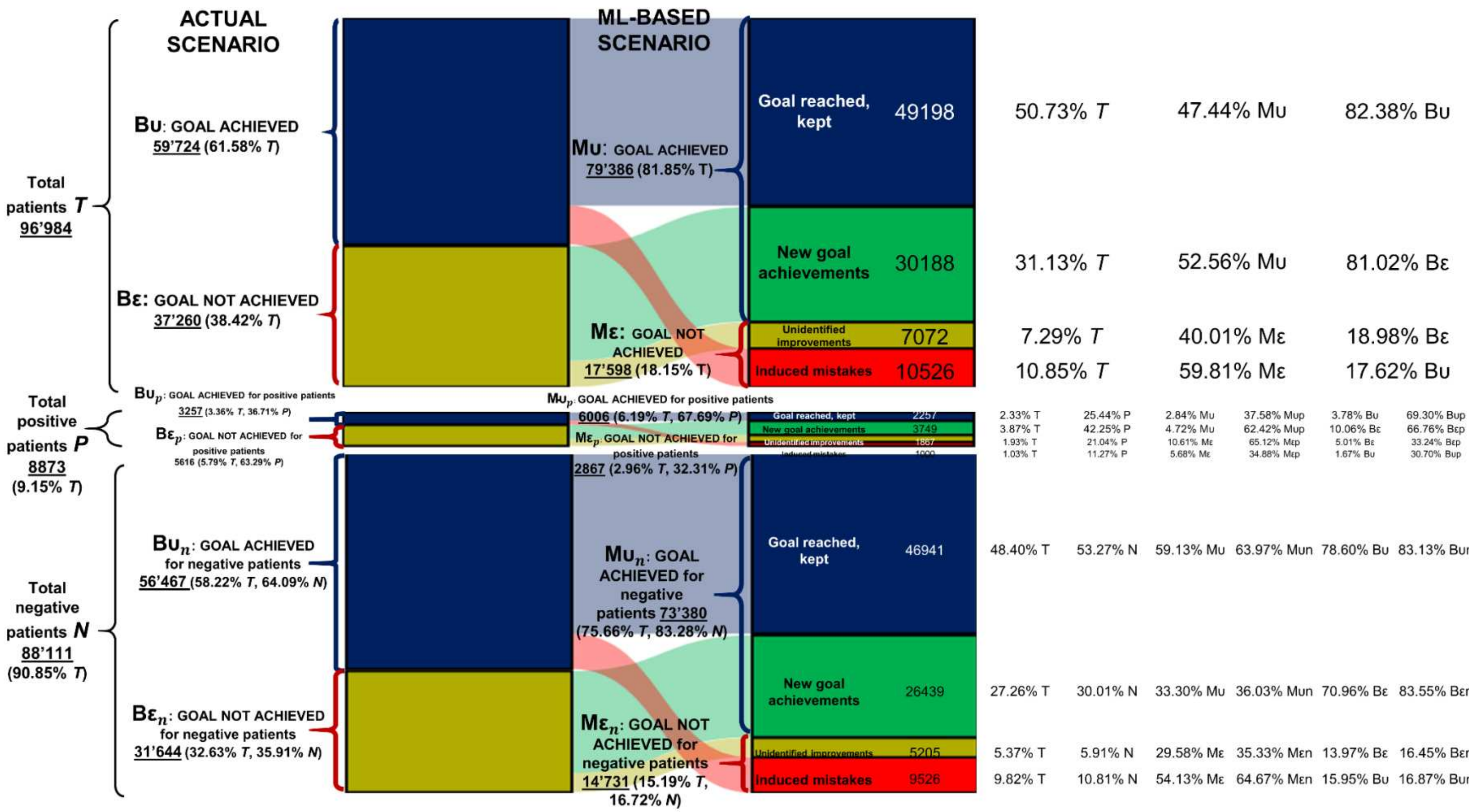

C
ACTUAL SCENARIO - high prevalence periods
Arbitrary criterion met - negatives
Arbitrary criterion met - positives
Arbitrary criterion not met - negatives
Arbitrary criterion not met - positives
Baseline % performance - positives
Baseline % performance - negatives
Baseline % performance


D
ML-based SCENARIO - high prevalence periods
Kt_n
Kt_p
Cr_n
Cr_p
Ud_n
Ud_p
Im_n
Im_p
Performance % of ML-based scenario - positives
Performance % of ML-based scenario - negatives
Performance % of ML-based scenario



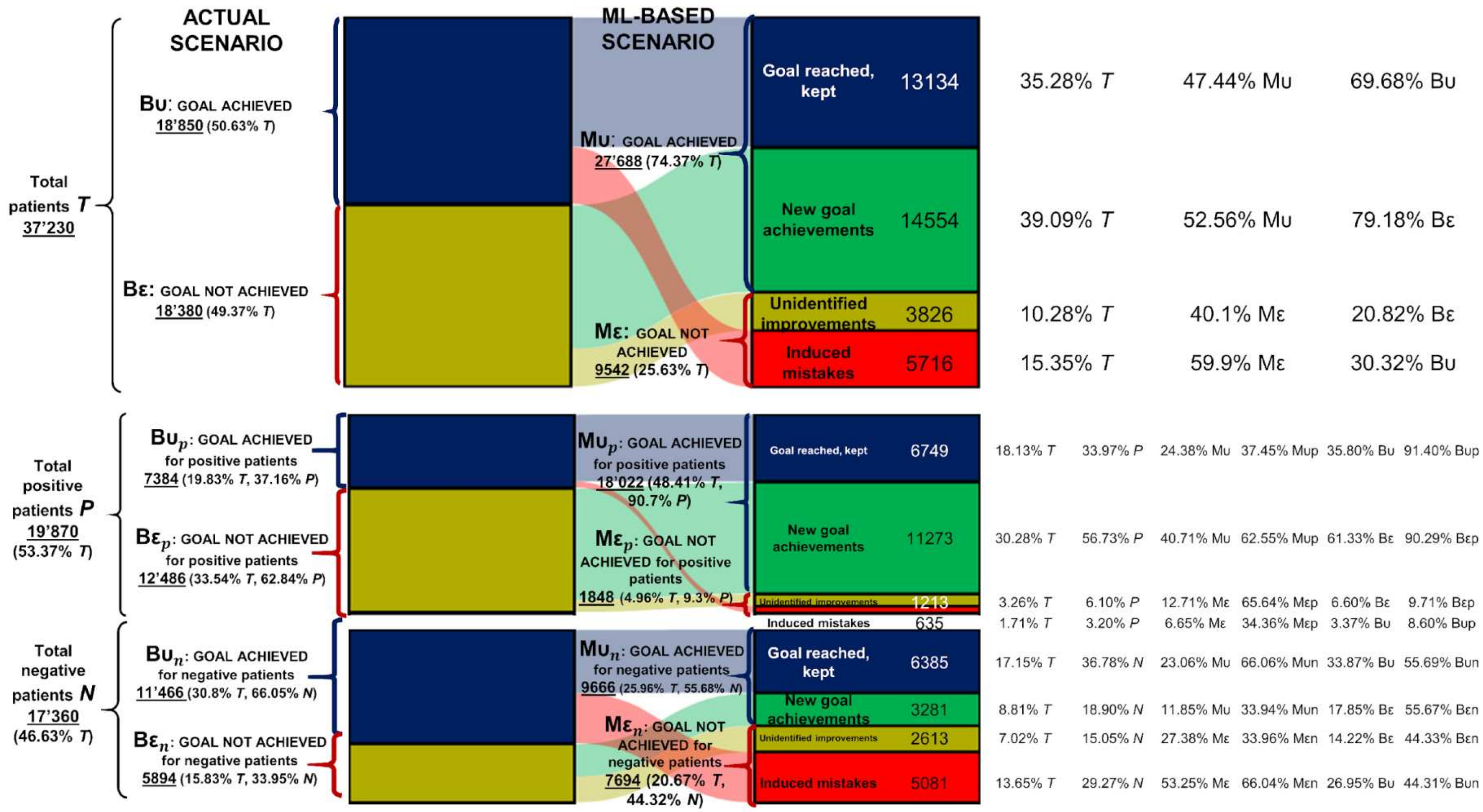
Management of patients rescued by an ambulance in the Lombardy region, Italy, between October 1st, 2020, and July 21st, 2021, separately for every week. Analysis of the destination towards COVID-19 specific hub hospitals or general hospitals, by comparing the actual scenario (panels A and C) with a hypothetical choice based on the machine learning model (panels B and D) re-trained each week on previous data, either for all patients during low prevalence periods (panels A and B) or for patients with severe features only during high prevalence periods (panels C and D); the subdivision between low and high prevalence periods is reported in Table 1 and Figure 3. The lower panel (E) reports the differential measures. In all panels, bars are referred to the left axis (absolute count), while markers are referred to the right axis (percentage values), always separately for patients which resulted (after rtPCR test) either positive or negative to COVID-19. In panels A and C, blue bars refer to patients that met the arbitrary criterion set (as described below Figure 2), while orange bars represent those for which the