Predicting COVID-19 Infection Severity Using CNN on Single Cell RNA-seq Data

Leonardo Falcon[1], Wenjian Huang[1], Karla Paniagua[1], Ricardo Ramirez[2], Mario Flores[1], Yu-Fang Jin[1]

[1]Department of Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, TX, [2]Department of Engineering, Houston Baptist University, Houston, TX
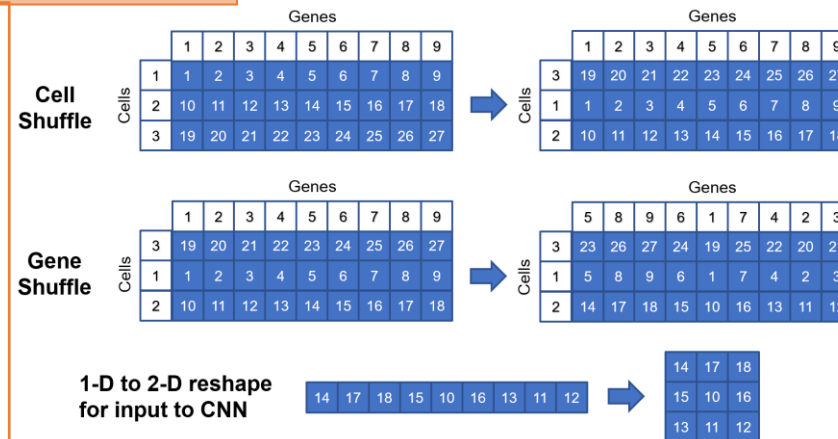
Introduction

Coronavirus Disease 19 (COVID-19) infection has demonstrated a highly variable severity among the infected patients. Some patients experience mild to no symptoms while others come down with severe illness, leading to one of the research problems to be investigated for COVID-19 treatment. The goal of this study is to uncover underlying regulatory mechanisms for different severities of COVID-19 infection using single-cell RNA sequencing (scRNA-Seq) data and deep learning algorithms. In particular, scRNA-Seq data collected from severe and mild cases were compared to uninfected cases (90,696 total cells) using convolutional neural network (CNN) approaches.

Method

Step 1: Raw Data Preparation

Filters:
- Cells with > 200 and < 6,000 genes
- Cells with UMI count (transcripts) > 1,000
- Cells with mitochondrial gene percentage < 10%
- Highly variable genes (regarding infection severity)

Filtered Dataset: 21,939 cells from uninfected healthy control, 7,316 cells from patients with mild symptoms, and 37,197 cells from patients with severe symptoms.
Total: 66,452 cells and expressions of 2,000 genes.

Step 2: Establishment of Embeddings

Goal: Embed 1-D gene expression data into an image.

Random Embedding: Randomly positioning a cell's genes into an image.

Chromosomal Order: Positioning genes into an image by their chromosomal order.
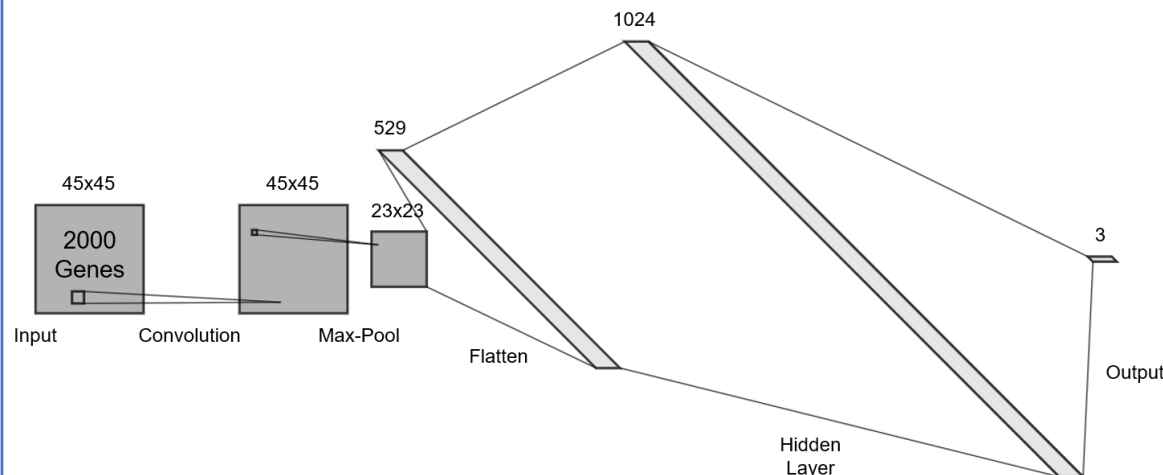

