

*Life Research***Early detection of COVID-19 using characteristic leucocyte differential count (CLDC)**Mark R Baker^{1*}, Jessica Rogge¹¹Complex Systems Research Division, Hypatia Solutions Ltd & MediChain Ltd, Impact Hub Kings Cross, London, UK.***Corresponding to:** Mark R Baker. Complex Systems Research Division, Hypatia Solutions Ltd & MediChain Ltd, Impact Hub King's Cross, 34b York Way, King's Cross London, N1 9AB, UK. Email: mark.baker@hypatia-solutions.com.**Abstract**

Background: COVID-19 is an acute infection of the respiratory tract that emerged in late 2019. Currently identified methods for identifying the severe acute respiratory syndrome coronavirus 2 virus include methods that detect the presence of the virus itself, such as reverse transcription PCR and isothermal amplification methods, and those that detect antibodies produced in response to the infection. Reverse transcription PCR and quantitative PCR are highly sensitive but have a narrow time window of sensitivity. **Methods:** We investigated a new method to detect the occurrence of severe acute respiratory syndrome coronavirus 2 by analyzing the early rise in leukocyte levels which has a characteristic set of ratios of leukocyte types which identify the viral pathogen and distinguish it from a number of others. We used the Albert Einstein Hospital, São Paulo data set and the Athena AI System to validate this method. **Results:** The sensitivity of the test is up to 98.67% prediction of positives from full blood count results. **Conclusion:** We have discovered an early test for SARS-CoV-2 which can be performed using a black-boxed AI to give high sensitivity prediction of COVID-19 infection.

Key words: COVID-19, SARS-CoV-2, H1N1, H5N1, Pandemic testing, Early-diagnosis, Mass screening, Home-testing

Author contributions:

Mark R Baker contributed to the text and Athena analysis; Jessica Rogge contributed to design and produce the figures.

Abbreviations:

COVID-19, coronavirus disease 2019; CLDC, characteristic leucocyte differential count; FBC, full blood count; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; qPCR, quantitative PCR; RT-PCR, reverse transcription-PCR; RT-qPCR, real-time quantitative PCR.

Competing interests:

Mark R Baker is a director of MediChain Ltd and Hypatia Solutions Ltd and owns all rights to the black boxed Athena methods are protected worldwide under patent pending GB2006742.7 and GB2006374.9

Citation:

Mark R Baker. Early detection of COVID-19 using characteristic leucocyte differential count (CLDC). *Life Research* 2020, 3 (3): 101–107.

Executive Editor: Yu-Ping Shi.

Submitted: 13 June 2020, **Accepted:** 10 July 2020, **Online:** 23 July 2020

Background

At present, researchers around the world have proposed many viral detection methods in response to coronavirus disease 2019 (COVID-19), such as reverse transcription-PCR (RT-PCR) and isothermal amplification methods, and those that detect antibodies produced in response to the infection [1–3]. RT-PCR and quantitative PCR (qPCR) are highly sensitive but have a narrow time window of sensitivity [4–7]. These existing tests show sensitivity 1–4 days after the initial infection, with sensitivity dropping off thereafter. Antibody tests become sensitive 2–4 weeks after the initial infection, due to the latency time needed to form antibodies. We found a new method of acquitting the occurrence of severe acute respiratory syndrome coronavirus 2 and resulting COVID-19, detecting the early rise in leukocyte levels which has a characteristic set of ratios of leukocyte types which identify the viral pathogen and distinguish it from a number of others [8].

At the time of writing countries all over the world are releasing citizens from lockdown. However, current levels of testing are not adequate in any country to allow reliably prevent a second or subsequent wave of COVID-19 or future waves (nominally COVID-2X) or to release from lockdown on the basis of testing. Rate of coronavirus COVID-19 tests performed worldwide as of June 2020 are less than 10% in almost every major country [9].

Here we describe a new test which can be 98.67% prediction of positives from full blood count (FBC) results, sensitive up to 14 days earlier than real-time quantitative PCR (RT-qPCR), at a cost at least an order of magnitude lower than other tests such as RT-qPCR and antibody tests.

Universal testing forms a powerful key and strategy to ending lockdown and preventing recurrence both of COVID 19, of mutants and future pandemics [10]. There are multiple types of COVID 9 tests although 2 are prevalent.