same criterion was not met; markers represent the percentage of patients that met the criterion on the total (circles), on positive patients (rhombuses) and on negative ones (squares). Panels B and D report the same values in case the decision was based on the suggestion of the implemented ML model, hence blue and green bars represent patients for whom the criterion was met, while orange and red bars represent those for whom it wasn't; blue and orange bars account for patients for whom the decision was the same in the actual scenario occurred and in the hypothetical ML-suggestion-based one, while green and red bars represent the number of patients for which the decision would have been changed (all categories are described in detail below Figure 2); markers once again represent the percentage of patients that met the criterion on the total (triangles), on positive patients (rhombuses) and on negative ones (squares). Panel E represents the differential in the performance between the currently occurred scenario and in the ML-based one: bars represent the difference in absolute values (separately for positive and negative patients), while markers represent the percentage variation in the absolute performance (triangles), in the relative (i.e., model – model / actual, see measures detailed description below Figure 2) performance (rhombuses) and in the relative error (squares), either for all patients (white markers), for positive ones (black markers) and for negative ones (grey markers).

Figure 7



Goal achievement based on arbitrary criteria (as described below Figure 2) to address patients to hospital resources, considering only patients in low prevalence periods (see Table 1, Figure 3), regarding the destination towards COVID-19 specific hub hospitals or general hospitals, comparing the actual scenario (left) with a hypothetical choice (right) based on the suggestion of a machine learning algorithm (random forest). The upper panel considers all patients, the middle panel is relevant to patients who resulted positive for SARS-CoV-2 infection, and the lower panel regards patients who resulted negative. A detailed description of all the parameters computed is provided below Figure 2.

Figure 8



Goal achievement based on arbitrary criteria (as described below Figure 2) to address patients to hospital resources, considering only patients with severe features in high prevalence periods (see Table 1, Figure 3), regarding the destination towards COVID-19 specific hub hospitals or general hospitals, comparing the actual scenario (left) with a hypothetical choice (right) based on the suggestion of a machine learning algorithm (random forest). The upper panel considers all patients, the middle panel is relevant to patients resulted positive for SARS-CoV-2 infection, and the lower panel regards patients resulted negative. A detailed description of all the parameters computed is provided below Figure 2.

Table 3, Detailed results of a machine learning model (random forest) used to predict the positivity to SARS-CoV-2 infection.

Each week is progressively used as the test set, while previous records are used for the training. The different parameters computed are described in detail below Figure 2 (‘perf.’ stands for ‘performance’). In the lower panel, TOTAL, MEDIAN, 25th Q(uartile), 75th Q(uartile) values are computed for the distribution of the different values across weeks.

