

Use of Machine Learning to Rapidly Predict Positivity to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) Using Basic Clinical Data

Thomas Langer (✉ thomas.langer@unimib.it)

Università degli Studi di Milano-Bicocca <https://orcid.org/0000-0002-9725-8520>

Martina Favarato

Università degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Riccardo Giudici

Azienda Socio Sanitaria Territoriale Grande Ospedale Metropolitano Niguarda

Gabriele Bassi

niguard

Roberta Garberi

Università degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Fabiana Villa

Università degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Hedwige Gay

Università degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Anna Zeduri

Università degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Sara Bragagnolo

Università degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Alberto Molteni

Università degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Andrea Beretta

Azienda Socio Sanitaria Territoriale Grande Ospedale Metropolitano Niguarda

Matteo Corradin

Azienda Socio Sanitaria Territoriale Grande Ospedale Metropolitano Niguarda

Mauro Moreno

Azienda Socio Sanitaria Territoriale Grande Ospedale Metropolitano Niguarda

Chiara Vismara

Azienda Socio Sanitaria Territoriale Grande Ospedale Metropolitano Niguarda

Carlo Federico Perno

Azienda Socio Sanitaria Territoriale Grande Ospedale Metropolitano Niguarda

Massimo Buscema

University of Colorado Denver

Enzo Grossi

Bracco SPA

Roberto Fumagalli

Universita degli Studi di Milano-Bicocca Scuola di Medicina e Chirurgia

Original research

Keywords: Severe Acute Respiratory Syndrome Coronavirus 2, Emergency Service, Hospital, Pandemics, Critical Care, Artificial Intelligence, Supervised Machine Learning

Posted Date: July 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-38576/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on December 1st, 2020. See the published version at <https://doi.org/10.1186/s13049-020-00808-8>.

Abstract

Objective: Reverse Transcription-Polymerase Chain Reaction (RT-PCR) for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) diagnosis currently requires quite a long time span. A quicker and more efficient diagnostic tool in emergency departments could improve management during this global crisis. Our main goal was assessing the accuracy of artificial intelligence in forecasting the results of RT-PCR for SARS-COV-2, using basic information at hand in all emergency departments.

Methods: This is a retrospective study carried out between February 22 and March 16 2020 in one of the main hospitals in Milan, Italy. We screened for eligibility all patients admitted with influenza-like symptoms tested for SARS-COV-2. Patients under 12 years old, with no leukocyte formula performed in the ED, were excluded. Input data through artificial intelligence were made up of a combination of clinical, radiological and routine laboratory data upon hospital admission.

Results: Among 199 patients subject to study (median [interquartile range] age 65 [46-78] years; 127 [63.8%] men), 124 [62.3%] resulted positive to SARS-COV-2. The best Machine Learning System reached an accuracy of 91.4% with 94.1% sensitivity and 88.7% specificity.

Conclusion: Our study suggests that properly trained artificial intelligence algorithms may be able to predict correct results in RT-PCR for SARS-COV-2, using basic clinical data. If confirmed, on a larger-scale study, this approach could have important clinical and organizational implications.

Introduction

At the end of 2019, an outbreak of pneumonia, of unknown origin at the time, turned out to have stemmed from Wuhan, Hubei, China and consequently spread throughout the world, reaching Italy in February 2020 (1, 2).

A new betacoronavirus, subsequently named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2), was identified as the cause of the epidemic (3).

The infection with SARS-COV-2 has, in humans, a broad spectrum of clinical presentations (4) and was named Coronavirus Disease 2019 (COVID-19).

Initial symptoms are nonspecific and similar to other seasonal viral diseases which entail fever, dry cough and fatigue. This could lead to a fairly impossible clinical diagnosis. Therefore, etiological diagnosis relies on Reverse Transcription-Polymerase Chain Reaction (RT-PCR) to detect the genome of SARS-COV-2.

There are several important limitations to RT-PCR. First, current techniques take up to 6-8 hours in order to obtain plausible results and often laboratories cannot handle the overload. Second, not all hospitals have the equipment and staff to run tests 24/7. Consequently, swabs are sent out to other facilities, thus slowing down the process and swamping central laboratories. Third, RT-PCR, on a retropharyngeal swab, may result falsely negative in the initial phase of the disease, in spite of the presence of typical symptoms (5-7).

Last of all, this technique carries a certain cost which could mean a considerable financial burden weighing upon both health systems and patients.

A direct consequence of these limitations is the time spent by a large number of patients awaiting results in the emergency department before a decision can be taken as to where admit them to, e.g. in wards and intensive care units focused on COVID-19 patients, or in “non-infective” wards of the hospital (8-12).

This is particularly troublesome for critically ill patients requiring immediate endotracheal intubation and mechanical ventilation. Specialized medical staff need to attend to these patients in emergency departments where, however, the health system is already under a lot of pressure.

Finding an easy and fast method of predicting positivity or negativity to SARS-COV-2, would prove to be of great clinical value. Algorithms have already been proposed using advanced imaging, e.g. chest computed tomography (CT) (13, 14). However, not all hospitals or countries can carry out a CT scan on every patient.

The main goal of this testing was to pinpoint the accuracy of artificial intelligence in predicting RT-PCR for SARS-COV-2 results using basic clinical, radiological and routine laboratory data at hand in all emergency departments.

Our theory was that using Artificial Neural Networks (ANNs), and other Machine Learning Systems (MLS), would lead to obtaining accurate results on RT-PCR testing for SARS-COV-2 and that these systems could possibly be applied in the future. The performance of different ANNs and MLS was analyzed so as to distinguish between patients resulting positive or negative to SARS-COV-2, thus identifying those variables which express the maximum amount of relevant information.

Methods

Study design and selection of Participants

This retrospective, single centre study, was approved by the Institutional Review Board of our hospital (N°3733). Mandatory informed consent was waived.

All patients admitted to the emergency department of our hospital between February 22 and March 16 2020, were screened for eligibility.

Symptoms of presentation compatible with COVID-19 (fever, sore throat, cough, dyspnoea, chest pain, headache, syncope, asthenia, arthralgia, diarrhoea, nausea and vomit) constituted the inclusion criteria. Age < 12 years and absence of evaluation of the leukocyte formula (defined as percentages of the five types of leukocytes: neutrophils, lymphocytes, eosinophils, basophils and monocytes) in the emergency department constituted exclusion criteria.

Data collection

Clinical data were taken from the Patient Data Management System of our hospital. Demographic data included age, gender, presence and type of comorbidities, medication currently being taken. Each drug was placed in a specific category.

In addition, information regarding vital signs upon admission to the emergency department, presence and type of ventilatory support, routinely performed blood tests, major electrocardiographic characteristics (presence of sinus rhythm and ST abnormalities) and chest X-rays (presence of any type of parenchymal involvement, presence of pleural effusion) were collected.

The final results of the RT-PCR swab for SARS-COV-2 were recorded. Should the outcome have proved negative in a symptomatic patient, our hospital made it mandatory for a second swab to be carried out after 48 hours and these swabs were checked to confirm negativity. The complete list of collected variables is summed up in Table 1.

