

Prediction of COVID-19 from hemogram results and age using machine learning

Elena Caires Silveira^{1*} 

¹Medical Student at Multidisciplinary Institute for Health, Federal University of Bahia (Universidade Federal da Bahia), Brazil

Article Info

Article type:

Research

Article History:

Received: 2020-08-17

Accepted: 2020-08-21

Published: 2020-08-23

* Corresponding author:

Elena Caires Silveira

Medical Student at Multidisciplinary Institute for Health, Federal University of Bahia (Universidade Federal da Bahia), Brazil

Email: elenacaires@gmail.com

Keywords:

Coronavirus Infections

COVID-19

Machine Learning

Blood Cell Count

ABSTRACT

Introduction: The rapid global dissemination of COVID-19 culminated in the mobilization of great technological efforts aimed at its better understanding and control. In this paper, Machine Learning gains notoriety, and its application has been widely documented for pathophysiological, diagnostic, therapeutic, prognostic and monitoring of COVID-19 purposes. The present paper aimed to build a model for the prediction of the diagnosis of COVID-19 based on blood count results and age of patients and to identify the main characteristics taken into account by the algorithm for the predictive decision.

Material and Methods: Anonymous data from 1157 patients made available by the COVID-19 Data Sharing / BR repository were used. The work took place in two distinct stages: description and analysis of the data; and construction of the predictive model.

Results: With the exception of hemoglobin measurement, mean corpuscular volume, red cell distribution width, mean platelet volume and neutrophil-lymphocyte ratio, there was a statistically significant association of all other hematological parameters assessed with COVID-19. The predictive model developed from the XGBoost classifier reached an accuracy of 80.0% with a sensitivity of 75.6% and specificity of 82.0%. The variables that had the greatest influence on the predictive decision were basophil, eosinophil and leukocyte measurements.

Conclusion: The present paper confirms the potential of using blood count results, a widely available and accessible test, in the context of the diagnostic evaluation and pathophysiological investigation of COVID-19. Conclusion: This work highlights the relevance of the systematization and dissemination of data related to COVID-19 for use in new research.

Cite this paper as:

Caires Silveira E. Prediction of COVID-19 From Hemogram Results and Age Using Machine Learning. Front Health Inform. 2020; 9: 39. DOI: [10.30699/fhi.v9i1.234](https://doi.org/10.30699/fhi.v9i1.234)

INTRODUCTION

In late 2019, a new type of coronavirus, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes the pandemic disease COVID-19, emerged in Wuhan province, Hubei, China [1]. Coronaviruses belong to the Coronaviridae family and members of Group IV ((+) ssRNA) in the Baltimore classification and are a group of enveloped and spiked surface viruses capable of infecting humans and a wide variety of animals [2, 3]. SARS-CoV-2 belongs to line B of beta-corona viruses and has a high degree of homology with previous SARS-CoV and MERS-CoV [4, 5]. Such virus enters host cells via endocytosis after binding to the receptor ACE2, and has a variable incubation period of 1 to 14 days

[6].

Viral transmission between individuals can occur through direct, close or indirect contact, through infected secretions or droplets [7-9]. Such secretions and droplets can still contaminate surfaces and objects, thus enabling transmission by fomites [10, 11]. In addition, SARS-CoV-2 airborne transmission was observed during medical procedures that lead to aerosol production, and the possibility of this type of transmission in the absence of aerosol-generating procedures is being investigated [12-14].

COVID-19 has a wide clinical spectrum, ranging from asymptomatic patients to cases of pneumonia, severe acute respiratory syndrome and multiple organ dysfunction [15]. In symptomatic patients, the most

common manifestations are fever, cough, dyspnea, fatigue, diarrhea, muscle pain and headache [15, 16]. Despite manifesting mainly as a respiratory infection, there is evidence that COVID-19 involves multiple systems, including the cardiovascular, respiratory, gastrointestinal, neurological, hematopoietic and immune systems [17].

