proceedings

MDPI

Proceedings

Study on Relevant Features in COVID-19 PCR Tests †

Plácido L. Vidal [1,2,*], Joaquim de Moura [1,2], Lucía Ramos [1,2], Jorge Novo [1,2] and Marcos Ortega [1,2]

1 Centro de investigación CITIC, Campus de Elviña, Universidade da Coruña, s/n, 15071 A Coruña, Spain; joaquim.demoura@udc.es (J.d.M.); l.ramos@udc.es (L.R.); jnovo@udc.es (J.N.); mortega@udc.es (M.O.)

2 Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, Xubias de Arriba, 84, 15006 A Coruña, Spain

* Correspondence: placido.francisco.lizancos.vidal@udc.es; Tel.: +34-981-16-70-00 (ext. 1330)

† Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Published: 26 August 2020

check for updates

Abstract: In the year 2020, the world suffered the effects of a global pandemic. COVID-19 is a disease that mainly affects the respiratory system of patients, even causing a disproportionate response of the immune system and further spreading the damage to other vital organs. The main means by which health care services detected this viral disease was through the use of Polymerase Chain Reactions (PCRs). These PCRs allow the detection of known chains of the genetic code of the virus in samples of sputum. In this work, we study PCR signal features that allow to automatize the analysis of hundreds of PCRs. The findings obtained from the study have shown these features to be capable of obtaining successful results in the detection of COVID-19 in PCR samples, with only a small fraction of the information extracted by the clinicians for that purpose.

Keywords: polymerase chain reaction; COVID-19; feature analysis

1. Introduction

SARS-CoV-2 is a strain of coronavirus responsible for the global COVID-19 pandemic of 2020. This virus causes a severe acute respiratory syndrome (SARS-CoV-2) and viral pneumonia that may leave lungs severely and irreparably damaged [1,2].

For the detection of this virus, sputum samples are analyzed. In these tests, using RT-qPCR or Reverse Transcription Quantitative Polymerase Chain Reaction, the RNA chains that are present in the obtained samples are transcribed into DNA. These tests use deoxyribonucleotides with fluorescent markers that, as soon as they successfully couple with the reference sequence, emit light. This way, by measuring the fluorescence emitted during each cycle, we can know if the reference gene has been found. The gene to be detected is called "E", common to the family to which COVID-19 belongs to group 2 coronavirus.

2. Methodology

To study these signals, we will analyze different features, in the search for one that best separates positive from negative patients. In this case, we have analyzed the mean, standard deviation and the percentiles 25% and 75% of the PCR signals of the patient (as well as the positive and negative reference signals for the PCR batch of that given sample). To study their separability, we will use a kernel density estimation with a gaussian kernel over a trained support vector machine (SVM) with a lineal kernel.

3. Results

Our dataset is composed by 65 positive and 65 negative patients. These signals have been generated with a LightCycler 480 Real-Time PCR System from Roche, using fluorophores with wavelength of excitation of 465 nm and wavelength of detection of 510 nm over 45 PCR cycles. In Figure 1, we can see the mean relative fluorescence returned by the positive PCRs and the negative PCRs. This figure clearly shows the general condition that separates a positive sample from a negative one: the excitation of the signal that is obtained after a certain number of PCR cycles. However, this slope and cycle threshold is highly dependent on the batch (thus the need for the reference signals).

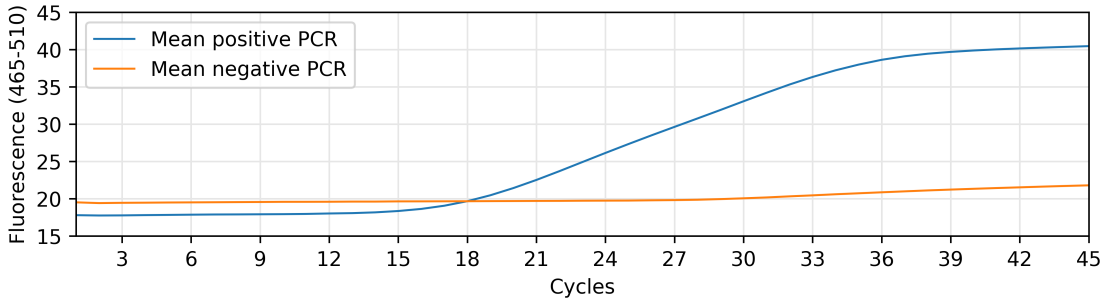

Figure 1. Mean PCR fluorescence signals per cycle of all positive and negative patients of the dataset.

Nonetheless, the best results were obtained by using only the standard deviation of the main signal, with an F1 score of 0.94. As shown in Figure 2, we can satisfactorily classify the majority of the patients without the necessity of the reference signals of the PCRs and with a minimal overlap between classes.

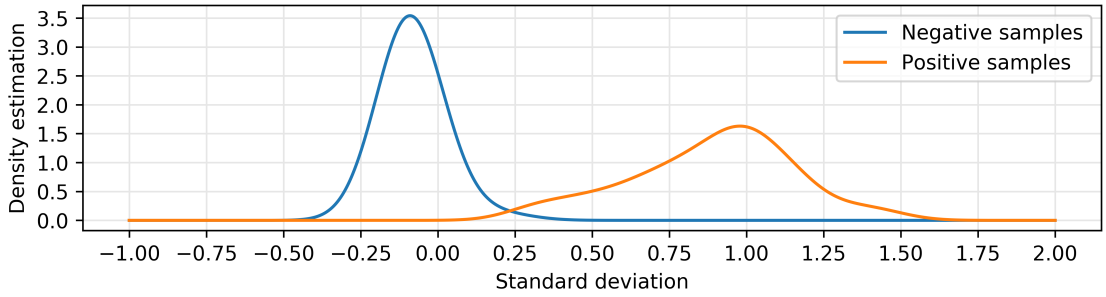

Figure 2. Resulting density estimation for the trained classifier using a kernel bandwith of 0.1.

As future work, we plan a more in-depth study with a larger dataset. This will allow us to generate more complex features, as well as using modern machine learning and statistical strategies to develop a robust system that can, effectively, further speed up the process of diagnosing this disease.

Author Contributions: All the authors contributed equally to this work and have read and agreed to the published version of the manuscript.

Funding: Instituto de Salud Carlos III, Government of Spain, DTS18/00136 research project; Ministerio de Ciencia, Innovación y Universidades, Government of Spain, RTI2018-095894-B-I00 research project, Ayudas para la formación de profesorado universitario (FPU), grant ref. FPU18/02271; CITIC, Centro de Investigación de Galicia ref. ED431G 2019/01, receives financial support from Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, through the ERDF (80%) and Secretaría Xeral de Universidades (20%).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. De Moura, J.; Novo, J.; Ortega, M. Fully automatic deep convolutional approaches for the analysis of Covid-19 using chest X-ray images. *medRxiv* **2020**, doi:10.1101/2020.05.01.20087254.
2. De Moura, J.; Ramos, L.; Vidal, P. L.; Cruz, M.; Abelairas, L.; Castro, E.; Novo, J.; Ortega, M. Deep convolutional approaches for the analysis of Covid-19 using chest X-Ray images from portable devices. *medRxiv* **2020**, doi:10.1101/2020.06.18.20134593.