Amongst new contenders are loop-mediated isothermal amplification: a simple, but less developed testing method, lateral flow: hand-held single-use assays providing results for an individual patient in as short as 15 minutes, and enzyme-linked immunosorbent assay: quick and technically simple assays that are easily read and while they offer relatively high throughput are not of the same magnitude as characteristic leucocyte differential count (CLDC) [11].

The PCR viral tests currently see if an individual has a COVID-19 infection. While it represents the gold standard for positive testing of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), false negative PCR testing of SARS-CoV-2 is common. Each patient needs to be evaluated based on their current symptoms, the exact timing and other relevant clinical information.

This means the negative PCR result does not mean the patient is free from infection. The problem is there for if the PCR test shows negative it may be that the patient is in the period which is up to 14 days prior to PCR being sensitive in which they may be able to infect others but not show up on PCR tests. Later on PCR does not show positive when the patient has previously had the disease. Variations include a real-time fluorescent RT-PCR kit for detecting the novel coronavirus (SARS-2019-nCoV) which has a U.S. Food and Drug Administration Emergency Use Authorization but broadly has the same strengths and weaknesses as other PCRs [12].

Antibody tests can test to see if a patient has had SARS-CoV-2 previously. The accuracy of any of many antibody tests with past COVID-19 infections depends on both the sensitivity and specificity of the individual tests and the underlying prevalence of COVID-19 infections in that population. Antibody tests are not generally sensitive till at least three weeks after the initial infection [13–15]. This means that as a means of determining who is released from quarantine like PCR they are a poor indicator and will allow mess infection. they are however sensitive for a period after the infection. Furthermore, these tests are limited in the numbers that can be applied due to the need for expensive reagents, and because of the possibility particularly with antibodies that new mutants and new strain will not be sensitive to the existing tests both allowing infective individuals to reenter the population and making large stockpiles that may be gathered of the tests essentially useless for testing purposes. Furthermore, any test which is a reagent like PCR or antibody tests is limited by the ability to the production of that agent and the availability. This generally makes personal testing on a regular basis extremely impractical.

Kucirka et al. [11] found that approximately two-thirds of patients who test negative 4 days after exposure are otherwise pre-symptomatic meaning that the majority of infected individuals will not be detected even when symptoms appear. Nearly four percent of patients tested negative on the first day of showing physical symptoms. The lowest false negative rate was actually found 3 days after the symptoms appeared, with a level of 20 percent. What is needed is a test that can detect SARS-CoV-2 much earlier than PCR, and more reliably than either antibody tests or PCR. The report here a test which fulfills both these criteria and has the potential to be sensitive not only to mutations and new strains of COVID-19 but to other pandemics as they arise. There is huge potential for ending lockdown, preventing future waves of COVID-19 and future pandemics.

Materials and methods

Data sources

The data set used in developing this technique was the anonymized data from patients seen at the Hospital Israelita Albert Einstein, at São Paulo, Brazil, from samples collected to perform the RT-PCR of SARS-CoV-2 [16]. The data had been anonymized and clinical data were standardized to have a mean of zero and a unit standard deviation. Five hundred and twelve patients out of a total of 5,644 had data fields which were complete enough for analysis. The data was open-sourced by the provider, the Albert Einstein hospital, in accordance with the declaration of Helsinki.

