ML-Based Pathogen Identification for COVID-19 Using Genomic Signatures

Anoushka Bhat, Rishov Chatterjee, Srisairam Achuthan Ph.D.

Center for Informatics

Outline

- Introduction
- Key goals of the project
- Solution overview
- Project plan and progress
- Ownership and engagement

Introduction

- When a new virus inflicts society, the first thing researchers do is find out what kind of virus it is.

- Classifying a virus is very much dependent on the genetic makeup of the virus.

- Machine learning gives us the ability to classify the various types of viruses based on patterns in their genetic makeup.

Key goals of the project

- Review techniques used in Randhawa's [PLOS paper](#)

- Pre-process and clean viral genome data.

- Construct several machine learning models to classify taxonomy of a virus's genetic sequence.

- Construct a machine learning model validation strategy.

Solution overview

- Retrieve viral genome sequences from data source provided in the PLOS paper.

- Pre-process the data and engineer new features from discussion with lab.

- Construct several machine learning models to classify viral taxonomy.

- Predict the taxonomy of the new COVID-19 viral genome data from latest experiments and discoveries.

- Create a more thorough analysis by introducing explainable AI metrics.

Project plan and progress

- GitHub repository will be used for version control.
- Notifications of updates will be sent to a Discord channel for updates.
- 1 weekly progress call will be conducted via Zoom.
- Project will be structured into parts via GitHub Project Boards.
- Lab Meeting will take place via Zoom for presenting and communicating the stages of your work.

Ownership and engagement

- Any work done for this project belongs to City of Hope.

- By agreeing to work on this project during your internship, you acknowledge that you must not share the details of your project with anyone outside of City of Hope.

- Any desire to share must receive written consent from Dr. Srisairam Achuthan and Rishov Chatterjee.

- Any violation of the above will result in immediate termination of the internship.

- Your agreement to this project and rules will be recorded via audio.