	Training set	Test set (positives-negatives)	Attributes* removed (missing)	Attributes* removed (significance)	AUC	Cut-off	Baseline perf.	Model-based perf.	Perf. differential	Perf. differential - Positives	Perf. differential - Negatives	Error relative reduction (pos-neg)	Goal reached kept (%bv)	New goal achievement (%be)	Unidentified improvements (%be)	Induced mistakes (%bv)
Nov 04, 2020	10124	8198 (4287-3911)	12, 14	6, 10, 11	.867	.493	1303 (48.38%)	2063 (76.61%)	760 (28.22%)	1110(55.5%-129.5%)	-350 (-50.6%--78.5%)	-54.7% (-97.%|142.3%)	907 (69.61%)	1156 (83.17%)	234 (16.83%)	396 (30.39%)
Nov 11, 2020	17200	8188 (4459-3729)	12, 14	2, 6, 11, 16	.871	.501	1413 (46.77%)	2418 (80.04%)	1005 (33.27%)	1256 (55.1%-139.1%)	-251 (-33.8%--49.2%)	-62.5% (-91.3%|107.7%)	1037 (73.39%)	1381 (85.88%)	227 (14.12%)	376 (26.61%)
Nov 18, 2020	24356	7512 (3648-3864)	12, 14	6, 15, 16	.865	.502	1075 (43.42%)	1928 (77.87%)	853 (34.45%)	1087 (61.%-176.2%)	-234 (-33.7%--51.1%)	-60.9% (-93.3%|99.2%)	736 (68.47%)	1192 (85.08%)	209 (14.92%)	339 (31.53%)
Nov 25, 2020	30150	6525 (2368-4157)	12, 14	1, 6, 11, 16	.850	.505	881 (48.22%)	1291 (70.66%)	410 (22.44%)	630 (57.1%-150.4%)	-220 (-30.4%--47.6%)	-43.3% (-92.%|84.3%)	554 (62.88%)	737 (77.91%)	209 (22.09%)	327 (37.12%)
Dec 02, 2020	33752	6100 (1670-4430)	12, 14	6, 16	.819	.496	793 (52.97%)	927 (61.92%)	134 (8.95%)	364 (54.7%-140.%)	-230 (-27.7%--43.2%)	-19.% (-89.7%|77.2%)	437 (55.11%)	490 (69.6%)	214 (30.4%)	356 (44.89%)
Dec 09, 2020	36272	6144 (1469-4675)	12, 14	1, 2, 6, 11	.823	.489	831 (57.27%)	835 (57.55%)	4 (.28%)	292 (51.8%-128.6%)	-288 (-32.5%--47.7%)	-.6% (-86.6%|101.8%)	434 (52.23%)	401 (64.68%)	219 (35.32%)	397 (47.77%)
Dec 16, 2020	38490	5731 (1110-4621)	12, 14	2, 6, 16	.833	.506	689 (55.79%)	613 (49.64%)	-76 (-6.15%)	244 (56.9%-148.8%)	-320 (-39.7%--61.%)	13.9% (-92.1%|113.9%)	290 (42.09%)	323 (59.16%)	223 (40.84%)	399 (57.91%)
Dec 23, 2020	40232	5669 (1066-4603)	12, 14	1, 11, 16, 19	.834	.517	650 (55.41%)	648 (55.24%)	-2 (-.17%)	225 (56.8%-152.%)	-227 (-29.2%--45.2%)	.4% (-90.7%|82.5%)	327 (50.31%)	321 (61.38%)	202 (38.62%)	323 (49.69%)
Dec 30, 2020	44448	5961 (1094-4867)	12, 13, 14	1, 6, 17	.815	.492	794 (59.17%)	954 (71.09%)	160 (11.92%)	179 (41.3%-107.8%)	-19 (-2.1%--3.%)	-29.2% (-67.%|6.8%)	554 (69.77%)	400 (72.99%)	148 (27.01%)	240 (30.23%)