1-D to 2-D reshape for input to CNN

Step 3: Configurations of CNN Model

Structure: One convolutional layer; 1,024 node hidden layer; three node fully connected output layer.
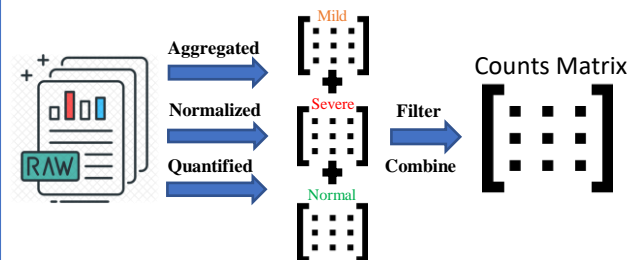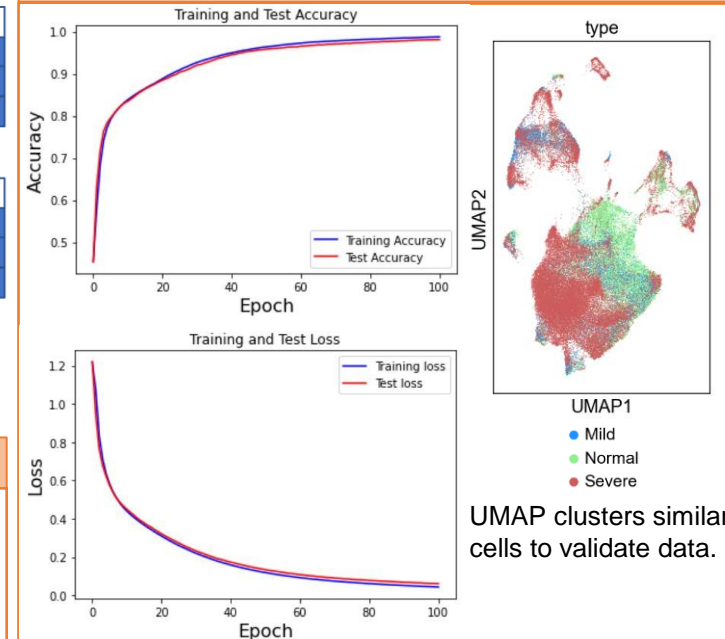
Input: 45x45 image representing the 2,000 gene expression levels of a cell.

Output: Three nodes representing the classification of uninfected, mild, or severe infection.

Results

UMAP clusters similar cells to validate data.

The CNN models predicted cells from patients with uninfected, mild, and severe symptoms with an accuracy of 98-99%.

Conclusion

CNN models have been developed to classify the severity of Covid-19 infection. The learned CNN features in the hidden layer can be extracted to identify the leading genes and functional modules for different severities of COVID-19 infection.

References

1. Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Adv Neural Inform Proc Syst. (2016) 3844–52.
2. Ramirez R, Chiu Y-C, Hererra A, Mostavi M, Ramirez J, Chen Y, Huang Y and Jin Y-F (2020) Classification of Cancer Types Using Graph Convolutional Neural Networks. Front. Phys. 8:203. doi: 10.3389/fphy.2020.00203