Statistical analysis and sample size

All data were tested for consistency with variance and normality of distribution using the Shapiro-Wilk test. Normally distributed data were expressed as mean \pm standard deviation, while non-normally distributed data were reported as median and interquartile range. Binary data were summed up in percentages, frequency of occurrence, and compared through Chi-Square test. Continuous variables were compared through Student T-Test or Wilcoxon Rank-Sum test, as appropriate.

Pearson's and Spearman's correlation was used to assess the correlation between collected variables (continuous and nominal, respectively) and RT-PCR for SARS-COV-2 results. A P-value lower than 0.05 was considered as statistically significant. Analysis was performed with SigmaPlot v.12.0 (Systat Software Inc., USA).

The study was carried out on a convenience sample of 200 patients admitted to the emergency department and tested for SARS-COV-2.

Machine Learning Methods

In order to predict and estimate the results of the RT-PCR for SARS-COV-2 using an input of the 74 variables under study (Table 1), different Machine Learning algorithms available on WEKA data mining software (15-17) and on Semeion Research Centre depository (Massimo Buscema, Deep Supervised ANNs, Semeion SW #12, version 33.0, 1999-2019) were trained. These classification tools were applied to predict RT-PCR results using the Training and Testing validation protocol, with the following steps:

1. Subdivision of the dataset into two sub-samples, A and B, each containing 50% of records and having an equal proportion of cases and controls (in our case SARS-COV-2 positive and SARS-COV-2 negative). The two sub-samples were obtained through the application of the TWIST algorithm, (i.e. they were not obtained by random extraction), in order to create two subsamples with similar probability density for all the input variables (see below). A homogeneity check was performed to confirm the substantial equivalence of the two subsets with regard to the distribution of variables. In the first run, A was used as Training Set and B as Testing Set.
2. Application of ANN on the Training Set. In this phase, the ANN learned to associate the input variables with those indicated as targets.

3. After the training phase, the weights matrix produced by the algorithm was saved and frozen together with all parameters used for the training.
4. The Testing Set was then shown to a virgin twin (same architecture and base parameters) ANN with the same weights matrix of the trained ANN, acting as final classifier. This operation took place for all records and the results (right or wrong classification) was not communicated to the classifier. This allowed to assess the generalization ability of trained ANN.
5. In a second run, another virgin ANN was applied to subset B which was used as training subset and then to subset A which was used as testing subset.
6. Therefore, the results are relevant to two sequences of training testing protocol: A-B and B-A.

Results were drawn up in terms of sensitivity (correct classification of positive patients), specificity (correct classification of negative patients), global accuracy (arithmetic mean between sensitivity and specificity). Overall results are expressed as average of the two experiments.

This crossover procedure allows to classify blindly all records with the trained algorithm ensuring the generalization capability of the model on records never seen before.

The machine learning algorithms developed at the University of Waikato, New Zealand available on the WEKA data mining software are listed in Table 2 (18-24), while two ANNs (Self Momentum Back Propagation and Sine Net) (25, 26) were implemented in “Supervised ANNs Software”, developed at the Semeion Research Center in Rome, Italy (Buscema; Supervised ANNs. Semeion software #12, version 33.0). Table 3 shows main features of Semeion Machine Learning Systems.

Table 2. Nicknames of Machine Learning Systems in WEKA software Package

Machine Learning software	Nick name
Logistic(18)	Logistic
Pruned decision tree(19)	J 48
Multi Layer Perceptron(21)	MLP
NaivBayes(20)	NaivBayes
RandomForest(22)	RandomForest
RotationForest(23)	RotationForest
Sequential Minimal Optimization(24)	SMO
K Nearest Neighborhood(27)	KNN47
Additive Logistic Regression(28)	Logit Boost48

Table 3. Main Features of Semeion Machine Learning Systems employed in the study

ANNs Architecture	Hidden layers	Learning Rate	Epochs	Initialization	Output Function	Nick name
Conic Net(29)	5 Layers= 4x12x12x12	0.01	1000	Auto- encoders	Soft Max	D_FF_Conic (4x12x12x12)
	5 Layers= 6x12x12x12	0.01	1000	Auto- encoders	Soft Max	D_FF_Conic (6x12x12x12)
Sine Net(30-32)	1 Layer= 48	0.1	2000	Random	Soft Max	D_FF_Sn (48)
Back Propagation(25)	0 Layer= L	0.1	1000	Random	Soft Max	D_FF_Bp (0)
	1 Layer= 24	0.1	1000	Random	Soft Max	D_FF_Bp (24)
	5 Layers= 16x16x16x16	0.01	2000	Auto- encoders	Soft Max	D_FF_Bp (16x16x16x16)
Bi-Modal Net	1 Layer= 48	0.1	1000	Random	Soft Max	D_FF_Bm (48)
Gauss Net(33)	1 Layer= 64	0.01	1000	Random	Soft Max	D_FF_GNet (64)

However, since noisy input attributes can sometimes hide the small meaningful information embedded in other attributes, a pruning procedure was used as a pre-processing tool to eliminate noisy variables before the outcome prediction of the main test. In order to conduct this procedure, an input selection algorithm named TWIST (Training With Input Selection and Testing) was applied (34, 35). This selection algorithm was recently developed at the Semeion Research Center in Rome, Italy (Buscema M (2006-2012) TWIST Input Search, Semeion software #39, version 5.1, 2006-2016).

TWIST algorithm

As previously shown (36), the TWIST algorithm is a complex algorithm able to search for the best distribution of the global dataset divided in two optimally balanced subsets containing a minimum number of input features, useful for optimal pattern recognition. TWIST is an evolutionary algorithm based on a paper about Genetic Doping Systems (17), which has already been applied to medical data with very promising results (36-40).

A detailed description of the algorithm is available in the Online supplement.

5-K-fold

In addition, a 5 K-fold cross-validation protocol was applied(41),in order analyze data also with a standard and popular approach. The dataset was randomly split in 5 groups (folds) with a similar number of subjects.Each unique groupwas used as a hold out, or validation dataset, and the remaining groups were used astraining datasets.The model fitted on the training set was evaluated on the validation set. Five different models for each employed machine learning machine systems were created, and each model provided an evaluation score. The skill of the 5 models was summarized as mean sensitivity, specificity, overall accuracy and balance accuracy.

Results

Three hundred forty-seven patients fulfilled the inclusion criteria, 148 patients presented exclusion criteria (9 patients <12 years, 139 patients without leukocyte formula), leaving 199 patients for the analysis.

Population description and classic statistics

Table 4 summarizes the main characteristics of the overall study population (n=199) and of the two subgroups, i.e. patients who tested positive (n=124) and negative (n=75) to SARS-COV-2.

Table 4.Characteristics of the study population in three groups: overall population, positive for SARS-COV-2 and negative for SARS-COV-2.