The rapid global spread of COVID-19, recognized by the World Health Organization (WHO) as a pandemic in March 2020 [18], culminated in the mobilization of major technological efforts aimed at its better understanding and control. In this context, the modality of Artificial Intelligence known as Machine Learning gains notoriety, and its application has been widely documented for pathophysiological, diagnostic, therapeutic, prognostic and monitoring of COVID-19 purposes [19-23].

Machine Learning is a scientific discipline arising from the intersection between statistics and computer science that focuses on the development of algorithms capable of learning from data through pattern recognition [24, 25]. In supervised learning, algorithms are trained from data sets composed of instances and labels (training sets), so that they can later be used to predict unknown labels for new instances provided, this predictive ability being verified from test sets, where the labels are not provided to the algorithm [25].

In the last decade, Machine Learning has aroused the interest of researchers and physicians due to its great possibility of contributing to Medicine and, conversely, the large volume of data generated by medical activities has aroused the interest of computer scientists. As highlighted by Cabitza and Banfi [26], the potential for applying Machine Learning to laboratory data for diagnostic and prognostic purposes deserves attention.

Considering the above, the present paper proposed to build a prediction model using the supervised classifier XGBoost for the diagnosis of COVID-19 based on blood count results, age and sex of patients and to identify the main characteristics (instances) taken into account by the algorithm for the predictive decision. The XGBoost is an efficient, flexible and portable library which implements machine learning algorithms providing decision-tree-based ensembling under the Gradient Boosting framework [27]. The choice of hematological data as instances for the prediction is supported by the identification of the correlation between hematological parameters and the diagnosis of COVID-19 in recent studies, as well as in the wide availability of the blood count, characterized by being a practical and accessible exam.

MATERIAL AND METHODS

The present paper aimed to construct a classification

algorithm oriented to the prediction of the diagnosis of COVID-19 based on parameters from the blood count, age and sex of the patients. For this purpose, anonymous data from 1157 patients made available by the repository COVID-19 Data Sharing/BR were used, referring to patients who underwent laboratory tests for the diagnosis of COVID-19 in the Fleury Network from May 29, 2020 to June 15, 2020.

The hematological parameters considered were: hemoglobin measurement (g/dL); erythrocytes (millions/mm³); hematocrit (%); mean corpuscular volume, or MCV (fL); mean corpuscular hemoglobin or MCH (pg); mean corpuscular hemoglobin concentration, or MCHC (g/dL); red cell distribution width, or RDW (%); global leukotic count (mm³); neutrophil count (mm³); lymphocyte count (mm³); eosinophil count (mm³); basophil count (mm³); monocyte count (mm³); platelet count (mm³); mean platelet volume, or MPV (fL); and neutrophil-lymphocyte ratio, or NLR. COVID-19 confirmed cases were those for which the result of the polymerase chain reaction test after reverse transcription (RT-PCR) for SARS-CoV-2 performed from a sample obtained from a nasopharyngeal swab was positive. Such a test is the gold standard method for diagnosing the disease. Only patients with data available for all evaluated variables were included in the study, with a maximum difference of 01 days between the collections of materials for both exams being admitted. After the collection, the work study place in two distinct stages: description and analysis of the data; and construction of the predictive model.

For the descriptive analysis of the data, the patients were grouped into positive and negative for COVID-19 according to the result of the RT-PCR test for SARS-CoV-2. Categorical variables were represented as proportions and continuous variables were represented as means and standard deviations within each group of patients. The proportions for categorical variables were compared between groups using the χ^2 test. For continuous variables, after denying the hypothesis of normal distribution using the Shapiro-Wilk test, a comparison between groups was performed using the Mann-Whitney U test. In both cases, p-values less than 0.05 were considered significant. For the construction of the predictive model, the data were segregated into two distinct sets: training set (which corresponded to 75% of the data), used during the making and improvement of the algorithm; and the independent test set (which corresponded to the remaining 25%), used to assess the model's predictive capacity. Therefore, the instances (hematological data, sex and age) and the labels (positive or negative result of the RT-PCR test) of the training set were presented to the algorithm, so that it could later predict, from only the instances of the test set, their respective labels.

