

Abstract

Human populations are constantly inundated with viruses, some of which are responsible for various deadly diseases. Molecular biology approaches have been employed extensively to identify pathogenic viruses despite the limitations of the approaches. Nevertheless, recent advances in the next generation sequencing technologies have led to a surge in viral genome sequence databases with potentials for Bioinformatics based virus identification. In this study, we have utilised the Gaussian radial basis function neural network to identify pathogenic viruses. To validate the neural network model, samples of sequences of four different pathogenic viruses were extracted from the ViPR corpus. Electron-ion interaction pseudopotential scheme was used to encode the extracted sample sequences while cepstral analysis technique was applied to the encoded sequences to obtain a new set of genomic features, here called Genomic Cepstral Coefficients (GCCs). Experiments were performed to determine the potency of the GCCs to discriminate between different pathogenic viruses. Results show that GCCs are highly discriminating and gave good results when applied to identify some selected pathogenic viruses.

Keywords

Cepstral Dengue Ebolavirus Electron Enterovirus Genomics Hepatitis Radial Neural Network