Median age in the overall population was 65 [46-78] years, with similar distribution in the two subgroups (65 [49-77] vs. 66 [38-82] years p=.94, for positive and negative patients, respectively).

Most patients were male (63.8%) with no significant difference between the two subgroups (62.9% vs. 65.3%, p=.85).

The most common comorbidities were hypertension (42.7%) and diabetes (14.6%) with similar prevalence between patients with and without SARS-COV-2. A lower prevalence of Chronic obstructive pulmonary disease (COPD) was observed in patients with COVID-19 (4.0% vs. 13.3% p=.03).

Regarding current medications, positive patients had lower chronic use of anti-epileptics and drugs for psychiatric disorders, as compared to patients that tested negative (0.8% vs. 12.0%, p=.002 and 0.8% vs. 8.0%, p=.02, respectively). Furthermore, administration of antibiotics before access to the emergency department was higher in the positive subgroups (29.0% vs. 14.7%, p=.03).

As for in-hospital clinical presentation, fever and coughing were more prevalent in patients with COVID-19 (96.0% vs. 73.3%, p<.001 and 73.4% vs. 52.0%, p=.004, respectively).

Several clinical findings turned out to be significantly different between the two population studies (Table 3). In particular, SARS-COV-2 positive patients had a slightly higher external body temperature (37.8 ± 0.8 vs. 37.2 ± 1.0 ,

p=<.001), lower prevalence of pleural effusion at chest X-ray (5.7% vs. 17.3%, p=.01) and a significant difference regarding the complete blood count and leukocyte formula (Table 3).

Table 5 sums up the positive and negative linear correlation index between descriptive variables and the results of RT-PCR for SARS-COV-2.

Main results: Prediction of the PCR Outcome with Machine Learning Algorithms

The TWIST system selected 42 variables of the original attributes. Selected variables are marked with an asterisk in Table 5. A global dataset of 42 input and 2 target attributes was thus generated. Thereafter, two optimal subsets were created, in order to apply the training and testing procedure described above.

Table 6 shows the results obtained by the application of an array of machine learning systems to the variables selected by TWIST system. These results are the average of two training-testing procedures (A-B and B-A sequences). Detailed predictive results for each experiment (A-B and B-A) is available in Table S1 in the Online Supplement.

Table 6 Predictive results with variables selection using Semeion(*) and WEKA () Machine learning systems.**

Machine learning system	Sensitivity %	Specificity %	Overall accuracy 1%	Balance accuracy 2 %	Variance %	PPV (%) ³	AUROC ⁴
D_FF_Conic(4x12x12x12)*	94.1	88.7	91.4	92.2	0.5	93.5	0.90
D_FF_Conic(6x12x12x12)*	92.5	90.2	91.3	91.7	0.0	94.1	0.91
D_FF_Bp(24) *	89.2	93.0	91.1	90.6	1.0	95.6	0.93
D_FF_Bp(16x16x16x16)*	93.2	88.7	91.0	91.7	1.0	93.4	0.92
D_FF_GNet(64)*	90.7	90.2	90.5	90.6	1.0	94.0	0.90
D_FF_Sn(48)*	91.7	88.9	90.3	90.6	1.0	93.2	0.92
D_FF_Bm(48)*	91.6	88.9	90.2	90.6	1.0	93.2	0.91
D_FF_Conic(48)*	91.6	88.7	90.2	90.6	0.0	93.3	0.92
D_FF_Bp(0)*	89.1	90.3	89.7	89.6	2.1	93.8	0.91
MLP**	81.0	84.8	82.9	82.3	3.1	89.8	0.90
RandomForest**	86.6	65.3	75.9	78.7	1.6	80.6	0.86
NaiveBayes**	85.8	64.5	75.1	78.1	4.2	81.0	0.83
RotationForest**	88.3	60.8	74.6	78.1	0.0	79.1	0.85
Logistic**	80.7	67.9	74.3	76.0	1.0	80.8	0.63
LogitBoost**	81.0	61.2	71.1	73.4	1.6	77.6	0.81
J48**	77.4	60.7	69.0	71.4	0.5	77.0	0.57
SMO**	96.8	27.9	62.4	70.3	9.9	61.5	0.65
kNN**	75.9	36.5	56.2	60.9	2.6	66.6	0.56
<p><i>The results are the average of two testing experiments with training-testing A-B and B-A sequence. A hundred cases were presented in subset A and ninety-nine cases in subset B.</i></p> <p>¹Aritmetic mean of sensitivity and specificity</p> <p>² Weighted mean of sensitivity and specificity</p> <p>³PPV-Positive Predictive Value.</p> <p>⁴AUROC- Area Under the Receiver Operator Curve.</p>							

The best machine learning system reached an accuracy of 91.4% with 94.1% sensitivity (correct prediction of positivity to SARS-COV-2) and 88.7% specificity (correct prediction of negativity to SARS-COV-2).

Table 7 shows the results obtained through applying a selected number of machine learning systems to the variables selected by the TWIST system. These results are the average of five Training-Testing procedures of a K-

fold cross-validation protocol. Detailed predictive results for each experiment are available in Table S2 in the Online Supplement.

Table 7 Predictive results with 5-K fold protocol,using Semeion(*) and WEKA () machine learning systems.**

Machine learning system	Sensitivity %	Specificity %	Overall accuracy ¹ %	Balance accuracy ² %	Variance %	PPV (%) ³	AUROC ⁴
D_FF_Conic(4x12x12x12)*	87.5	84.6	86.0	86.5	3.4	91.0	0.84
D_FF_Conic(6x12x12x12)*	89.2	86.2	87.7	88.0	2.6	92.1	0.86
D_FF_Bp(24) *	88.3	77.6	83.0	84.3	6.0	87.1	0.81
D_FF_Sn(48)*	88.3	83.2	85.8	86.4	4.3	90.4	0.85
RandomForest**	90.0	58.0	74.0	78.1	3.2	78.9	0.83
Logistic**	82.5	62.5	72.5	75.1	5.6	79.7	0.74

The results are the average of five testing experiments.

¹*Aritmetic mean of sensitivity and specificity*

²*Weighted mean of sensitivity and specificity*

³*PPV-Positive Predictive Value.*

⁴*AUROC- Area Under the Receiver Operator Curve.*

Discussion

We studied 199 adult patients admitted to the emergency department of the largest hospital in Milan, Lombardy, with symptoms compatible with COVID-19 during the first 3 weeks of SARS-COV-2 outbreak in Italy. In the present manuscript, we describe this population and highlight the differences between patients who actually tested positive to SARS-COV-2 and those who did not. Few attempts in applying artificial intelligence to rapidly predict positivity/negativity to SARS-COV-2 were made since the outbreak, using mostly CT imaging and lab results, collected in Chinese population(42-44). Nevertheless, we present the first European attempt and promising results applying artificial intelligence to rapidly predict positivity/negativity to SARS-COV-2 using only basic clinical data, available in the vast majority of emergency departments all over the world. The wide application of this decision support tool could have a major clinical and organizational impact during the current pandemic.