Available data fields

Patient ID, patient age quantile, SARS-Cov-2 exam result, patient admitted to regular ward (1 = yes, 0 = no), patient admitted to semi-intensive unit (1 = yes, 0 = no), patient admitted to intensive care unit (1 = yes, 0 = no), hematocrit, hemoglobin, platelets, mean platelet volume, red blood cells, lymphocytes, mean corpuscular hemoglobin concentration, leukocytes, basophils, mean corpuscular hemoglobin, eosinophils, mean corpuscular volume, monocytes, red blood cell distribution width, serum glucose, respiratory syncytial virus, influenza A, influenza B, parainfluenza 1, coronavirus NL63, rhinovirus/enterovirus, mycoplasma pneumoniae, coronavirus HKU1, parainfluenza 3, chlamydomphila pneumoniae, adenovirus, parainfluenza 4, coronavirus 229 elowing viral (and other) pathogens, coronavirus OC43, influenza A H1N1 2009, bordetella pertussis, metapneumovirus, parainfluenza 2, influenza B (rapid test), influenza A (rapid test), alanine transaminase, aspartate transaminase, gamma-glutamyltransferase, total bilirubin, direct bilirubin, indirect bilirubin, alkaline phosphatase, ionized calcium, *Strepto* A, magnesium, pCO₂ (venous blood gas analysis), Hb saturation (venous blood gas analysis), base excess (venous blood gas analysis), pO₂ (venous blood gas analysis), FiO₂ (venous blood gas analysis), total CO₂ (venous blood gas analysis), pH (venous blood gas analysis), HCO₃ (venous blood gas analysis), rods #, segmented, promyelocytes, metamyelocytes, myelocytes, myeloblasts, urineresterase, urine-aspect, urine-pH, urine-hemoglobin, urine-bile pigments, urine-ketone bodies, urine-nitrite, urine-density, urine-urobilinogen, urine-protein, urine-sugar, urine-leukocytes, urine-crystals, urine-red blood cells, urine-hyaline cylinders, urine-granular cylinders, urine-yeasts, urine-color, partial thromboplastin time, relationship (patient/normal), international normalized ratio, lactic dehydrogenase, prothrombin time, activity, vitamin B12, Creatine phosphokinase, ferritin, arterial lactic acid, lipase dosage, D-dimer, albumin, Hb saturation (arterial blood gases), pCO₂ (arterial blood gas analysis), base excess (arterial blood gas analysis), pH (arterial blood gas analysis), total CO₂ (arterial blood gas analysis), HCO₃ (arterial blood gas analysis),

pO₂ (arterial blood gas analysis), arterial FiO₂, phosphor, ctO₂ (arterial blood gas analysis).

Statistical analysis

The analysis was carried out using the black-boxed MediChain/Hypatia Athena AI using neural net, heuristic and algorithmic approaches. MediChain's new S3ER filter screens negatives to minimize false negatives and indicate the need for a retest. Proportions vary on filter thresholds selected and the quality of original blood test data. The general approach was the auto filter incomplete data sets and feed each parameter and ratio of parameters into nodes in the input layer of an Athena Network (S2 version). The network was then trained to predict RT-PCR positive outcomes.

Results

Using age adjust sets the Athena S3ER algorithm gave 85.06% prediction of positives from FBC results. This was despite data from poor quality data sets. Better data will result in improved sensitivity. Gender information not available in the test set but would be expected to improve this result. Twenty-five percent of age sets gave 100% prediction. Using the eosinophil thresholds in addition to the leukocyte/monocyte differences (the S3ER filter) allows us to flag the need for the retest. This gives a 98.67% prediction of positives from FBC results. The S3ER filter indicates which negatives may have poor data quality and should be retested (Amber category). In this case, there appear to be 1.33% false negatives using the S3ER flag. Platelet measurements would further improve predictive quality. The tests appear highly robust despite sampling artifacts in the test data such as quantization which could be reduced with a better blood measurement protocol.

Figure 1 shows that while there is a different clustering of the platelet/leucocyte ratio in patients clustering positive vs patients clustering negative, by itself, it is not a decisive distinguishing method as there is overlap between to populations. Figure 2 shows that while the eosinophil levels are also different, they too are not distinctive by themselves. Figure 3 shows that there Leucocyte/monocyte gap also has the same features, a difference which is suggestive, but not clearly defining. It is only by combining the three in a heuristically adjusted and preconditioned algorithm which can be implemented through a combination of a special deep learning net and an AI ensemble, that the strong predictive outcomes can be achieved.

In addition, although numbers were low and we are still in the case study phase regarding specificity, however, CDLC tests proved negative to the following viral pathogens. The respiratory syncytial virus, influenza A, influenza B, parainfluenza 1, coronavirus NL63, rhinovirus/enterovirus, mycoplasma pneumoniae, coronavirus HKU1, parainfluenza 3, chlamydomphila