Jan 06, 2021	46314	6088 (1273-4815)	12, 13, 14	1, 6	.833	.518	809 (54.33%)	1160 (77.9%)	351 (23.57%)	210 (41.%-120.%)	141 (14.4%-22.2%)	-51.6% (-62.3%|-41.1%)	642 (79.36%)	518 (76.18%)	162 (23.82%)	167 (20.64%)
Jan 13, 2021	48496	6102 (1226-4876)	12, 13, 14	2, 6	.830	.481	1634 (57.88%)	2295 (81.3%)	661 (23.41%)	234 (31.6%-76.7%)	427 (20.5%-32.1%)	-55.6% (-53.8%|-56.6%)	1352 (82.74%)	943 (79.31%)	246 (20.69%)	282 (17.26%)
Jan 20, 2021	50558	5813 (949-4864)	12, 13, 14	1, 6	.827	.479	697 (57.32%)	983 (80.84%)	286 (23.52%)	107 (28.5%-67.3%)	179 (21.3%-33.3%)	-55.1% (-49.3%|-59.3%)	571 (81.92%)	412 (79.38%)	107 (20.62%)	126 (18.08%)
Jan 27, 2021	52152	5925 (966-4959)	12, 13, 14	2, 6	.818	.480	3072 (58.64%)	4328 (82.61%)	1256 (23.97%)	254 (28.8%-74.5%)	1002 (23.%-36.7%)	-58.% (-47.%|-61.6%)	2564 (83.46%)	1764 (81.4%)	403 (18.6%)	508 (16.54%)
Feb 03, 2021	53836	5869 (895-4974)	12, 13, 14	1, 2, 6	.810	.479	3503 (59.69%)	4760 (81.1%)	1257 (21.42%)	247 (27.6%-72.6%)	1010 (20.3%-31.9%)	-53.1% (-44.5%|-55.8%)	2861 (81.67%)	1899 (80.26%)	467 (19.74%)	642 (18.33%)
Feb 10, 2021	55358	6024 (993-5031)	12, 13, 14	1, 6	.830	.492	3603 (59.81%)	4927 (81.79%)	1324 (21.98%)	351 (35.3%-100.3%)	973 (19.3%-29.9%)	-54.7% (-54.6%|-54.7%)	2965 (82.29%)	1962 (81.04%)	459 (18.96%)	638 (17.71%)
Feb 17, 2021	57070	6013 (1162-4851)	12, 13, 14	1, 2, 6	.857	.487	3533 (58.76%)	5004 (83.22%)	1471 (24.46%)	463 (39.8%-114.3%)	1008 (20.8%-32.2%)	-59.3% (-61.2%|-58.5%)	2981 (84.38%)	2023 (81.57%)	457 (18.43%)	552 (15.62%)
Feb 24, 2021	59064	6466 (1581-4885)	12, 13, 14	1, 6	.872	.482	3277 (57.06%)	4748 (82.67%)	1471 (25.61%)	576 (39.9%-111.6%)	895 (20.8%-32.4%)	-59.7% (-62.%|-58.2%)	2717 (82.91%)	2031 (82.36%)	435 (17.64%)	560 (17.09%)
Mar 03, 2021	61862	7297 (2330-4967)	12, 13, 14	1, 2, 6	.880	.475	907 (51.16%)	1395 (78.68%)	488 (27.52%)	492 (50.3%-142.6%)	-4 (-.5%--.7%)	-56.4% (-77.6%|1.7%)	693 (76.41%)	702 (81.06%)	164 (18.94%)	214 (23.59%)
Mar 10, 2021	65908	7820 (2944-4876)	12, 13, 14	1, 6	.878	.489	986 (46.51%)	1717 (80.99%)	731 (34.48%)	713 (54.8%-157.7%)	18 (2.2%-3.4%)	-64.5% (-84.%|-6.3%)	768 (77.89%)	949 (83.69%)	185 (16.31%)	218 (22.11%)
Mar 17, 2021	71070	8109 (3153-4956)	12, 13, 14	1, 6	.892	.488	1101 (47.15%)	1846 (79.06%)	745 (31.91%)	831 (57.9%-168.2%)	-86 (-9.5%--14.2%)	-60.4% (-88.4%|29.3%)	823 (74.75%)	1023 (82.9%)	211 (17.1%)	278 (25.25%)
Mar 24, 209	76610	7818 (2964-4854)	12, 13, 14	1, 6, 17	.884	.495	1049 (47.42%)	1751 (79.16%)	702 (31.74%)	765 (55.3%-155.5%)	-63 (-7.6%--11.3%)	-60.4% (-85.9%|23.2%)	778 (74.17%)	973 (83.66%)	190 (16.34%)	271 (25.83%)
