





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/emmermed-2022-212853>).

¹SOREU, Agenzia Regionale Emergenza Urgenza (AREU), Milano, Italy

²Department of Anesthesia, Critical Care and Pain Medicine, ASST Grande Ospedale Metropolitano Niguarda, Milano, Italy

³Electronics, Information and Biomedical Engineering Department, Politecnico di Milano, Milano, Italy

⁴Department 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



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

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/emmermed-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 km². 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 (SpO₂), 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 \pm 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

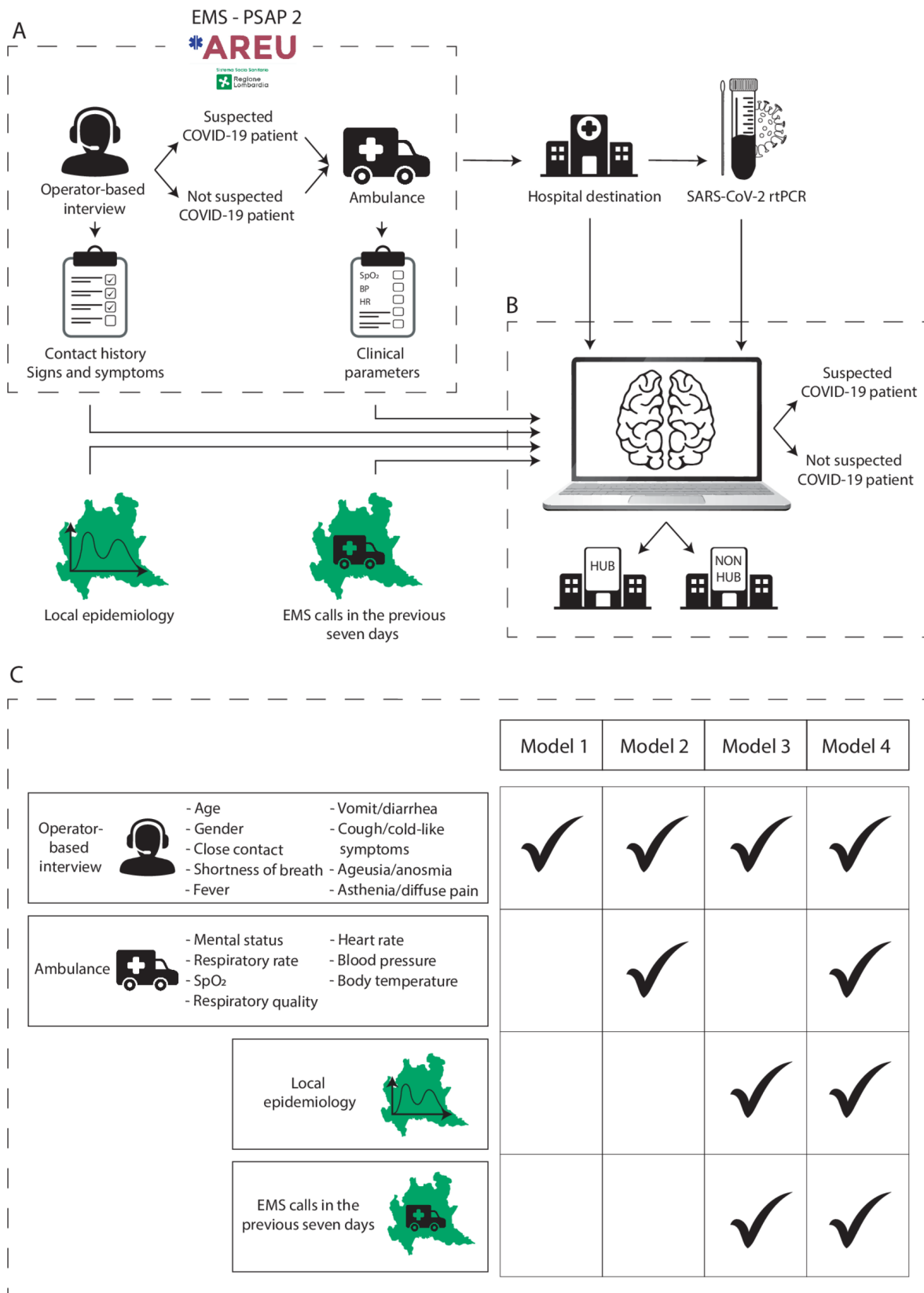


Figure 1 Model description. The performance of the operator-based interview in detecting patients with COVID-19 is evaluated by matching the results of the available SARS-CoV-2 rtPCR (box A). The machine learning models are implemented considering different combinations of the variables of the operator-based 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; SpO₂, 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, SpO₂<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 SpO₂<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)		Validation set (n=103 336)	
		Missing, n (%)		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; SpO₂, 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 SpO₂ 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 rtPCR-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)
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)
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)
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 \geq 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