The supervised classifier XGBoost, a scalable machine

learning system available as an open source package - was adopted as the predictor, given its good performance in characterizing patterns and the possibility to list the degree of importance of each variable for the predictions [28].

The evaluation of the model was based on the comparison between the real labels of each test patient (not presented to the algorithm) and the labels provided by the algorithm for such. Accuracy, specificity, sensitivity, F1-score and ROC-curve together with the value corresponding to the area under the curve (AUC) were used as evaluation parameters.

All steps of statistical analysis and development of the predictive model were performed in Python (version 3.6.9). The methodology adopted is schematized in Fig 1.

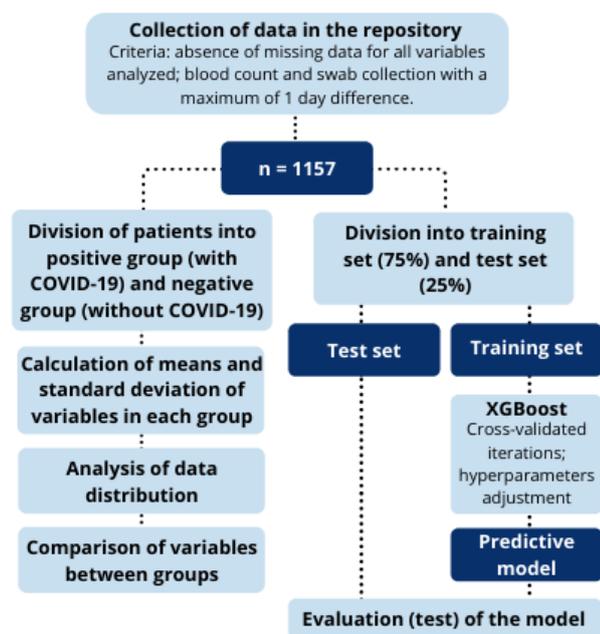


Fig 1: Methodological design of the study

RESULTS

Among the 1157 patients analyzed, 349 presented positive results for the RT-PCR test (positive group), being therefore, for the purposes of the study, considered as cases of COVID-19, while 808 presented negative results (negative group). The positive group consisted of 164 (45.9%) males and 185 (54.1%) females, with an average age of 50.4 years (standard deviation = 16.8). The negative group had 342 (42.3%) males and 466 (57.7%) females, with an average age of 45.7 years (standard deviation = 17.8). The comparison of sex proportions using the χ^2 test revealed that there was no statistically significant difference (95% CI: -2.5947% to 9.8295%, $p = 0.257$) between the two groups. Regarding age, there was a statistically significant difference between the distributions in the two groups (p

<0.001) according to the Mann-Whitney U test.

The means and standard deviations of each hematological variable in the positive and negative groups are available in Table 1, as well as the p values from the Mann-Whitney U test for comparison between groups.

Table 1: Comparative analysis of hematological parameters of patients with COVID-19 (RT-PCR positive) and without COVID-19 (RT-PCR negative)

Hematological parameter	Mean \pm standard deviation		p-value*
	With COVID-19 (n=349)	Sem COVID-19 (n=808)	
Hemoglobin	13,87 \pm 1,44	13,81 \pm 1,50	0,118
Erythrocytes	4,76 \pm 0,53	4,71 \pm 0,52	0,015
Hematocrit	41,68 \pm 4,08	41,35 \pm 4,14	0,032
MCV	87,89 \pm 4,95	88,15 \pm 5,16	0,181
MCH	29,23 \pm 1,84	29,75 \pm 9,76	0,024
MCHC	33,19 \pm 1,95	33,39 \pm 1,42	0,006
RDW	13,16 \pm 1,20	13,08 \pm 1,16	0,078
Leukocytes	5849,80 \pm 6063,18	7214,44 \pm 5184,46	$< 0,001$
Neutrophils	3190,69 \pm 1772,83	4133,72 \pm 2287,76	$< 0,001$
Lymphocytes	1673,09 \pm 669,92	2172,97 \pm 2422,24	$< 0,001$
Eosinophils	98,19 \pm 168,75	171,57 \pm 183,31	$< 0,001$
Basophils	21,83 \pm 15,67	37,09 \pm 21,82	$< 0,001$
Monocytes	566,56 \pm 257,63	603,25 \pm 329,23	0,042
Platelets	224982,81 \pm 76408,50	253806,68 \pm 66371,22	$< 0,001$
NLR	2,27 \pm 1,96	2,43 \pm 2,31	0,071