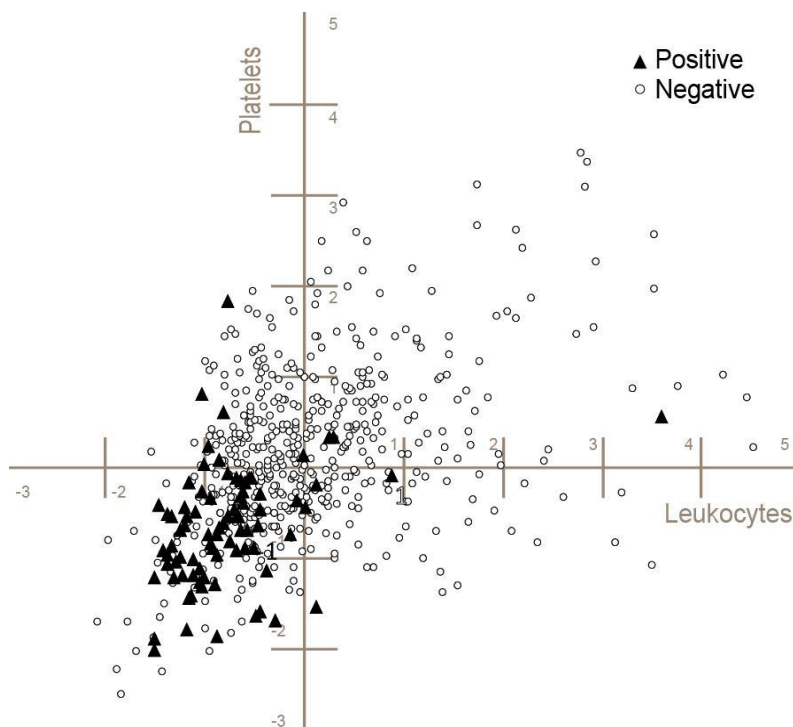


Figure 1 Platelets vs leucocyte count in patients. Solid triangles tested SARS-CoV-2 positive with RT-PCR. Open circles tested negative. The 2-tailed, 2-sample unequal variance t-test gives a $P < 0.001$, but it is notable that by themselves these data do not allow clear distinction of SARS-CoV-2 positive patients. All clinical data were standardized at source to have a mean of 0 and a unit standard deviation. The overlap of clusters may be due to false RT-PCR negatives owing to the narrow time window of RT-PCR sensitivity. RT-PCR, reverse transcription-PCR.

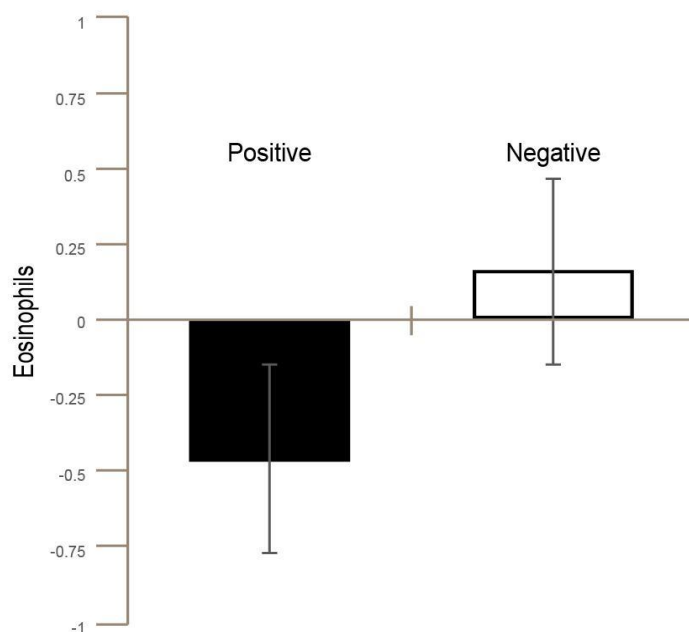


Figure 2 Eosinophil differences in patients tested SARS-CoV-2 positive with RT-PCR. Difference between eosinophil levels in patients tested SARS-CoV-2 positive with RT-PCR (solid) and patients tested negative (open). A trend is indicated, but themselves these data do not allow clear distinction of SARS-CoV-2 positive patients. All clinical data were standardized at source to have a mean of zero and a unit standard deviation. Again, the overlap of error bars may be due to false RT-PCR negatives owing to the narrow time window of RT-PCR sensitivity. RT-PCR, reverse transcription-PCR.