In our study, several differences were observed between the two study groups (Table 3). However, none of these, or a combination of them, allows, so far, to clearly differentiate between patients with COVID-19 and patients with other diseases, having a similar clinical presentation. Our data underline the key finding of Coronavirus-induced

alterations in the white blood cell differential count (Table 3). On the one hand, in contrast to other reports (45-48) we did not observe a marked lymphocytopenia, possibly because of the early stage of the viral disease. On the other, other subtypes such as eosinophils might play a key role in COVID-19 (49).

When applying artificial intelligence to our dataset, in particular ANNs and MLS, we were able to predict with high sensitivity and specificity the results of RT-PCR (Table 6).

Artificial Neural Networks allow forecasting through understanding of the relationship between variables, in particular through the application of nonlinear relationships (36, 50, 51). These systems initially learn from a set of data with a known solution (training). Thereafter, the networks, inspired by the analytical processes of the human brain, are able to reconstruct imprecise rules, which may be underlying a complex dataset (testing). Machine learning systems and, in particular, ANNs analyse real-world data very efficiently. The internal validity of their assessment is provided by uniquely strict validation protocols, seldom used in classical statistics (50, 52, 53).

In the present manuscript, it was possible to predict with reasonable accuracy the status of being positive or negative to SARS-COV-2 based on 42 simple variables. This was achieved using the TWIST algorithm, which does not have, at the moment, the same popularity of other techniques, such as K Fold, Boosting and others. Nevertheless, it has been used extensively in the past 15 years in different context(54-56). The reason of its low diffusion is partly that TWIST is very complex to program, as it includes two evolutionary algorithms that work together managing a huge population of ANNs, kNN and Naive Bayes algorithms. The execution of TWIST needs therefore to be programmed in C language to be sufficiently fast. Thus, for its complexity and for needed running time, TWIST is not suitable for programming in Python, R or similar languages.

TWIST system allowed reaching a global accuracy of 91.4% with the best machine learning system: 94.1% sensitivity (correct prediction of positivity to SARS-COV-2) and 88.7% specificity (correct prediction of negativity to SARS-COV-2). Considering the eight best machine-learning systems their average performance was the following: sensitivity of 91.8%, specificity of 89.6% and global accuracy of 90.8%.

Comparing the two testing procedures (A-B and B-A), explained in the mathematical section, the differences in predicting values between these two experiment is small, therefore reasonably excluding overfitting of the model(57).

In order to analyze our dataset also with more popular and widely applied procedure, we applied a 5 k-fold cross-validation protocol, using a selected number of machine learning systems (Table 7). With this type of analysis, the best machine learning system obtained an overall accuracy of 87.7% with a sensitivity and specificity of 89.2% and 86.2%, respectively. Global average performance was the following: sensitivity of 87.6%, specificity of 75.4% and a global accuracy of 81.5%.

Comparing these results with those obtained by the same machine learning systems, using the AB -BA Train-Testing protocol (shown in Table 6), the latter allows to obtain a slightly better predictive results, reasonably related to the optimal splitting of the records, with an average performance of 89.1% specificity, 82.2% specificity and 85.7% global accuracy. The high variance of results obtained with the K Fold protocol and the low variance of the same results using TWIST protocol is suggestive of the high polarization affecting the K Fold protocol with this kind of data.

This is the reason why we have chosen to rely on an optimized distribution of records in training and testing subsets, rather than on a random allocation. Nevertheless, also the application of a standard K fold cross-validation, *i.e.* a system widely available, was able to predict accurately the results of SARS-COV-2 RT-PCR.

It is useful to analyse variables selected by AI, as they certainly bear specific clinical information. As mentioned above, the white blood cells and their differential count are certainly very informative.

Indeed, total white blood cell count ($R = -0.46$), lymphocytes ($R = 0.26$) and eosinophils ($R = -0.34$) correlated either negatively or positively with COVID-19 and were included in the model.

In addition, the final model included also variables with very low correlation with RT-PCR results, such as dyspnea ($R = -0.07$), basophils ($R = -0.06$), mean cell haemoglobin ($R = -0.06$), non-invasive ventilation ($R = -0.05$), monocytes ($R = 0.05$), age ($R = 0.04$), female sex ($R = 0.02$) and headache ($R = 0.02$). The fact that these variables have been included in the model confirms the ability of ANN to handle highly nonlinear functions.

Other authors have applied AI for the diagnosis of SARS-COV-2. Rao et al. employed an AI framework to a mobile phone-based survey, exclusively based on pre-hospital clinical symptoms and demographic characteristics to assess the probability of SARS-COV-2 infection (58). Three different research groups tried to predict positivity to SARS-COV-2 using, among other variables, CT scans (14, 42, 59). Chest CT scan was analysed via deep learning by Li et al. to differentiate SARS-COV-2 induced viral pneumonia from other lung disease (60). Two other research groups developed from machine learning models free online applications, using only lab test results (43, 44).

Our model significantly differs from the abovementioned. First, it relies on basic clinical information, available in almost every emergency department. The required information is quickly obtainable for every patient at hospital admission. For this reason, we decided to include chest X-Ray rather than CT in our model. Indeed, despite CT being certainly more sensitive in identifying alterations typical of viral pneumonia (13), not every SARS-COV-2 suspect will have access to a CT scan. Second, our study is the first one analysing data from an European country. While there is no evidence so far, it is possible that different ethnicities will show slightly different responses to viral invasion.

This study has certainly limitations. First, it is a retrospective, single-center study. Second, the sample size is limited. Third, some fundamental clinical data, such as arterial blood gas analysis were not available for all patients and this information was thus not included in the model. Given the typical profound hypoxemia of patients with COVID-19, it is conceivable that adding these variables to the system could further improve its accuracy. Finally, while no definitive data are available regarding the accuracy of RT-PCR testing for SARS-COV-2 (61), several studies have described a certain percentage of false negative results (5-7). However, every negative result was re-tested after 48 hours. This methodological aspect should reduce the risk of falsely negative results.

Possible clinical and organizational implications

Facing a highly contagious viral outbreak requires a complex effort in terms of political, economic, social and health systems re-organization (62, 63).

A fundamental aspect is to define a clear management protocol, in order to separate infected from non-infected patients, *i.e.*, those admitted for other clinical conditions. Indeed, it is of paramount importance to set up clearly separate pathways, in order to avoid the spread of viral infections within the hospital. A quick and reliable system to identify SARS-COV-2 infected patients is therefore fundamental. Currently, the gold standard for the diagnosis is a RT-PCR assay searching for SARS-COV-2 genome (9). This type of molecular assay has certainly several limitations. During the first month of outbreak in Italy, the processing of samples became more efficient, theoretically reducing the technical time needed for the result. Despite this, the problem of delayed diagnosis still exists. This is due to the availability RT-PCR machines, considering the high demand during the outbreak. Furthermore, laboratories of referral hospitals, such as ours, analyse samples also arriving from smaller hospitals not equipped for SARS-COV-2 testing. Finally, of course, RT-PCR is not a perfect test, and false negative result have been described, even in the presence of strong clinical suspicion for COVID-19 (5-7).