Mar 31, 2021	81810	7686 (2656-5030)	12, 13, 14	1, 6	.884	.499	994 (47.86%)	1658 (79.83%)	664 (31.97%)	602 (49.9%-143.7%)	62 (7.1%-10.8%)	-61.3% (-76.5%|-20.9%)	766 (77.06%)	892 (82.36%)	191 (17.64%)	228 (22.94%)
Apr 07, 2021	86500	7114 (2095-5019)	12, 13, 14	1, 6, 17	.888	.486	904 (50.73%)	1455 (81.65%)	551 (30.92%)	462 (50.8%-145.3%)	89 (10.2%-15.2%)	-62.8% (-78.%|-31.1%)	735 (81.31%)	720 (82.%)	158 (18.%)	169 (18.69%)
Apr 14, 2021	90248	6404 (1480-4924)	12, 13, 14	1, 6, 17	.867	.493	718 (50.42%)	1148 (80.62%)	430 (30.2%)	319 (52.%-161.1%)	111 (13.7%-21.3%)	-60.9% (-76.7%|-38.3%)	560 (77.99%)	588 (83.29%)	118 (16.71%)	158 (22.01%)

<table>
<tr><td>Apr 21, 2021</td><td>87636</td><td>6532 (1251-5281)</td><td>12, 14</td><td>1, 2</td><td>.889</td><td>.493</td><td>772 (55.22%)</td><td>915 (65.45%)</td><td>143 (10.23%)</td><td>293 (56.7%-156.7%)</td><td>-150 (-17.%--25.6%)</td><td>-22.8% (-88.8%|50.7%)</td><td>461 (59.72%)</td><td>454 (72.52%)</td><td>172 (27.48%)</td><td>311 (40.28%)</td></tr>
<tr><td>Apr 28, 2021</td><td>89722</td><td>6385 (1128-5256)</td><td>12, 14</td><td>1, 2</td><td>.882</td><td>.492</td><td>757 (55.83%)</td><td>882 (65.04%)</td><td>125 (9.22%)</td><td>279 (58.%-161.3%)</td><td>-154 (-17.6%--26.4%)</td><td>-20.9% (-90.6%|52.9%)</td><td>445 (58.78%)</td><td>437 (72.95%)</td><td>162 (27.05%)</td><td>312 (41.22%)</td></tr>
<tr><td>May 05, 2021</td><td>91580</td><td>6408 (968-5440)</td><td>12, 14</td><td>/</td><td>.879</td><td>.502</td><td>3844 (59.99%)</td><td>5332 (83.21%)</td><td>1488 (23.22%)</td><td>422 (43.6%-127.1%)</td><td>1066 (19.6%-30.4%)</td><td>-58.% (-66.4%|-55.3%)</td><td>3202 (83.3%)</td><td>2130 (83.07%)</td><td>434 (16.93%)</td><td>642 (16.7%)</td></tr>
<tr><td>May 12, 2021</td><td>93158</td><td>6067 (599-5468)</td><td>12, 14</td><td>/</td><td>.871</td><td>.497</td><td>3787 (62.42%)</td><td>5123 (84.44%)</td><td>1336 (22.02%)</td><td>234 (39.1%-108.8%)</td><td>1102 (20.2%-30.9%)</td><td>-58.6% (-60.9%|-58.1%)</td><td>3167 (83.63%)</td><td>1956 (85.79%)</td><td>324 (14.21%)</td><td>620 (16.37%)</td></tr>
<tr><td>May 19, 2021</td><td>94172</td><td>5943 (467-5476)</td><td>12, 14</td><td>1</td><td>.864</td><td>.487</td><td>3706 (62.36%)</td><td>5077 (85.43%)</td><td>1371 (23.07%)</td><td>178 (38.1%-106.6%)</td><td>1193 (21.8%-33.7%)</td><td>-61.3% (-59.3%|-61.6%)</td><td>3188 (86.02%)</td><td>1889 (84.44%)</td><td>348 (15.56%)</td><td>518 (13.98%)</td></tr>