Data source: COVID-19 Data Sharing/BR repository. *p-values were determined using the Mann-Whitney U test. MCV = mean corpuscular volume; MCH = mean corpuscular hemoglobin; MCHC = mean corpuscular hemoglobin concentration; RDW = red cell distribution width; NLR = neutrophil-lymphocyte ratio.

With the exception of hemoglobin measurement, mean corpuscular volume, red cell distribution width, mean platelet volume and neutrophil-lymphocyte ratio, there was a statistically significant difference for the values of all other hematological parameters analyzed between the two groups. The significance for the values of leukocytes, neutrophils, lymphocytes, eosinophils, basophils and platelets was more expressive, for which the p-value was less than 0.001.

The predictive model, an XGBoost classifier, was trained with data from 867 patients, among whom 259 belonged to the positive group and 608 belonged

to the negative group. The values of hyperparameters that demonstrated greater balance in minimizing bias errors and variance errors, as well as in the relationship between sensitivity and specificity, were: learning rate (`learning_rate`)=0.005; maximum depth of each decision tree (`max_depth`)=3; number of decision trees (`n_estimators`)=420; regularization value (`reg_lambda`)=0.5.

In the independent test set, consisting of data from 290 patients (of which 90 belonged to the positive group and 200 belonged to the negative group), the model reached an accuracy of 80.0% with a sensitivity of 75.6% (CI: 65.4% to 84.0%), specificity of 82.0% (CI: 76.0% to 87.1%), positive predictive value of 65,38%, (CI: 57,88% to 72,20%), negative predictive value of 88,17% (CI: 83,75% to 91,51%), F1-score of 70.1% and AUC value of 81.1%. The confusion matrix regarding model predictions and true results (RT-PCR test results for SARS-CoV-2 in a nasopharyngeal swab sample) is shown in Fig 2, and the ROC curve for the model is shown in Fig 3.