For these reasons, applying AI as a rapid decision support tool for the diagnosis of COVID-19 and therefore to speed up the sorting of infected from non-infected patients would be of great clinical help. Indeed, a simple online software, fed with basic clinical data, easily obtainable in almost every emergency department, could apply trained ANNs to predict with high accuracy the RT-PCR result. The results obtained from this software should of course be integrated with available clinical data.

The application of AI to clinical practice is still limited for its complexity and for limited in-hospital availability of technical infrastructures and support. This, of course, could be particular troublesome for small centres with limited resources. The decision support software that could integrate the information contained in the present manuscript could ideally retrieve data directly from the electronic patient management system. Otherwise, data could be manually entered in an online software, which however significantly increases the risk of errors. An active support from and collaboration with the local information technology infrastructure is therefore fundamental in order to be able, in the future, to integrate AI into clinical practice.

Finally, it is conceivable that the information obtained from the present study might be useful also at the end of the current pandemic. Indeed, it is likely that SARS-COV-2 might become a seasonal virus. In this regard, the early identification would be a key factor to reduce the risk of a further epidemic outbreak.

In summary, our study suggests that basic clinical data might be sufficient for properly trained ANNs and MLS algorithms to predict with good accuracy the positivity and negativity to SARS-COV-2. If confirmed, this could have important clinical and organizational implications. Indeed, while not directly changing the treatment of COVID-19 patients, it could reduce the time patients spend unnecessarily in the emergency department, could reduce the workload of intensive care staff and, finally, reduce the risk of collapsing healthcare systems.

List Of Abbreviations

SARS-COV-2: Severe Acute Respiratory Syndrome Coronavirus 2

COVID-19: Coronavirus Disease 2019

RT-PCR: Reverse Transcription- Polymerase Chain Reaction

CT : Computed Tomography

ANNs: Artificial Neural Networks

MLS: Machine Learning Systems

TWIST: Training With Input Selection and Testing

COPD: Chronic obstructive pulmonary disease

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of our hospital (N°3733). Mandatory informed consent was waived, considering the complete data anonymization.

Consent for publication

Mandatory informed consent was waived.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

No funding.

Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by TL, MF, RG, FV, HG, EG, AM, AZ, SB, MB. The first draft of the manuscript was written by TL, MF, EG, FV, RG, AM and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to acknowledge Dr. Arturo Chiericato (Department of Anaesthesia and Intensive Care Medicine, Niguarda Ca' Granda, Milan, Italy), Dr. Massimo Puoti (Department of Infectious Disease, Niguarda Ca' Granda, Milan, Italy) and the following collaborators for their help and important contribution in data analysis: Giulia Massini and Francesca della Torre, Semeion Research Center of Sciences of Communication, Rome, Italy. Finally, the authors wish to thank Mrs. Convery Geraldine for proofreading this manuscript.

References

1. WHO. Pneumonia of unknown cause – China 2020 [Available from: <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/>].
2. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli 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.
3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-33.
4. Mission W-CJ. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) 2020 [Available from: [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19))].
5. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*. 2020.
6. Guo L, Ren L, Yang S, Xiao M, Chang, Yang F, et al. Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clin Infect Dis*. 2020.
7. Lippi G, Simundic AM, Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). *Clin Chem Lab Med*. 2020.
8. Li LQ, Huang T, Wang YQ, Wang ZP, Liang Y, Huang TB, et al. 2019 novel coronavirus patients' clinical characteristics, discharge rate and fatality rate of meta-analysis. *J Med Virol*. 2020.
9. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*. 2020.
10. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507-13.
11. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med*. 2003;348(20):1967-76.
12. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. 2019;17(3):181-92.

13. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020:200642.
14. Mei X, Lee H-C, Diao K-y, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nature Medicine*. 2020.
15. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA data mining software: An update. *SIGKDD Explor Newsl*. 2008;11:10-8.
16. Buscema M, Grossi E, Intraligi M, Garbagna N, Andriulli A, Breda M. An optimized experimental protocol based on neuro-evolutionary algorithms application to the classification of dyspeptic patients and to the prediction of the effectiveness of their treatment. *Artif Intell Med*. 2005;34(3):279-305.
17. Buscema M. Genetic doping algorithm (GenD): theory and applications. *Expert Systems*. 2004;21(2):63-79.
18. W. HD, Stanley L. *Applied Logistic Regression* 2000.
19. J. R, Quinlan. *C4.5: Programs for Machine Learning*: Morgan Kaufmann Publishers Inc.; 1993.
20. George H. John, Langley P. *Estimating Continuous Distributions in Bayesian Classifiers*. 2013.
21. Collobert R, Bengio S. Links between Perceptrons, MLPs and SVMs. *Icml '04*. 2004:23.
22. F L. *Implementing Breiman's Random Forest Algorithm into Weka* 2005.
23. Rodriguez JJ, Kuncheva LI, Alonso CJ. Rotation forest: A new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell*. 2006;28(10):1619-30.
24. Keerthi SS, Gilbert EG. Convergence of a Generalized SMO Algorithm for SVM Classifier Design. 2002;46(1):351-60.
25. Buscema M. Back propagation neural networks. *Subst Use Misuse*. 1998;33(2):233-70.
26. Buscema M, Terzi S, Breda M. Using sinusoidal modulated weights improve feed-forward neural network performances in classification and functional approximation problems. 2006;3:885-93.
27. Wang J, Zucker J-d. Solving the multiple-instance problem: A lazy learning approach 2000. 1119-26 p.
28. Friedman J, Hastie T, Tibshirani R. Additive Logistic Regression: A Statistical View of Boosting. 2000;28:337-407.
29. Buscema PM, Massini G, Fabrizi M, Breda M, Della Torre F. The ANNS approach to DEM reconstruction. *Computational Intelligence*. 2018;34(1):310-44.
30. Buscema M, Terzi S, Breda M. Improve Feed-Forward Neural Network Performances in Classification and Functional Approximation Problems. *WSEAS Transactions on Information Science & Applications*. 2006;3(5):885-93.
31. Buscema M, inventor Sine Net : an artificial neural network 2003.
32. Buscema M, Terzi S, Breda M, editors. A feed Forward sine based neural network for functional approximation of a waste incinerator emissions. 8th WSEAS Int Conference on Automatic Control, Modeling and Simulation 2006 March 12 th -14 th, 2006.; Praga.
33. Buscema PM. Gauss Net Equations. *Semeion Archives* 2015.
34. Buscema M, Breda M, Lodwick W. Training with Input Selection and Testing (TWIST) Algorithm: A Significant Advance in Pattern Recognition Performance of Machine Learning. *Journal of Intelligent Learning Systems and Applications*. 2013;5:29-38.
35. Pace F, Riegler G, de Leone A, Pace M, Cestari R, Dominici P, et al. Is it possible to clinically differentiate erosive from nonerosive reflux disease patients? A study using an artificial neural networks-assisted algorithm. *Eur J*