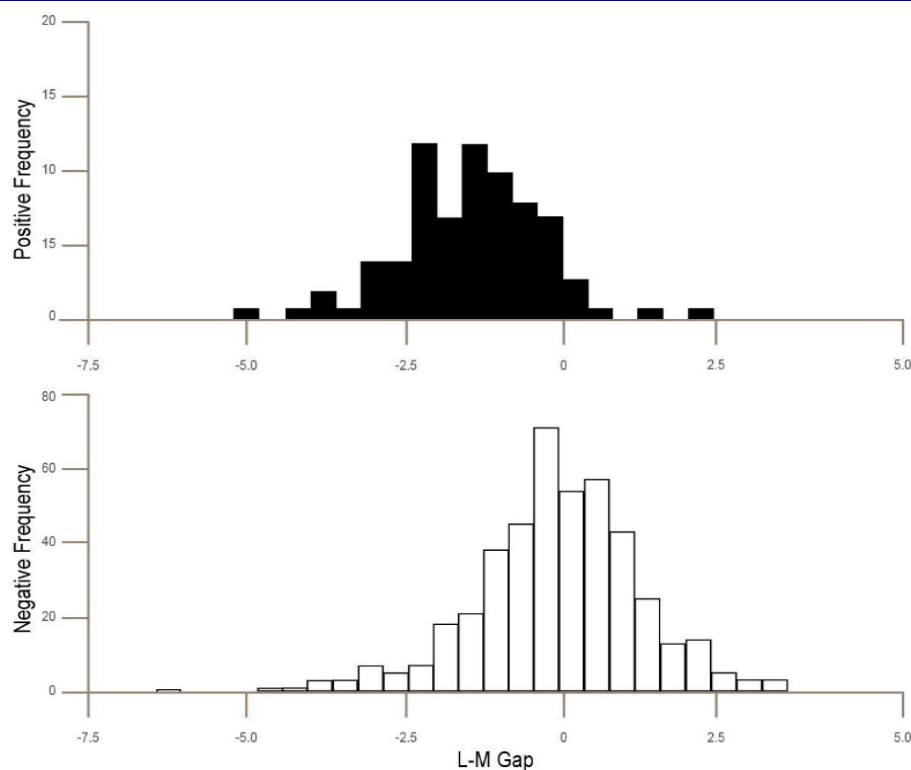


Figure 3 The Leucocyte monocyte gap (L-M Gap) in patients. Solid bars tested SARS-CoV-2 positive with RT-PCR. Open bars tested negative. And again it is notable that by themselves these data do not allow clear distinction of SARS-CoV-2 positive patients. All clinical data were standardized at source to have a mean of zero and a unit standard deviation. Again the overlap of bars may be due to false RT-PCR negatives owing to the narrow time window of RT-PCR sensitivity. RT-PCR, reverse transcription-PCR.

pneumoniae, adenovirus, parainfluenza 4, coronavirus 229E, coronavirus OC43, Influenza A H1N1 2009, *Bordetella pertussis*, metapneumovirus, parainfluenza, "influenza B, rapid test", "influenza A, rapid test", indicating very high specificity within the Albert Einstein data set (Table 1). Because the CLDC test is based on early immune response it could be sensitive up to 14 days earlier than PCR and 21 days earlier than IgM or IgG giving a substantial benefit in terms of being able to detect infection early and stop the spread (Figure 4).

Discussion

What we have discovered is a powerful, early test for SARS-CoV-2 which can be performed using a black-boxed AI to give up to 98.67% prediction of positives from standard results. The overall test using multiple blood factors gives a significance of $P < 0.001$ ($P = 0.000077$). Figures 1–3 show that while individual cell differences are present it is only by using a suitable algorithm that combines all that we can predict reliably creating a CLDC. The overlap of clusters may be due to false RT-PCR negatives owing to the narrow time window of RT-PCR sensitivity. The underlying theory of analysis is that at the time of initial infection before new antibodies are formed and before the virus becomes prevalent the body responds by creating new leukocytes

which it circulates in the blood.

