Sistema Socio Sanitario
Ospedale Niguarda
Regione Lombardia

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.

M. Favarato[1], H. Gay[1], R. Garberi[1], F. Villa[1], A. Zeduri[1], S. Bragagnolo[1], A. Molteni[3], M. Buscema[4], E. Grossi[5], R. Giudici[2], G. Bassi[2], T. Langer[1,2], R. Fumagalli[1,2]

Affiliations

[1] Department of Medicine and Surgery, University of Milan-Bicocca, Monza, Italy;

[2]Department of Anaesthesia and Intensive Care Medicine, Niguarda Ca' Granda, Milan, Italy;

[3] Department of General oncologic and mini-invasive Surgery, Niguarda Ca' Granda, Milan, Italy;

[4]Semeion Research Center of Sciences of Communication, Rome, Italy;

[5]Villa Santa Maria Foundation, Tavernerio, Italy

Background

RT-PCR for SARS-COV-2 diagnosis currently requires quite a long time span. A ***quicker and more efficient diagnostic*** tool in emergency departments (ED) could improve management during this global crisis.

Our main goal was assessing the accuracy of ***artificial intelligence (AI)*** in ***predicting the results of RT-PCR for SARS-COV-2***, using basic information at hand in all ED.

Methods

This is a ***retrospective study*** carried out between February 22, 2020 and March 16, 2020 in one of the main hospitals in Milan, Italy.

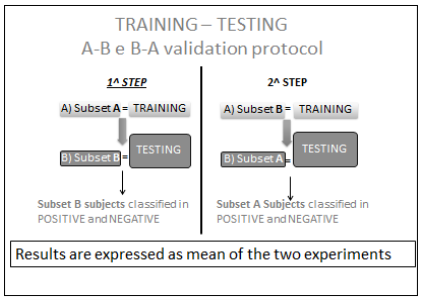
Inclusion criteria:

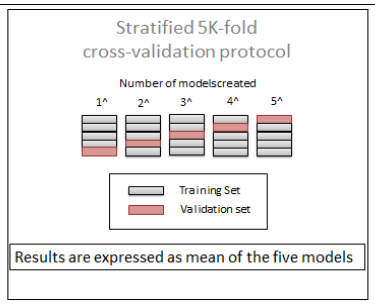
- patients admitted to the ED of Niguarda hospital with influenza-like symptoms
- a performed SARS-COV-2 test in ED.

Exclusion criteria:

- Patients < 12 years old
- absence of leukocyte formula in the ED.

We created a dataset with 74 variables, collected from a combination of ***clinical, radiological and routine laboratory*** findings upon hospital admission. The RT-PCR results constituted the target variable of the prediction model.





Different Machine Learning algorithms were trained using both the ***Training and Testing AB-BA sequence*** *and* ***the K-fold cross-validation protocol***.

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 AI model selected 42 out of the 74 collected variables.

The best Machine Learning System reached an ***accuracy of 91.4%*** with ***94.1% sensitivity and 88.7% specificity***. Other AI algorithm had similar results but were less efficient.

Machine Learning System	Sensitivity	Specificity	Overall accuracy	AUROC
D_FF_Conic (4x12x12x12)	94.1	88.7	91.4	0.90
D_FF_Conic (6x12x12x12)	92.5	90.2	91.3	0.91
D_FF_Bp (24)	89.2	93.0	91.1	0.93
D_FF_Bp (16x16x16x16)	93.2	88.7	91.0	0.92
D_FF_GNet (64)	90.7	90.2	90.5	0.90
Mean results	**91.9**	**90.2**	**91.1**	**0.91**

The following variables hold a ***considerable weight*** *in the mathematical model:*

- ***Leukocyte***
- ***Lymphocyte***
- ***Eosinophil***
- ***Basophil***
- ***Monocyte***
- ***Age***
- ***Female sex***

Conclusion

Our exploratory 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 in larger multicentre studies, this could have important clinical and organizational implications.