- Gastroenterol Hepatol. 2010;22(10):1163-8.
36. Coppede F, Grossi E, Migheli F, Migliore L. Polymorphisms in folate-metabolizing genes, chromosome damage, and risk of Down syndrome in Italian women: identification of key factors using artificial neural networks. *BMC Med Genomics*. 2010;3:42.
 37. Lahner E, Intraligi M, Buscema M, Centanni M, Vannella L, Grossi E, et al. Artificial neural networks in the recognition of the presence of thyroid disease in patients with atrophic body gastritis. *World J Gastroenterol*. 2008;14(4):563-8.
 38. Buri L, Hassan C, Bersani G, Anti M, Bianco MA, Cipolletta L, et al. Appropriateness guidelines and predictive rules to select patients for upper endoscopy: a nationwide multicenter study. *Am J Gastroenterol*. 2010;105(6):1327-37.
 39. Street ME, Grossi E, Volta C, Faleschini E, Bernasconi S. Placental determinants of fetal growth: identification of key factors in the insulin-like growth factor and cytokine systems using artificial neural networks. *BMC Pediatr*. 2008;8:24.
 40. Buscema M, Grossi E, Capriotti M, Babiloni C, Rossini P. The I.F.A.S.T. model allows the prediction of conversion to Alzheimer disease in patients with mild cognitive impairment with high degree of accuracy. *Curr Alzheimer Res*. 2010;7(2):173-87.
 41. Little M, Varoquaux G, Saeb S, Lonini L, Jayaraman A, Mohr D, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. *GigaScience*. 2017;6.
 42. Feng C, Huang Z, Wang L, Chen X, Zhai Y, Zhu F, et al. A Novel Triage Tool of Artificial Intelligence Assisted Diagnosis Aid System for Suspected COVID-19 pneumonia In Fever Clinics. *medRxiv*. 2020:2020.03.19.20039099.
 43. Wu J, Zhang P, Zhang L, Meng W, Li J, Tong C, et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. *medRxiv*. 2020:2020.04.02.20051136.
 44. Meng Z, Wang M, Song H, Guo S, Zhou Y, Li W, et al. Development and utilization of an intelligent application for aiding COVID-19 diagnosis. *medRxiv*. 2020:2020.03.18.20035816.
 45. Li YX, Wu W, Yang T, Zhou W, Fu YM, Feng QM, et al. [Characteristics of peripheral blood leukocyte differential counts in patients with COVID-19]. *Zhonghua nei ke za zhi*. 2020;59(0):E003.
 46. Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*. 2020.
 47. Wan S, Xiang Y, Fang W, Zheng Y, Li B, Hu Y, et al. Clinical Features and Treatment of COVID-19 Patients in Northeast Chongqing. *Journal of Medical Virology*.n/a(n/a).
 48. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama*. 2020.
 49. Liu F, Xu A, Zhang Y, Xuan W, Yan T, Pan K, et al. Patients of COVID-19 may benefit from sustained lopinavir-combined regimen and the increase of eosinophil may predict the outcome of COVID-19 progression. *Int J Infect Dis*. 2020.
 50. Vomweg TW, Buscema M, Kauczor HU, Teifke A, Intraligi M, Terzi S, et al. Improved artificial neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Med Phys*. 2003;30(9):2350-9.
 51. Penco S, Grossi E, Cheng S, Intraligi M, Maurelli G, Patrosso MC, et al. Assessment of the role of genetic polymorphism in venous thrombosis through artificial neural networks. *Ann Hum Genet*. 2005;69(Pt 6):693-706.

52. Andriulli A, Grossi E, Buscema M, Festa V, Intraligi NM, Dominici P, et al. Contribution of artificial neural networks to the classification and treatment of patients with uninvestigated dyspepsia. *Dig Liver Dis.* 2003;35(4):222-31.
53. Mecocci P, Grossi E, Buscema M, Intraligi M, Savare R, Rinaldi P, et al. Use of artificial networks in clinical trials: a pilot study to predict responsiveness to donepezil in Alzheimer's disease. *J Am Geriatr Soc.* 2002;50(11):1857-60.
54. Cosmi V, Mazzocchi A, Milani GP, Calderini E, Scaglioni S, Bettocchi S, et al. Prediction of Resting Energy Expenditure in Children: May Artificial Neural Networks Improve Our Accuracy? *Journal of clinical medicine.* 2020;9(4).
55. Buscema M, Grossi E, Intraligi M, Garbagna N, Andriulli A, Breda M. An Optimized Experimental Protocol Based on Neuro-Evolutionary Algorithms. Application to the Classification of Dyspeptic Patients and to the Prediction of the Effectiveness of Their Treatment. *Artificial Intelligence in Medicine.* 2005;34:279-305.
56. Podda GM, Grossi E, Palmerini T, Buscema M, Femia EA, Della Riva D, et al. Prediction of high on-treatment platelet reactivity in clopidogrel-treated patients with acute coronary syndromes. *International journal of cardiology.* 2017;240:60-5.
57. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed).* 2015;350:g7594.
58. Rao A, Vazquez JA. Identification of COVID-19 Can be Quicker through Artificial Intelligence framework using a Mobile Phone-Based Survey in the Populations when Cities/Towns Are Under Quarantine. *Infect Control Hosp Epidemiol.* 2020:1-18.
59. Xiong Z, Fu L, Zhou H, Liu JK, Wang AM, Huang Y, et al. [Construction and evaluation of a novel diagnosis process for 2019-Corona Virus Disease]. *Zhonghua Yi Xue Za Zhi.* 2020;100(0):E019.
60. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. *Radiology.* 2020:200905.
61. WHO. Coronavirus disease (COVID-19) technical guidance: Laboratory testing for 2019-nCoV in humans 2020 [Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance>].
62. Grasselli G, Pesenti A, Cecconi M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. *Jama.* 2020.
63. Spina S, Marrazzo F, Migliari M, Stucchi R, Sforza A, Fumagalli R. The response of Milan's Emergency Medical System to the COVID-19 outbreak in Italy. *The Lancet.* 2020;395(10227):e49-e50.