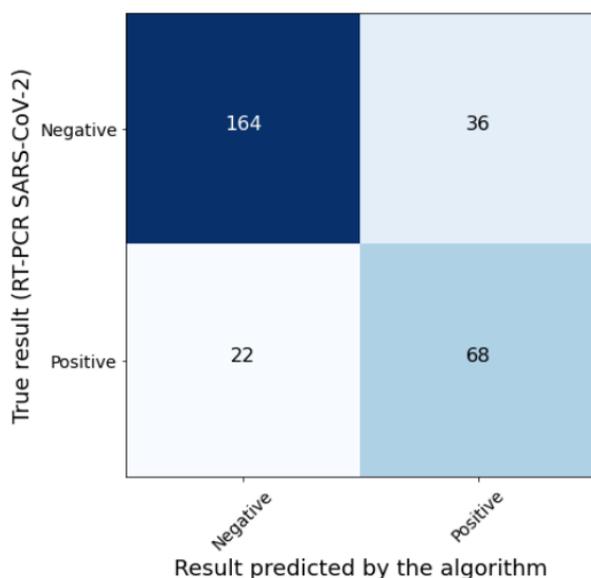


Fig 2: Confusion matrix for the predictive model

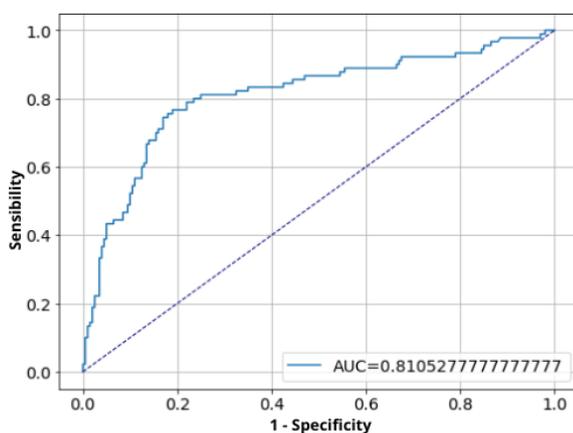


Fig 3: ROC curve and AUC-score for the predictive model

The variables that had the greatest influence on the predictive decision of the model were the basophil dosage (17.7%), the eosinophil dosage (11.0%) and the leukocyte dosage (7.5%). Those that exercised less influence were sex (which did not exert any influence), the mean platelet volume (2.1%) and the mean corpuscular volume (3.8%). The percentage importance of each instance for the prediction of the algorithm is shown in Table 2.

Table 2: Degree of relevance of each variable for the predictive model

Variable	Predictive relevance (%)
Basophils	17,7
Eosinophils	11,0
Leukocytes	7,5
Lymphocytes	6,5
Erythrocytes	6,3
Platelets	6,0
MCHC	5,5
Hematocrit	5,3
Age	5,3
Neutrophils	5,1
Monocytes	5,1
MCH	4,8
Hemoglobin	4,2
NLR	4,0
RDW	4,0
MCV	3,8
MPV	2,1
Sex	0,0

DISCUSSION

The present study demonstrated the association of COVID-19 with the change in the values distribution of several hematological parameters, in particular the global leukocyte count and the counts of neutrophils, lymphocytes, eosinophils, basophils and platelets. For all of these parameters, lower averages were found in the group of patients with COVID-19. Such findings are consistent with those described by Ferrari et al. [29] and Dorn et al. [24].

The predictive model developed in the study was able to achieve a specificity of 82.0% and a sensitivity of 75.6% in the independent test set, demonstrating the potential of using blood count results, a widely available and accessible test, in the context of the clinical evaluation of COVID-19. In this context, it is valid to point out that the use of CBC data has already been proposed as a parameter of great utility in the diagnosis and management of viral pandemics [30]. Furthermore, the performance obtained by the model reflects the possibility of recognizing hematological patterns that point to the diagnosis of COVID-19, which reinforces the clinical significance of the change in the distribution of several hematological parameters associated with COVID-19 observed in the analytical stage of the study.

Stratification of data according to the predictive value of each of them is another major contribution of the model, given that it may not only guide a careful clinical observation of the parameters identified as most relevant (such as basophil and eosinophil measurements) in the face of suspicion of COVID-19, but also contribute to understanding the pathophysiological process involved in the disease. It has already been described in the literature that eosinophils and basophils can play a central role in the organism's response to SARS-CoV-2 [31, 32], and the present study corroborates the relevance of investigating this role. This study presents as one of its main limitations the fact that the developed predictive model - as well as all supervised machine learning models - is totally data-driven, so that, in the face of new data with distinct characteristics patterns, performance may be compromised. In addition, the model would benefit from the availability of more data. Another limitation is the non-availability of information regarding the clinical context of symptoms and severity of patients in the repository used since COVID-19 has a wide spectrum of clinical presentations, it is inferable that it also has a wide laboratory spectrum - there are studies which describe different laboratory profiles for different classifications of disease severity -, making it difficult to recognize patterns in the absence of such information [15, 17, 31]. Reports of variable sensitivity of the RT-PCR test for SARS-CoV-2 must also be considered: since the test was the definer of the true labels in the study, the occurrence of false negatives in the training set and/or the test set could reflect inadequacies in the development and/or evaluation of the model, respectively [33-35].

