Predicting the Likelihood of a Patient Developing Mild or Severe COVID-19 Symptoms using Machine Learning

Mitchell, Michael (School: SOS-Hermann Gmeiner International College)
Anaglate, Senam (School: SOS-Hermann Gmeiner International College)

The 2019 coronavirus has proven to have a devastating global effect. With symptoms ranging from those of the common cold to pneumonia, it is evident that scientists must decipher the point at which a case becomes severe to swiftly develop an apt treatment plan. To meet this need, machine learning and RNA sequenced cells were utilized to develop a model that predicts the likelihood of cells developing a mild or severe case. Firstly, severe and mild immune cell data from the lungs and blood were imported, filtered, transformed, and integrated into training and testing sets. After, the data sets were passed through dimensionality reduction techniques - Principal Component Analysis and UMAP - to improve visualization of features. They were then run through a clustering algorithm. When running the classifier, the number of principal components (npcs) values was varied to increase the accuracy of the predictions. Furthermore, the cell data was analyzed to better understand the body's response to the virus. The optimum npcs value for the classifier was 8 with 73% and 88% accuracy for mild and severe cases respectively in the lungs, and 63% and 53% for mild and severe cases respectively, in the blood. Further analysis showed that the majority of the immune response took place in the lungs. The study revealed that severe cases are caused by high levels of inflammatory cytokines in the lungs, machine learning is an effective tool in diagnosing the virus, and lung cells are the most appropriate when running this classifier.