[BIOINFO 2024]

Integrative analysis of SARS-CoV-2 variants and clinical features for predicting disease progression using neural networks

Soojung Ha[1], Juho Park[1], and Kyuri Jo[1,*]

[1] *Department of Computer Engineering, Chungbuk National University*
**Corresponding author: kyurijo@chungbuk.ac.kr*

The COVID-19 pandemic has significantly impacted the global spread and pathological evolution of the virus, resulting in the emergence of diverse viral variants. Analysis of these variants using sequencing data from COVID-19 patients is crucial for understanding key pathological features such as disease severity and clinical progression. In this study, SARS-CoV-2 genome sequences from 74 COVID-19 patients are analyzed to identify mutation patterns and genotypes from variants. Furthermore, a neural network-based classification model is developed to discriminate patient status such as deterioration and recovery phase, by integrating various features from the SARS-CoV-2 genome, gene expression data, and clinical information of patients. Multidimensional scaling (MDS) technique is used to generate input features from the SARS-CoV-2 genome and discover key clinical features that have correlation with mutation patterns. By combining RNA sequencing data, clinical characteristics and machine learning techniques, this study aims not only to elucidate the characteristics of COVID-19 variants, but also to contribute to the development of robust predictive models for clinical outcomes.