<tr><td>May 26, 2021</td><td>94962</td><td>5790 (322-5468)</td><td>12, 14</td><td>1</td><td>.830</td><td>.503</td><td>3618 (62.49%)</td><td>4842 (83.63%)</td><td>1224 (21.14%)</td><td>106 (32.9%-94.6%)</td><td>1118 (20.4%-31.9%)</td><td>-56.4% (-50.5%|-57.%)</td><td>3017 (83.39%)</td><td>1825 (84.02%)</td><td>347 (15.98%)</td><td>601 (16.61%)</td></tr>
<tr><td>Jun 02, 2021</td><td>95508</td><td>5910 (209-5701)</td><td>12, 14</td><td>1, 2</td><td>.852</td><td>.506</td><td>3718 (62.91%)</td><td>5028 (85.08%)</td><td>1310 (22.17%)</td><td>63 (30.1%-77.8%)</td><td>1247 (21.9%-34.3%)</td><td>-59.8% (-49.2%|-60.4%)</td><td>3164 (85.1%)</td><td>1864 (85.04%)</td><td>328 (14.96%)</td><td>554 (14.9%)</td></tr>
<tr><td>Jun 09, 2021</td><td>95866</td><td>5901 (180-5721)</td><td>12, 14</td><td>1</td><td>.823</td><td>.513</td><td>3787 (64.18%)</td><td>4988 (84.53%)</td><td>1201 (20.35%)</td><td>40 (22.2%-58.8%)</td><td>1161 (20.3%-31.2%)</td><td>-56.8% (-35.7%|-58.%)</td><td>3199 (84.47%)</td><td>1789 (84.63%)</td><td>325 (15.37%)</td><td>588 (15.53%)</td></tr>
<tr><td>Jun 16, 2021</td><td>95224</td><td>5743 (129-5614)</td><td>12, 14</td><td>2</td><td>.854</td><td>.493</td><td>3664 (63.8%)</td><td>4863 (84.68%)</td><td>1199 (20.88%)</td><td>31 (24.%-60.8%)</td><td>1168 (20.8%-32.3%)</td><td>-57.7% (-39.7%|-58.4%)</td><td>3062 (83.57%)</td><td>1801 (86.63%)</td><td>278 (13.37%)</td><td>602 (16.43%)</td></tr>
<tr><td>Jun 23, 2021</td><td>95448</td><td>5658 (83-5574)</td><td>12, 14</td><td>2</td><td>.780</td><td>.501</td><td>3598 (63.6%)</td><td>4753 (84.02%)</td><td>1155 (20.42%)</td><td>16 (19.3%-50.%)</td><td>1139 (20.4%-31.9%)</td><td>-56.1% (-31.4%|-56.7%)</td><td>2973 (82.63%)</td><td>1780 (86.45%)</td><td>279 (13.55%)</td><td>625 (17.37%)</td></tr>
<tr><td>Jun 30, 2021</td><td>95588</td><td>5438 (75-5363)</td><td>12, 14</td><td>/</td><td>.625</td><td>.491</td><td>3401 (62.54%)</td><td>5083 (93.47%)</td><td>1682 (30.93%)</td><td>-13 (-17.3%--35.1%)</td><td>1695 (31.6%-50.4%)</td><td>-82.6% (34.2%|-84.8%)</td><td>3182 (93.56%)</td><td>1901 (93.32%)</td><td>136 (6.68%)</td><td>219 (6.44%)</td></tr>