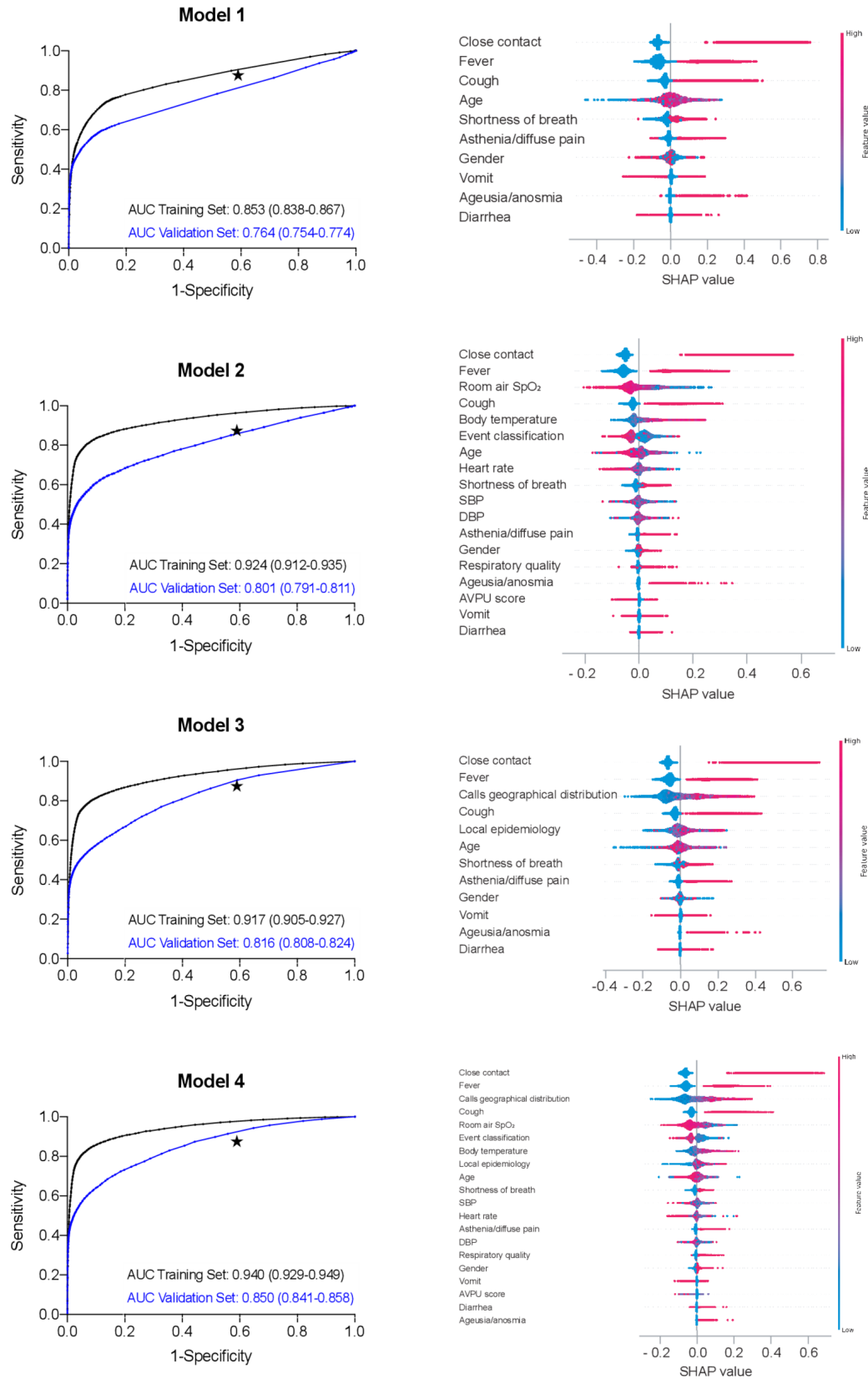


Figure 2 Machine learning models performance. *Left*: receiver operating characteristic (ROC) curves of the four models implemented, as compared with the performance of the operator-based interview (*). The box within each graph reports the AUC with the respective 95% CI, relevant to the random forest 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; SpO₂, pulse oximeter oxygen saturation.

