World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:17, No:07, 2023

Open Science Index, Biotechnology and Bioengineering Vol:17, No:07, 2023 publications.waset.org/abstracts/167146.pdf

Application of MALDI-MS to Differentiate SARS-CoV-2 and Non-SARS-CoV-2 Symptomatic Infections in the Early and Late Phases of the Pandemic

Authors : Dmitriy Babenko, Sergey Yegorov, Ilya Korshukov, Aidana Sultanbekova, Valentina Barkhanskaya, Tatiana Bashirova, Yerzhan Zhunusov, Yevgeniya Li, Viktoriya Parakhina, Svetlana Kolesnichenko, Yeldar Baiken, Aruzhan Pralieva, Zhibek Zhumadilova, Matthew S. Miller, Gonzalo H. Hortelano, Anar Turmuhambetova, Antonella E. Chesca, Irina Kadyrova

Abstract : Introduction: The rapidly evolving COVID-19 pandemic, along with the re-emergence of pathogens causing acute respiratory infections (ARI), has necessitated the development of novel diagnostic tools to differentiate various causes of ARI. MALDI-MS, due to its wide usage and affordability, has been proposed as a potential instrument for diagnosing SARS-CoV-2 versus non-SARS-CoV-2 ARI. The aim of this study was to investigate the potential of MALDI-MS in conjunction with a machine learning model to accurately distinguish between symptomatic infections caused by SARS-CoV-2 and non-SARS-CoV-2 during both the early and later phases of the pandemic. Furthermore, this study aimed to analyze mass spectrometry (MS) data obtained from nasal swabs of healthy individuals. Methods: We gathered mass spectra from 252 samples, comprising 108 SARS-CoV-2-positive samples obtained in 2020 (Covid 2020), 7 SARS-CoV- 2-positive samples obtained in 2023 (Covid 2023), 71 samples from symptomatic individuals without SARS-CoV-2 (Control non-Covid ARVI), and 66 samples from healthy individuals (Control healthy). All the samples were subjected to RT-PCR testing. For data analysis, we employed the caret R package to train and test seven machine-learning algorithms: C5.0, KNN, NB, RF, SVM-L, SVM-R, and XGBoost. We conducted a training process using a five-fold (outer) nested repeated (five times) ten-fold (inner) cross-validation with a randomized stratified splitting approach. Results: In this study, we utilized the Covid 2020 dataset as a case group and the non-Covid ARVI dataset as a control group to train and test various machine learning (ML) models. Among these models, XGBoost and SVM-R demonstrated the highest performance, with accuracy values of 0.97 [0.93, 0.97] and 0.95 [0.95; 0.97], specificity values of 0.86 [0.71; 0.93] and 0.86 [0.79; 0.87], and sensitivity values of 0.984 [0.984; 1.000] and 1.000 [0.968; 1.000], respectively. When examining the Covid 2023 dataset, the Naive Bayes model achieved the highest classification accuracy of 43%, while XGBoost and SVM-R achieved accuracies of 14%. For the healthy control dataset, the accuracy of the models ranged from 0.27 [0.24; 0.32] for k-nearest neighbors to 0.44 [0.41; 0.45] for the Support Vector Machine with a radial basis function kernel. Conclusion: Therefore, ML models trained on MALDI MS of nasopharyngeal swabs obtained from patients with Covid during the initial phase of the pandemic, as well as symptomatic non-Covid individuals, showed excellent classification performance, which aligns with the results of previous studies. However, when applied to swabs from healthy individuals and a limited sample of patients with Covid in the late phase of the pandemic, ML models exhibited lower classification accuracy.

Keywords : SARS-CoV-2, MALDI-TOF MS, ML models, nasopharyngeal swabs, classification

Conference Title : ICBCBBE 2023 : International Conference on Bioinformatics, Computational Biology and Biomedical Engineering

Conference Location : Zurich, Switzerland

Conference Dates : July 24-25, 2023