The role of AI in genomic tests for detecting COVID-19

Several reviews address the broader applications of AI in the context of COVID-19, and the role of AI in clinical and genomic diagnostics including variant calling, variant classification, genome annotation, and phenotype-to-genotype correspondence (Alimadadi et al., 2020; Dias and Torkamani 2019; Syeda et al, 2021).

Thus far, AI has had limited applications in the specific context of genomic diagnosis of COVID-19 infections, and the various applications, ranging from assay development to taxonomic classification, are summarized in Table 1.

Table 1. Applications of AI in genomic tests for detecting and assessing COVID-19 infections

Diagnostic application	AI methodology	Reference
ASSAY DEVELOPMENT AND MONITORING		
Primer design for accurate detection of SARS-CoV-2.	Convolutional neural network (CNN) composed of one convolutional layer with 12 different filters or weights (each with window size 21, and an even padding of 10 steps on each side) with maxpooling (pool size 148 and stride 1), a fully connected layer (196 rectified linear units with dropout probability 0.5), and a final softmax layer with 5 units, to differentiate the different classes of Coronavirus strains. The optimizer is Adaptive Momentum (ADAM), with learning rate 10^{-5} and a batch size of 50 samples, run for 1000 epochs.	Lopez-Rincon et al 2021
Anomaly Identification during real-time RT-PCR for detecting SARS-CoV-2.	Four different ML algorithms tested [K-neighbor classifier, support vector machine for classification (SVC), decision tree classifier, and random forest classifier (RFC) models] and RFC identified as the best model.	Villareal-Gomez et al 2021
Predict the activity of a probe-target combination in CRISPR-based diagnostics.	Deep convolutional neural network (CNN).	Metsky et al., 2020a; Metsky et al., 2020b
VIRAL CO-INFECTIONS		
Classification of SARS-CoV-2 and co-infecting RNA viruses from RNA sequence data.	PACIFIC, a deep-learning algorithm based on an embedding layer, a convolutional neural network (CNN), and a bi-directional long short-term memory (BiL- STM) network that ends in a fully connected layer.	Mateos et al 2021

DIAGNOSTIC TARGET IDENTIFICATION

Possible diagnostic target identification based on determinants of pathogenicity and zoonotic transmission.	Multiple support vector machines (SVMs) were trained [using the Python library scikit-learn with a linear kernel function] on all 5-nt sliding windows in the identified high- confidence alignment regions, using a cross-validation technique.	Gussow et al 2020

GENETIC RISK SCORE FOR DISEASE SEVERITY

Predicting severity of reaction to COVID-19 infection based on a genetic risk score.	Machine learning algorithm, XGBoost, used to build a classifier based only on their 88-number classification. (88 numbers characterizing the chromosomal-scale length variability of their germ line DNA. Each number represented one quarter of the 22 autosomes). [NOTE: AUC = 0.51, too low for a useful genetic test based on this approach].	Toh and Brody 2020

HOST RESPONSE BIOSIGNATURE

Determining a host response biosignature for SARS-CoV-2 infection based on RNA profiling.	Classifiers developed using scikit-learn (version 1.2.2) in Python. 13 different classifier models, including Linear Support Vector Machine, Linear Discriminant Analysis, and Deep Neural Network, were trained in parallel using a cross-validation approach.	Ng et al 2021

REGIONAL AND TERRITORIAL SEQUENCE DIVERSITY

Assessment of regional and territorial diversity of SARS-Cov-2 RNA sequence.	Support Vector Machine (SVM) (supervised statistical learning and machine learning method for classification analysis) with Top-N (ranking technique or N-best method). The multi-class SVM is defined on binary SVM with one-versus-one max-wins voting strategy or one-versus-all winner-takes-all way.	Liu 2021

TAXONOMIC CLASSIFICATION

Taxonomic classification of whole virus genomes for SARS-CoV-2 that can be used for real time classification of COVID-19 pathogens.	MLDSP-GUI (Machine Learning with Digital Signal Processing with an interactive Graphical User Interface) augmented by a decision tree approach to the supervised machine learning component and a Spearman's rank correlation coefficient analysis for result validation.	Randhawa et al 2020

References

Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19 Physiol Genom 2020;52:20002. https://doi.org/10.1152/physiolgenomics.00029.2020

Das JK, Tradigo G, Veltri P, et al. Data science in unveiling COVID-19 pathogenesis and diagnosis: evolutionary origin to drug repurposing. Brief Bioinform 2021;Feb 17:bbaa420. doi:10.1093/bib/bbaa420. PMID: 33592108.

Dias R, Torkamani A. Artificial intelligenNce in clinical and genomic diagnostics. Genome Med 2019;11:70. https://doi.org/10.1186/s13073-019-0689-8

Gussow AB, Auslander N, Faure G, et al. Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. PNAS 2020;117:15193-9.

Liu J. SARS-Cov-2 RNA sequence classification based on territory information. arXiv:2101.03323 [q-bio.QM].

Lopez-Rincon A, Tonda A, Mendoza-Maldonado L, et al. Classification and specific primer design for accurate detection of SARS-CoV-2 using deep learning. Sci Rep 2021;11:947. doi: 10.1038/s41598-020-80363-5. PMID: 33441822.

Mateos PA, Balboa RF, Easteal S, Eyras E, Patel HR. PACIFIC: a lightweight deep-learning classifier of SARS-CoV-2 and co-infecting RNA viruses. Sci Rep 2021;11:3209. doi: 10.1038/s41598-021-82043-4. PMID: 33547380.

Metsky HC, Welch NL, Haradhvala NJ, Rumker L, Zhang YB, Pillai PP, Yang DK, Ackerman CM, Weller J, Blainey PC, Myhrvold C, Mitzenmacher M, Sabeti PC. Diagnostic design with machine learning model-based optimization. bioRxiv 2020a. doi: https://doi.org/10.1101/2020.11.28.401877.

Metsky HC, Freije CA, Kosovo-Thoroddsen T-SF, Sabeti PC, Myhrvold C. CRISPR-based surveillance for COVID-19 using genomically-comprehensive machine learning design. bioRxiv 2020b. doi: https://doi.org/10.1101/2020.02.26.967026

Ng DL, Granados AC, Santos YA et al. A diagnostic host response biosignature for COVID-19 from RNA profiling of nasal swabs and blood. Sci Adv 2021; Feb:EABE5984.

Randhawa GS, Soltysiak MPM, El Roz H, de Souza CPE, Hill KA, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. PLoS ONE 2020;15(4): e0232391. https://doi.org/10.1371/journal.pone.0232391.

Syeda H, Syed M, Sexton K, et al. Role of machine learning techniques to tackle the COVID-19 crisis: Systematic review. JMIR Med Inform 2021;9:e23811. doi:10.2196/23811

Toh C, Brody JP. Evaluation of a genetic risk score for severity of COVID-19 using human chromosomal-scale length variation. Hum Genomics. 2020 Oct 9;14:36. doi: 10.1186/s40246-020-00288-y.

Villarreal-González R, Acosta-Hoyos AJ, Garzon-Ochoa JA, Galán-Freyle NJ, Amar-Sepúlveda P, Pacheco-Londoño LC. Anomaly identification during polymerase chain reaction for detecting SARS-CoV-2 using artificial intelligence trained from simulated data. molecules. Molecules 2021;26:20. doi: 10.3390/molecules26010020.