Tables

Table 1.
Variables under study (n=74)

Demographic Data
Sex
Age
Comorbidities
Hypertension
Human immunodeficiency virus
Ischemic cardiomyopathy
Chronic liver disease
Atrial fibrillation
Autoimmune/inflammatory disease
Diabetes mellitus
Chronic obstructive pulmonary disease (COPD)
Chronic kidney disease
Asthma
Solid organ tumor
Chronic Interstitial lung disease
Hematologic malignancy
Smoking (active)
Current Medications
Beta blocker
Anticoagulant therapy
Calcium channel antagonist
Anti-platelets therapy
Sartans
Anti-epileptic therapy
Diuretics
Drugs for psychiatric disorders
Angiotensin converting enzyme inhibitor
Immunosuppressant drugs
Hydroxymethylglutaryl-CoA reductase inhibitors

5-alpha-reductase inhibitors
Anti-arrhythmic drugs
Antibiotics before Emergency Department
Symptoms
Fever
Sore throat
Cough
Chest pain
Headache
Gastrointestinal symptoms
Asthenia
Dyspnea
Arthralgia
Syncope
Clinical findings
Glasgow coma scale (GCS)
Basophils percentage
Sinus rhythm
Total red blood cell count (10 ⁶ /μL)
Normal ST segment
Hemoglobin (g/dL)
Heart rate (beats per minute)
Hematocrit (%)
Respiratory rate (breaths/min)
Mean red blood cell volume (μm ³)
Systolic blood pressure (mmHg)
Mean corpuscular hemoglobin (pg)
Diastolic blood pressure (mmHg)
Mean corpuscular hemoglobin concentration (g/dL)
Body temperature (degree Celsius)
Red blood cell distribution width (%)
Oxygen supplementation (nasal cannula, mask)

Platelets (103/ μ L)
Non-Invasive Ventilation
Creatinine (mg/dL)
Chest X-rays opacity
Glycaemia (mg/dL)
Chest X-ray pleural effusion
Urea (mg/dL)
Total white blood cell count (103/ μ L)
Calcium (mg/dL)
Neutrophils percentage
Sodium (mEq/L)
Lymphocytes percentage
Total bilirubin (mg/dL)
Monocytes percentage
C-reactive protein (mg/dL)
Eosinophils percentage
Swab test for SARS-COV-2 (TARGET)
Positive
Negative

Table 4.

Characteristics of the study population in three groups: overall population, positive and negative for SARS-COV-2.

Demographic characteristics	Overall population (n=199)	Positive (n=124)	Negative (n=75)	p value
Male ♂ no. (%)	127 (63.8)	78 (62.9)	49 (65.3)	0.85
Age, years	65 [46-78]	65 [49-77]	66 [38-82]	0.94
Comorbidities (no. (%))				
Hypertension	85 (42.7)	55 (44.4)	30 (40)	0.65
Ischemic cardiomyopathy	19 (9.6)	10 (8.1)	9 (12)	0.51
Atrial fibrillation	19 (9.6)	11 (8.9)	8 (10.7)	0.87
Diabetes mellitus	29 (14.6)	16 (12.9)	13 (17.3)	0.51
Chronic kidney disease	11 (5.5)	8 (6.5)	3 (4)	0.68
Solid organ tumour	16 (8)	7 (5.7)	9 (12)	0.18
Hematologic malignancy	6 (3)	4 (3.2)	2 (2.7)	0.84
Human immunodeficiency virus	1 (0.5)	0 (0)	1 (1.3)	0.80
Chronic liver disease	7 (3.5)	3 (2.4)	4 (5.3)	0.49
Autoimmune/inflammatory disease	6 (3)	3 (2.4)	3 (4)	0.84
Chronic obstructive pulmonary disease (COPD)	15 (7.5)	5 (4)	10 (13.3)	0.03
Asthma	13 (6.6)	9 (7.3)	4 (5.4)	0.83
Chronic Interstitial lung disease	1 (0.5)	1 (0.8)	0 (0)	0.80
Smoking (active)	12 (6)	4 (3.2)	8 (10.7)	0.07
Current medication (no. (%))				
Beta blockers	40 (20.1)	24 (19.4)	16 (21.3)	0.88
Calcium channel antagonists	27 (13.6)	18 (14.5)	9 (12)	0.77
Sartans	24 (12.1)	19 (15.3)	5 (6.7)	0.11
Diuretics	35 (17.6)	21 (16.9)	14 (18.7)	0.91
Angiotensin converting enzyme inhibitors	33 (16.6)	21 (16.9)	12 (16)	0.98
Hydroxymethylglutaryl-CoA (HMG-CoA) reductase inhibitors	28 (14.1)	18 (14.5)	10 (13.3)	0.98
Anti-arrhythmic drugs	11 (5.5)	6 (4.8)	5 (6.7)	0.82
Anticoagulant drugs	25 (12.6)	15 (12.1)	10 (13.3)	0.97
Antiplatelets drugs	33 (16.6)	20 (16.1)	13 (17.3)	0.98

Anti-epileptic drugs	10 (5.0)	1 (0.8)	9 (12)	0.002
Drugs for psychiatric disorders	7 (3.5)	1 (0.8)	6 (8)	0.02
Immunosuppressant drugs	12 (6)	8 (6.5)	4 (5.3)	0.99
5-alpha-reductase inhibitors	12 (6)	9 (7.3)	3 (4)	0.53
Antibiotics before Emergency Department	47 (23.6)	36 (29)	11 (14.7)	0.03
Symptoms/no. (%)				
Fever	174 (87.4)	119 (96)	55 (73.3)	<0.001
Cough	130 (65.3)	91 (73.4)	39 (52)	0.004
Headache	12 (6.0)	8 (6.5)	4 (5.3)	0.99
Asthenia	18 (9.1)	9 (7.3)	9 (12)	0.38
Arthralgia	7 (3.5)	6 (4.8)	1 (1.3)	0.37
Sore throat	8 (4)	3 (2.4)	5 (6.7)	0.27
Gastrointestinal symptoms	27 (13.6)	18 (14.5)	9 (12)	0.77
Chest pain	7 (3.5)	7 (5.7)	0 (0)	0.09
Dyspnea	75 (37.7)	43 (34.7)	32 (42.7)	0.33
Syncope	5 (2.5)	2 (1.6)	3 (4)	0.57
Clinical findings				
Glasgow coma scale, score \bar{x} median [IQR]	15 [15-15]	15 [15-15]	15 [15-15]	>0.99
Sinus rhythm \bar{x} no. (%)	185 (93)	117 (94.4)	68 (90.7)	0.48
Normal ST segment \bar{x} no. (%)	198 (99.5)	123 (99.2)	75 (100)	0.80
Heart rate, beats per min \bar{x} median [IQR]	90 [80-105]	90 [83- 105]	90 [80-105]	0.87
Respiratory rate, breaths per min \bar{x} median [IQR]	18 [16-22]	18 [16-22]	18 [16-24]	0.80
Systolic blood pressure, mmHg \bar{x} mean \pm SD	131 \pm 22	133 \pm 19	128 \pm 25	0.09
Diastolic blood pressure, mmHg \bar{x} median [IQR]	75 [65-80]	75 [70-80]	70 [65-80]	0.09
Body temperature, Celsius \bar{x} mean \pm SD	37.6 \pm 0.9	37.8 \pm 0.8	37.2 \pm 1	<0.001
Oxygen supplementation (nasal cannula. mask) \bar{x} no. (%)	51 (25.6)	27 (21.8)	24 (32)	0.15
Non-Invasive Ventilation \bar{x} no. (%)	8 (4)	4 (3.2)	4 (5.3)	0.72
Chest X-rays opacity \bar{x} no. (%)	158 (79.4)	104 (83.9)	54 (72)	0.07
Chest X-rays pleural effusion \bar{x} no. (%)	20 (10.1)	7 (5.7)	13 (17.3)	0.01
Total white blood cell count, $10^3/\mu\text{L}$ \bar{x} median [IQR]	6.64 [4.65-9.65]	5.44 [4.21-7.23]	9.28 [6.87-13.64]	<0.001

