

Covid-19 Diagnosis Using DNA Sequences of a Hybrid Method based on Machine Learning to Identify Biomarkers

Ajay Yadav*

Department of Biochemistry, University of Mansoura, Egypt

Corresponding author: Ajay Yadav

Abstract

Even if some people don't have any chronic illnesses or don't fall within the Covid-19 risk age range, they are more susceptible to the coronavirus. Some specialists believe that the patient's immune system is to blame, while others believe that the patient's genetic background may be a factor. To ascertain the connection between Covid-19 and genes, it is crucial to identify corona from DNA signals as early as feasible. As a result, it will be possible to determine how changes in the corona disease-related genes affect the disease's severe course. This study proposes a revolutionary intelligent computer method for the first time to distinguish coronavirus from nucleotide signals. The suggested approach offers a multi-layered Singular Value Decomposition, Discrete Wavelet Transform statistical feature extractor, and Entropy-based mapping approach are combined to create a feature extraction framework to extract the most potent features. The Relief approach then chooses distinguishing characteristics. The classifiers used are support vector machine and k closest neighbourhood (k-NN). The technique identified Covid-19 from DNA signals with the greatest classification accuracy rate of 98.84% using an SVM classifier. The suggested technique for determining Covid-19 using RNA or other signals is prepared to be tested with a different database.

Keywords: Covid-19; Big data analysis; Machine learning; Linear algebra; Biomedical signal processing

Received: 02-Feb-2023, Manuscript No. Ipmedt-23-13458; **Editor assigned:** 06-Feb-2023, PreQC No. Ipmedt-23-13458; **Reviewed:** 20-Feb-2023, QC No. Ipmedt-23-13458; **Revised:** 22-Feb-2023, Manuscript No. Ipmedt-23-13458 (R); **Published:** 28-Feb-2023, DOI: 2572-5475-09.01-124

Introduction

Beginning in December, coronavirus was first detected in the Wuhan region of China. It is a contagious virus that spreads from person to person and causes respiratory infections. The World Health Organization has designated the virus as Severe Acute Respiratory Syndrome Coronavirus as its official name. Aiming to define its effects on functionality or pathogenesis, SARS-CoV-2 has some significant modifications in the amino acid sequence despite being generally similar to SARS-CoV [1]. The WHO classified Covid-19 a worldwide health emergency on January 30, 2020. The virus was deemed a global epidemic on March 11, 2020 [2]. The varying number of open reading frames and distinctive spike protein architectures of coronaviruses enable human infection [3]. Sequence evaluation revealed [4]. That the ORFs ORF1a, ORF1b, ORF3a, ORF6, ORF7a, ORF7b, and ORF8

may be used to segment the SARS-CoV-2 genome [5]. Because the N protein surrounds the RNA genome, it has a helical tubular structure. The nucleocapsid E protein, which is encircling this helix, is linked to other structural proteins like the M and S proteins [6]. The SARS-CoV spike protein's surface glycoproteins are crucial for binding to the host receptor [7]. While some people contract the coronavirus sickness and become seriously ill, others simply have minor symptoms [8].

Discussion

SARS-CoV-2 infection of the lower respiratory tract quickly transforms severe survivors' acute respiratory distress syndrome into a condition requiring mechanical oxygen support [9]. Several long-term health issues, including diabetes, hypertension, and Heart failure can be a factor, and genes can affect how people's

✉ AjayYadav76@gmail.com

Department of Biochemistry, University of Mansoura, Egypt

Citation: Yadav A (2021) Covid-19 Diagnosis Using DNA Sequences of a Hybrid Method based on Machine Learning to Identify Biomarkers. Mol Enzy Drug Targ, Vol. 9 No. 1: 124.

bodies respond to viruses [10]. An infection occurs when a human, a bacterium, and the surrounding environment come into contact. Except for the coronavirus vaccine, no reliable technique has been discovered that can stop the pandemic from spreading and totally eliminate the virus. These include increased social distance, antibody testing for infections, and contact tracking to locate and isolate affected individuals. Coronaviruses are classified into the four main genera of alpha, beta, gamma, and delta and have a single-stranded RNA genome spanning from bases. Infection-related consequences are impacted by genomic differences. The sensitivity or resistance to Covid-19 problems can be affected by genomic variables. Finding genetic variables that influence Covid-19 is crucial for pandemic research because the acquired information can be applied to treat patients or prevent diseases before they start. Covid-19 can be predicted from DNA sequences using a variety of detection techniques, including polymerase chain reaction, microarray, and isothermal techniques. These procedures, however, take a lot of time and money to complete and are based on the laboratory setting. The most significant aspect that sets this study apart from others is that it provides a low-cost, machine learning-based method to identify Covid-19 from DNA sequences without the requirement for a lab setting.

Conclusion

There is no study to date that uses DNA nucleotide cues to predict Covid19 infection. This study is the first to generate spectrogram images from DNA nucleotide signals and to use SVD and DWT algorithms to extract features from the images. This research possibly persuades the scientists to apply the suggested technique to various datasets. The following could serve as a summary of this paper's significant contributions: In this investigation, a distinct feature extraction structure was employed, incorporating DWT, SVD, and statistical features. For feature selection, a hybrid model based on the Relief algorithm and a statistical feature extractor was utilised. The association between Covid-19 and genes will be discovered, which will help researchers understand how genes affect how severe the coronavirus sickness is. The suggested method for detecting Covid-19 infection using nucleotide signals achieved a classification accuracy rate of 98.84% with an SVM classifier.

Acknowledgement

None

Conflict of Interest

No conflict of interest.

References

- 1 Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.
- 2 Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, et al. (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* 10: 104742.
- 3 Li F (2013) Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res* 100: 246-254.
- 4 Lu G, Wang Q, Gao GF (2015) Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol* 23: 468-478.
- 5 Ortega JT, Serrano ML, Suárez AI, Baptista J, Pujol FH, et al. (2019) Antiviral activity of flavonoids present in aerial parts of *Marcetia taxifolia* against Hepatitis B virus, Poliovirus, and Herpes simplex virus in vitro. *EXCLI J* 18: 1037-1048.
- 6 Pedretti A, Villa L, Vistoli G (2004) VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. *J Comput Aided Mol Des* 18: 167-173.
- 7 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781-1802.
- 8 Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, et al. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 30: 1771-1173.
- 9 Satija N, Lal SK (2007) The molecular biology of SARS coronavirus. *Ann NY Acad Sci* 1102: 26-38.
- 10 Shang J, Wan Y, Liu C, Yount B, Gully K, et al. (2020) Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS Pathog* 16: e1008392.