This is a common response found in almost all infections and elevated leukocyte levels are typically a

Table 1 Specificity of CLDC testing

Test	Specificity
Respiratory Syncytial Virus	No false positives
Influenza A	No false positives
Influenza B	No false positives
Parainfluenza 1	No false positives
CoronavirusNL63	No false positives
Rhinovirus/enterovirus	No false positives
Mycoplasma pneumoniae	No false positives
Coronavirus HKU1	No false positives
Parainfluenza 3	No false positives
Chlamydomphila pneumoniae	No false positives
Adenovirus	No false positives
Parainfluenza 4	No false positives
Coronavirus229E	No false positives
CoronavirusOC43	No false positives
Influenza A H1N1 2009	No false positives
Bordetella pertussis	No false positives
Metapneumovirus	No false positives
Parainfluenza 2	No false positives
Influenza B, rapid test	No false positives
Influenza A, rapid test	No false positives

CLDC, characteristic leucocyte differential count.

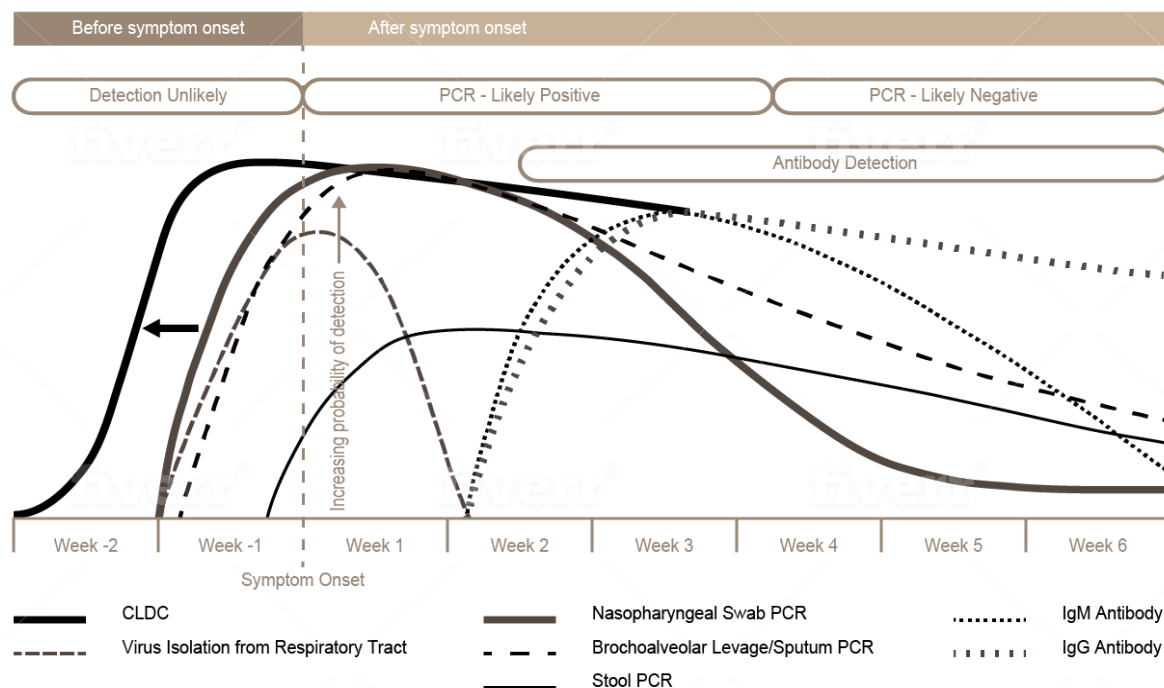


Figure 4 Time dependency of different SARS-CoV-2 tests. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

very early sign of infection or disease. In the case of SARS CoV-2, those are characteristic pattern types of leukocytes formed throughout the immune reaction [17, 18]. These specific types of leukocytes are implicated in the creation of antibodies and their numbers change before antibodies are actually formed. For this reason, we pick up a change in CLDC before the virus is prevalent in the body and therefore before the PCR test can detect it. It is these leukocytes that are creating the antibodies. Before the body has created antibodies it must create leukocytes and that is what we are detecting.

Earlier indications using the same data set suggested low levels of leucocytes and platelets were predictive and that a weak, but not publishable correlation between leucocyte/monocyte ratios and later SARS-Cov-2 exam result when tested with RT-PCR [17, 19]. Data is being acquired in the context of an international emergency and available datasets are limited. The data used here has many limitations but was the best available source at the time. The ability of the methods to perform structure discovery so robustly with such a modest set is a credit to the power and integrity of the algorithm family and AI used. Results given here are for sweep S3 and S3ER.