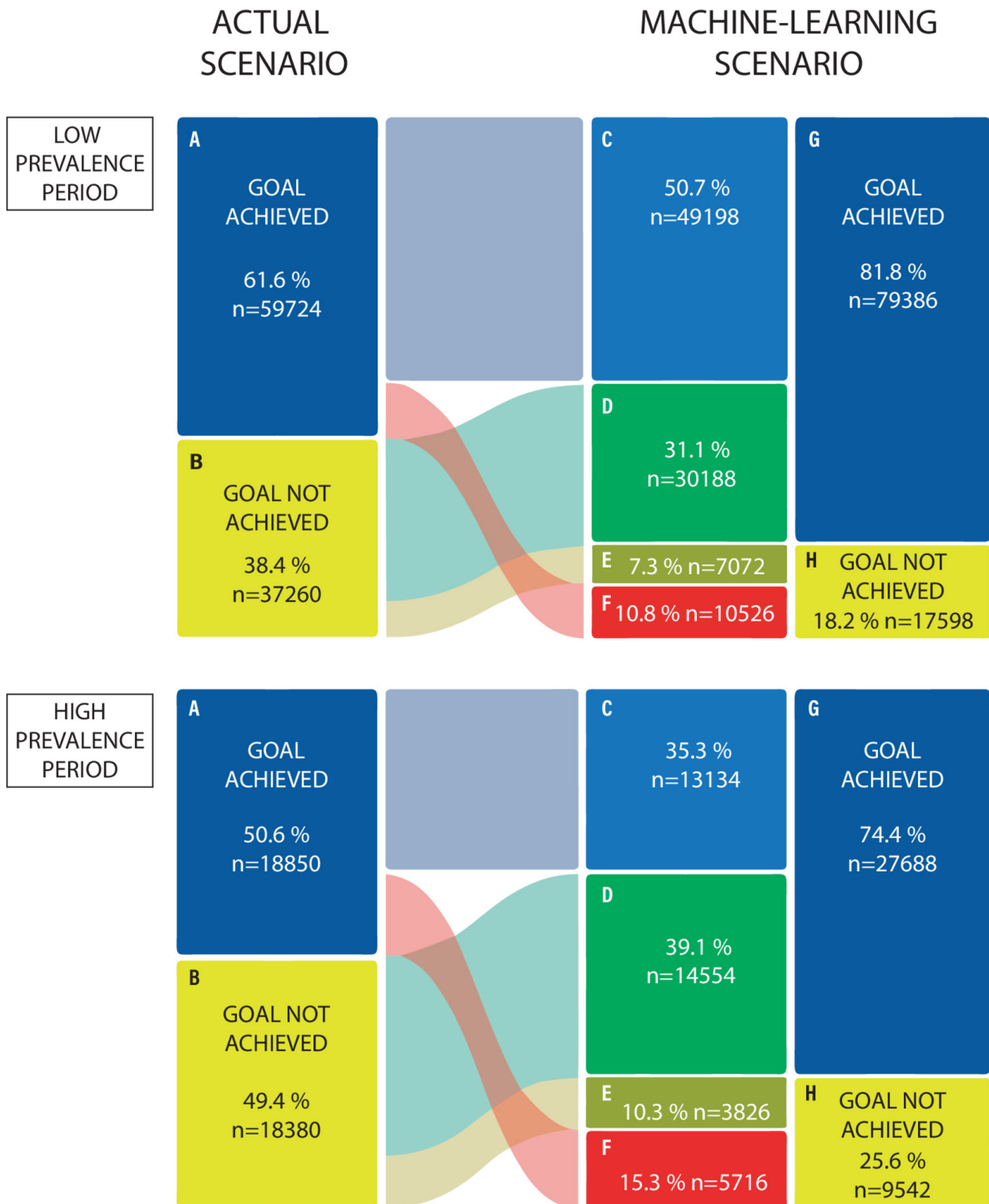


Figure 3 Real-world scenario simulation. *Upper panel:* a first simulation was performed in the low prevalence period (n=96 984). Here, positive patients would be addressed to hub hospitals, while negative patients to non-hub hospitals. Based on actual data (ie, actual scenario), the goal was achieved 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 Migliari) 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.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Stefano Spina <http://orcid.org/0000-0003-3210-1714>

Roberto Fumagalli <http://orcid.org/0000-0003-0398-1329>

REFERENCES

- Arsenault C, Gage A, Kim MK, *et al*. COVID-19 and resilience of Healthcare systems in ten countries. *Nat Med* 2022;28:1314–24.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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>
- 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.
- Zieliński W. The shortest Clopper–Pearson confidence interval for binomial probability. *Communications in Statistics - Simulation and Computation* 2009;39:188–93.
- Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;2017:4766–75.
- Grasselli G, Zanfrillo 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Trevethan R. Specificity, and predictive values: foundations, Pliabilities, and pitfalls in research and practice. *Front Public Health* 2017;5:307.
- 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.