Neutrophils, % ☒ median [IQR]	72.5 [64.4-81.7]	70.4 [62.5-79.9]	76.7 [68.4-85.3]	0.001
Lymphocytes, % ☒ median [IQR]	18.0 [10.5-25.1]	20.6 [12.9-27.7]	14.5 [7.4-19.8]	<0.001
Monocytes, % ☒ median [IQR]	7.9 [5.6-10.2]	8.0 [6-10.9]	7.6 [5.4-9.7]	0.18
Eosinophils, % ☒ median [IQR]	0 [0-0.3]	0 [0-0.2]	0.2 [0-1.2]	<0.001
Basophils, % ☒ median [IQR]	0.3 [0.2-0.4]	0.2 [0.2-0.4]	0.3 [0.2-0.5]	0.03
Total red blood cells count, 10 ⁶ /μL ☒ median [IQR]	4.80 [4.29-5.25]	4.88 [4.42-5.28]	4.52 [4.01-5.14]	0.01
Haemoglobin, g/dL ☒ mean ± SD	13.7 ± 1.9	14.0 ± 1.6	13.3 ± 2.2	0.01
Haematocrit, % ☒ mean ± SD	41.6 ± 5.3	42.1 ± 4.5	40.8 ± 6.3	0.10
Mean red blood cell volume, μm ³ ☒ median [IQR]	88.3 [85.4-91.4]	87.7 [84.5-90.4]	89.7 [87-92.6]	0.002
Mean corpuscular haemoglobin, pg ☒ median [IQR]	29.2 [28.1-30.5]	29.2 [27.9-30.5]	29.3 [28.6-30.2]	0.56
Mean corpuscular haemoglobin concentration, g/dL ☒ median [IQR]	33.1 [32-33.9]	33.2 [32.5-34.1]	32.8 [31.6-33.8]	0.01
Red blood cells distribution width,%☒ median [IQR]	13.2 [12.4-14.4]	13.1 [12.3-13.9]	13.6 [12.5-15.5]	0.01
Platelets, 10 ³ /μL ☒ median [IQR]	193 [162-244]	183 [145-234]	221 [175-283]	<0.001
Creatinine, mg/dL ☒ median [IQR]	1.0 [0.8-1.3]	0.98 [0.82-1.21]	1.09 [0.83-1.45]	0.19
Glicemia, mg/dL ☒ median [IQR]	120 [104-142]	120 [104-138]	117 [101-158]	0.80
Urea, mg/dL ☒ median [IQR]	35 [23-54.5]	34 [22-49]	41.5 [26-64.5]	0.07
Calcium, mg/dL☒ median [IQR]	9 [8.7-9.4]	8.9 [8.6-9.2]	9.3 [8.8-9.8]	<0.001
Sodium, mEq/L ☒ median [IQR]	138 [135-140]	138 [134-140]	138 [136-141]	0.37
Total bilirubin, mg/dL ☒ median [IQR]	0.5 [0.4-0.7]	0.5 [0.3-0.6]	0.5 [0.4-0.7]	0.05
C-reactive protein, mg/dL ☒ median [IQR]	4.1 [1.2-9]	4 [1.3-8.9]	4.5 [0.9-9.3]	0.90

Variables are expressed as median [interquartile range], mean ± Standard Deviation or as number (percentages). Clinical findings refer to data gathered in the Emergency Room.

Table 5 Study variables positively and negatively correlated with SARS-COV-2 positivity.

Positive linear correlation			
VARIABLES	Pearson R	VARIABLES	Pearson R
*Fever	0.33	*Sinus Rhythm	0.06
*Body temperature (degree Celsius)	0.27	*Chronic kidney disease	0.05
Lymphocytes percentage	0.26	*Monocytes percentage	0.05
Cough	0.22	Hypertension	0.04
*Total red blood cells count ($10^6/\mu\text{L}$)	0.19	Asthma	0.04
*Haemoglobin (g/dL)	0.17	*Age	0.04
Glasgow coma scale (GCS)	0.17	Gastrointestinal symptoms	0.04
*Antibiotics before Emergency Department	0.16	*Calcium channel antagonist	0.04
*Chest pain	0.15	Mean corpuscular haemoglobin concentration (%)	0.03
*Chest X-rays opacity	0.14	Respiratory rate (breaths/min)	0.03
*Sartans	0.13	*Female	0.02
*Systolic blood pressure(mmHg)	0.12	*Headache	0.02
Haematocrit (%)	0.12	*Immunosuppressant drugs	0.02
Arthralgia	0.09	*Hydroxymethylglutaryl-CoA (HMG-CoA) reductase inhibitors	0.02
*Diastolic blood pressure (mmHg)	0.07	Hematologic malignancy	0.02
P wave present and normal	0.07	*Angiotensin converting enzyme inhibitor	0.01
*5-alpha-reductase inhibitors	0.07	Heart rate (beats per minute)	0.00
Chronic Interstitial lung disease	0.06		
Negative linear correlation			
VARIABLES	Pearson R	VARIABLES	Pearson R
*Total white blood cells count ($10^3/\mu\text{L}$)	-0.46	*Dyspnoea	-0.07
*Eosinophils percentages	-0.34	*Ischemic cardiomyopathy	-0.06
Platelets ($10^3/\mu\text{L}$)	-0.33	*Basophils percentage	-0.06
*Anti-epileptic therapy	-0.25	Diabetes mellitus	-0.06
*Calcium (mg/dL)	-0.22	Creatinine (mg/dL)	-0.06
Mean red blood cell volume (μm^3)	-0.21	*Mean corpuscular haemoglobin (pg)	-0.06

Neutrophils percentages	-0.19	*Non Invasive Ventilation	-0.05
*Drugs for psychiatric disorders	-0.19	Autoimmune/inflammatory disease	-0.04
*Chest X-rays pleural effusion	-0.19	Red blood cells distribution width (%)	-0.04
*Chronic obstructive pulmonary disease	-0.17	Anti-arrhythmic drugs	-0.04
*Smoking (active)	-0.15	Syncope	-0.04
*Urea (mg/dL)	-0.14	Glycaemia (mg/dL)	-0.03
*Total bilirubin (mg/dL)	-0.14	*Diuretics	-0.03
Oxygen supplementation	-0.11	Atrial fibrillation	-0.03
*Solid organ tumour	-0.11	C-reactive protein (mg/dL)	-0.03
*Sore throat	-0.10	*Male	-0.02
Human Immunodeficiency Virus	-0.09	Beta blocker	-0.02
Asthenia	-0.08	Anticoagulant therapy	-0.02
Chronic liver disease	-0.08	*Anti-platelets therapy	-0.02
*Sodium (mEq/L)	-0.07		
<i>Asterisks indicate variables included in the TWIST model.</i>			

Supplementary Files

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

- [OnlineSupplement.docx](#)