Conclusion

In this study, we have discovered an early test for SARS-CoV-2 which can be performed using a black-boxed AI to give up to 98.67% prediction blood routine examination results. This study also shows a strong predictive correlation between age-matched normalized

leucocyte/monocyte ratios with eosinophils and platelets used to determine whether retest is required in suspected patients.

References

1. Wang DW, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020, 323: 1061–1069.
2. Colaneri M, Sacchi P, Zuccaro V, et al. Clinical characteristics of coronavirus disease (COVID-19) early findings from a teaching hospital in Pavia, North Italy, 21 to 28 February 2020. *Euro Surveill* 2020, 25: 2000460.
3. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020, 395: 507–513.
4. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, et al. False-negative results of initial RT-PCR assays for covid-19: a systematic review. *MedRxiv Preprint* 2020.
5. Arevalo-Rodriguez I, Steingart KR, Tricco AC, et al. Current methods for development of rapid reviews about diagnostic tests: an international survey. *BMC Med Res Methodol* 2020, 20: 115.
6. Arevalo-Rodriguez I, Moreno-Nunez P, Nussbaumer-Streit B, et al. Rapid reviews of medical tests used many similar methods to systematic reviews but key items were rarely reported: a scoping review. *J Clin Epidemiol* 2019,

- 116: 98–105.
7. Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* 2020, In press.
 8. Mitra A, Dwyre DM, Schivo M, et al. Leukoerythroblastic reaction in a patient with COVID-19 infection. *Am J Hematol* 2020, In press.
 9. Coronavirus Pandemic (COVID-19) Oxford: The Global Change Data Lab [Internet]. Total COVID-19 tests per 1,000 people [cited 13 June 2020]. Available from: <https://ourworldindata.org/coronavirus-testing>.
 10. Oxford, England: Centre for Evidence-Based Medicine [Internet]. What tests could potentially be used for the screening, diagnosis and monitoring of COVID-19 and what are their advantages and disadvantages? [cited 11 June 2020]. Available from: https://www.cebm.net/covid-19/what-tests-could-potentially-be-used-for-the-screening-diagnosis-and-monitoring-of-covid-19-and-what-are-their-advantages-and-disadvantages/https://www.cebm.net/wp-content/uploads/2020/04/CurrentCOVIDTests_descriptions-FINAL.pdf.
 11. Kucirka LM, Lauer SA, Laeyendecker O, et al. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med* 2020, In press.
 12. BGI [Internet]. Real-time fluorescent RT-PCR kit for detecting SARS-2019-nCoV 2020 [cited 11 June 2020]. Available from: <https://www.bgi.com/us/sars-cov-2-real-time-fluorescent-rt-pcr-kit-ivd/>.
 13. Kohmer N, Westhaus S, Rühl C, et al. Clinical performance of different SARS-CoV-2 IgG antibody tests. *J Med Virol* 2020, In press.
 14. Traugott M, Aberle SW, Aberle JH, et al. Performance of SARS-CoV-2 antibody assays in different stages of the infection: comparison of commercial ELISA and rapid tests. *J Infect Dis* 2020, 222: 362–366.
 15. Whitman JD, Hiatt J, Mowery CT, et al. Test performance evaluation of SARS-CoV-2 serological assays. *MedRxiv Preprint* 2020.
 16. Kaggle [Internet]. Hospital Israelita Albert Einstein aSP, Brazil. Diagnosis of COVID-19 and its clinical spectrum hospital Israelita Albert Einstein, at São Paulo, Brazil, 2020 [cited 11 June 2020]. Available from: <https://www.kaggle.com/einsteindata4u/covid19>.
 17. Sun DW, Zhang D, Tian RH, et al. The underlying changes and predicting role of peripheral blood inflammatory cells in severe COVID-19 patients: a sentinel? *Clin Chim Acta* 2020, 508: 122–129.
 18. Coronahack [Internet]. Discussions in Coronahack 2020 in the MaPP team led by Dr Mark Baker [cited 10 June 2020]. Available from: <https://www.coronahack.co.uk/>.
 19. Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 2020, 323: 1843–1844.

Reviewer information *Life Research* thanks Besmira Zama and the other anonymous reviewer(s) for the contribution to the peer review of this paper.