CONCLUSION

REFERENCES

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*. 2020; 5(4): 536-44. PMID: 32123347 DOI: 10.1038/s41564-020-0695-z [PubMed]
2. Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health*. 2020; 25(3): 278-80. PMID: 32052514 DOI: 10.1111/tmi.13383 [PubMed]
3. Ouassou H, Kharchoufa L, Bouhrim M, Daoudi NE, Imtara H, Bencheikh N, et al. The pathogenesis of coronavirus disease 2019 (COVID-19): Evaluation and prevention. *J Immunol Res*. 2020; 2020(10): 1-7.
4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet*. 2020; 395(10224): 565-74. PMID: 32007145 DOI: 10.1016/S0140-6736(20)30251-8 [PubMed]
5. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020; 579(7798): 270-73. PMID: 32015507 DOI: 10.1038/s41586-020-2012-7 [PubMed]
6. 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. PMID: 31978945 DOI: 10.1056/NEJMoa2001017 [PubMed]
7. Liu J, Liao X, Qian S, Yuan J, Wang F, Liu Y, et al. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, 2020. *Emerg Infect Dis*. 2020; 26(6): 1320-3. PMID: 32125269 DOI: 10.3201/eid2606.200239 [PubMed]
8. World Health Organization. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19) [Internet]. 2020 [cited: 20 Jul 2020]. Available from: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on->

This study, in line with the emerging technological efforts to contribute to the management of the international health crisis resulting from COVID-19, developed a predictive model capable of predicting with 80% accuracy the test result for the disease only from results of blood count and patient age. In this context, it is important to highlight the relevance of systematizing and disseminating data related to the disease so that they can be used in related research. It is also important that researchers from different areas of knowledge engage in such studies.

ACKNOWLEDGEMENT

This work used data made available by the repository COVID-19 Data Sharing/BR, available at: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>. I thank and congratulate the São Paulo Research Foundation and the collaborating/participant institutions - University of São Paulo, Fleury Institute, Sírio-Libanês Hospital and Israelita Albert Einstein Hospital for the initiative, which makes a notable contribution to the development of research on this theme.

AUTHOR'S CONTRIBUTION

The author agree on this final form of the manuscript.