<tr><td>Jul 07, 2021</td><td>95476</td><td>5393 (64-5329)</td><td>12, 14</td><td>/</td><td>.582</td><td>.495</td><td>3439 (63.77%)</td><td>4990 (92.53%)</td><td>1551 (28.76%)</td><td>-3 (-4.7%--13.%)</td><td>1554 (29.2%-45.5%)</td><td>-79.4% (7.3%|-81.2%)</td><td>3170 (92.18%)</td><td>1820 (93.14%)</td><td>134 (6.86%)</td><td>269 (7.82%)</td></tr>
<tr><td>Jul 14, 2021</td><td>95584</td><td>5215 (81-5134)</td><td>12, 14</td><td>/</td><td>.799</td><td>.509</td><td>3284 (62.97%)</td><td>4880 (93.58%)</td><td>1596 (30.6%)</td><td>-7 (-8.6%--14.9%)</td><td>1603 (31.2%-49.5%)</td><td>-82.7% (20.6%|-84.5%)</td><td>3070 (93.48%)</td><td>1810 (93.73%)</td><td>121 (6.27%)</td><td>214 (6.52%)</td></tr>
<tr><td>Jul 21, 2021</td><td>95728</td><td>3141 (84-3057)</td><td>12, 14</td><td>/</td><td>.840</td><td>.507</td><td>1983 (63.13%)</td><td>2927 (93.19%)</td><td>944 (30.05%)</td><td>-2 (-2.4%--4.3%)</td><td>946 (30.9%-48.8%)</td><td>-81.5% (5.3%|-84.5%)</td><td>1836 (92.59%)</td><td>1091 (94.21%)</td><td>67 (5.79%)</td><td>147 (7.41%)</td></tr>
<tr><td>TOTAL</td><td>/</td><td>240’100 (53’478-186’620)</td><td>/</td><td>/</td><td>/</td><td>/</td><td>136336 (58.01%)</td><td>199355 (84.82%)</td><td>63019 (26.81%)</td><td>20087 (40.8%-115.6%)</td><td>42932 (23.1%-36.1%)</td><td>-63.8% (-63.%|-64.3%)</td><td>116230 (85.25%)</td><td>83125 (84.22%)</td><td>15580 (15.78%)</td><td>20106 (14.75%)</td></tr>
<tr><td>MEDIAN</td><td>46859</td><td>5827 (500-4824)</td><td>/</td><td>/</td><td>.824</td><td>.488</td><td>3611 (58.73%)</td><td>5179 (84.26%)</td><td>1522 (25.62%)</td><td>368 (34.7%-93.6%)</td><td>1100 (22.4%-34.5%)</td><td>-63.3% (-55.7%|-63.6%)</td><td>3122 (84.7%)</td><td>2033 (83.62%)</td><td>437 (16.38%)</td><td>559 (15.3%)</td></tr>
<tr><td>25TH Q</td><td>93918</td><td>6530 (1989-5317)</td><td>/</td><td>/</td><td>.872</td><td>.502</td><td>3472 (55.89%)</td><td>5026 (82.07%)</td><td>1467 (24.74%)</td><td>121 (27.9%-71.7%)</td><td>955 (19.5%-30.6%)</td><td>-67.3% (-66.%|-70.2%)</td><td>2978 (82.61%)</td><td>1913 (81.75%)</td><td>278 (12.56%)</td><td>439 (12.04%)</td></tr>
<tr><td>75TH Q</td><td>10124</td><td>8198 (4287-3911)</td><td>/</td><td>/</td><td>.867</td><td>.493</td><td>3787 (62.47%)</td><td>5393 (87.74%)</td><td>1820 (29.27%)</td><td>641 (43.2%-125.4%)</td><td>1313 (25.3%-39.3%)</td><td>-59.1% (-45.8%|-53.5%)</td><td>3315 (87.96%)</td><td>2303 (87.44%)</td><td>542 (18.25%)</td><td>687 (17.39%)</td></tr>
</table>

References

1 Unal I. Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Comput Math Methods Med* 2017; **2017**: 3762651.