CONFLICTS OF INTEREST

The author declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

- covid-19-final-report.pdf
9. Luo L, Liu D, Liao X, Wu X, Jing Q, Zheng J, et al. Modes of contact and risk of transmission in COVID-19 among close contacts (pre-print). *MedRxiv*. 2020. DOI: 10.1101/2020.03.24.20042606 [[PubMed](#)]
 10. Van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl J Med*. 2020; 382(16): 1564-7. PMID: 32182409 DOI: 10.1056/NEJMc2004973 [[PubMed](#)]
 11. Wu S, Wang Y, Jin X, Tian J, Liu J, Mao Y. Environmental contamination by SARS-CoV-2 in a designated hospital for coronavirus disease 2019. *Am J Infect Control*. 2020;48(8):910-4. PMID: 32407826 DOI: 10.1016/j.ajic.2020.05.003 [[PubMed](#)]
 12. World Health Organization. Advice on the use of masks in the context of COVID-19 [Internet]. 2020 [cited: 20 Jul 2020]. Available from: [https://www.who.int/publications/i/item/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-\(2019-ncov\)-outbreak](https://www.who.int/publications/i/item/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak)
 13. Stadnytskyi V, Bax CE, Bax A, Anfinrud P. The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc Natl Acad Sci*. 2020; 117(22): 11875-7. PMID: 32404416 DOI: 10.1073/pnas.2006874117 [[PubMed](#)]
 14. Somsen GA, van Rijn C, Kooij S, Bem RA, Bonn D. Small droplet aerosols in poorly ventilated spaces and SARS-CoV-2 transmission. *Lancet Respir Med*. 2020; 8(7): 658-9.
 15. Singhal T. A review of coronavirus disease 2019 (COVID-19). *Indian J Pediatr*. 2020; 87(4): 281-6. PMID: 32166607 doi: 10.1007/s12098-020-03263-6 [[PubMed](#)]
 16. Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang YQ, et al. Lymphopenia predicts disease severity of COVID-19: A descriptive and predictive study. *Signal Transduct Target Ther*. 2020; 5(1): 33. PMID: 32296069 DOI: 10.1038/s41392-020-0148-4 [[PubMed](#)]
 17. Terpos E, Ntanasios-Stathopoulos I, Elalamy I, Kastritis E, Sergentanis TN, Politou M, et al. Hematological findings and complications of COVID-19. *Am J Hematol*. 2020;95(7):834-47. PMID: 32282949 DOI: 10.1002/ajh.25829 [[PubMed](#)]
 18. World Health Organization. Director general's opening remarks at the media briefing on COVID19 [Internet]. 2020 [cited: 1 Jul 2020]. Available from: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---29-june-2020>
 19. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. *Physiol Genomics*. 2020; 52(4): 200-2. PMID: 32216577 DOI: 10.1152/physiolgenomics.00029.2020 [[PubMed](#)]
 20. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy. *Radiology*. 2020; 296(2): E65-71. PMID: 32191588 DOI: 10.1148/radiol.2020200905 [[PubMed](#)]
 21. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet*. 2020; 395(10225): 689-97. PMID: 32014114 DOI: 10.1016/S0140-6736(20)30260-9 [[PubMed](#)]
 22. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020; 14(4): 337-9.
 23. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *Lancet Digit Health*. 2020; 2(4): e166-7. PMID: 32289116 DOI: 10.1016/S2589-7500(20)30054-6 [[PubMed](#)]
 24. Deo RC. Machine learning in medicine. *Circulation*. 2015; 132(20): 1920-30. PMID: 26572668 DOI: 10.1161/CIRCULATIONAHA.115.001593 [[PubMed](#)]
 25. Kan A. Machine learning applications in cell image analysis. *Immunol Cell Biol*. 2017; 95(6): 525-30. PMID: 28294138 DOI: 10.1038/icb.2017.16 [[PubMed](#)]
 26. Cabitza F, Banfi G. Machine learning in laboratory medicine: Waiting for the flood? *Clin Chem Lab Med*. 2018; 56(4): 516-24. PMID: 29055936 DOI: 10.1515/cclm-2017-0287 [[PubMed](#)]
 27. XGBoost Developers. XGBoost documentation [Internet]. 2020 [cited: 1 Aug 2020]. Available from: <https://xgboost.readthedocs.io/>.
 28. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *International Conference on Knowledge Discovery and Data Mining*. Arxiv; 2016.
 29. Ferrari D, Motta A, Strollo M, Banfi G, Locatelli M. Routine blood tests as a potential diagnostic tool for COVID-19. *Clin Chem Lab Med*. 2020; 58(7): 1095-9. PMID: 32301746 DOI: 10.1515/cclm-2020-0398 [[PubMed](#)]
 30. Shimoni Z, Glick J, Froom P. Clinical utility fo the full blood count in identifying patients with pandemic Influenza A (H1N1). *Journal of Infection*. 2013; 66(6): 545-7. PMID: 23318262 DOI: 10.1016/j.jinf.2013.01.001 [[PubMed](#)]
 31. Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, et al. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis*. 2020; 12: ciaa248. PMID: 32161940 DOI: 10.1093/cid/ciaa248 [[PubMed](#)]
 32. Rodriguez L, Pekkarinen P, Tadepally LK, Tan Z, Consiglio CR, Pou C, et al. Systems-level immunomonitoring from acute to recovery phase of severe COVID-19. *Cell reports. Medicine*. 2020 :100078. PMID: 32838342 DOI: 10.1016/j.xcrm.2020.100078 [[PubMed](#)]
 33. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-

-
- 19 detection: Issues affecting the results. *Expert Rev Mol Diagn.* 2020; 20(5): 453-4. PMID: 32297805 DOI: 10.1080/14737159.2020.1757437 [[PubMed](#)]
34. West CP, Montori VM, Sampathkumar P. COVID-19 testing: The threat of false-negative results. *Mayo Clin Proc.* 2020; 95(6): 1127-9. PMID: 32376102 DOI: 10.1016/j.mayocp.2020.04.004 [[PubMed](#)]
35. Xiao AT, Tong YX, Zhang S. False negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence. *J Med Virol.* 2020. [Epub ahead of print]. PMID: 32270882 DOI: 10.1002/jmv.25855 [[PubMed](#)]