Name: Jolene Branch

Date: August 1, 2021

Class: DSC680-T301-2217

Assignment: 8.2 Project 3: Proposal and Data Selection – Milestone 1

This project falls under the image recognition and healthcare domains. (References with annotations at end).

This unbalanced binary classification project will use a Kaggle competition dataset (<https://www.kaggle.com/wobotintelligence/face-mask-detection-dataset> ) and convnet concepts from the book 'Deep Learning With Python' by Francois Chollet (2018). This dataset consists of over 10,000 files in .png, .jpg, .jpeg, and other formats. Interestingly, it has more than two classes of images, which will require some sort of subsetted dataset to be used for the purposes of this project.

I propose to analyze this dataset using Python and concepts learned during the course of this program, then comparing my results to see if my model classifies better than any of the Kaggle competition entries (that received at least one upvote!) I also want to see if the widespread wearing of face masks causes known, established image detection programs, such as the Viola Jones Algorithm, to fail. (Mujtaba, 2020). What level of accuracy can I achieve? Is there any literature on human visual accuracy (such as at screening checkpoints) for comparison?

Unlike the previous two projects, this one should not require extensive data cleaning. I will not be photo-shopping images! I will import the data into Python. Due to the size of the dataset, this involves a zipped folder, which will need to be uncompressed at some point. I will attempt to use the Bellevue University VMWare due to the potential size limitations of my personal computer. I will split the images into test, train, and validation datasets and attempt to train a convnet from scratch, as I read about in class. I will use Keras ImageDataGenerator to read images from directories. As an employee in a healthcare facility with a ‘campus’ style layout, the cost of staffing a mask/symptom screener at the main entrance to each building during COVID-19 is staggering. Yet one person sitting in front of a bank of monitors would not be able to ‘eyeball’ each person entering every building, especially during peak change of shift times. Any progress toward automation of the screening process could have immediate financial benefits.

I anticipate challenges in handling the size of the dataset and in properly loading it. I have already spent three days working on that and am not where I need to be. It seems that at each step I am working through error handling issues. I am skeptical of the ability to generate interesting graphics from this project beyond training and validation accuracy/loss plots.

References:

Asrani, P., Hasan, G. M., Sohal, S. S., & Hassan, M. D. I. (2020, November 24). *Molecular Basis of Pathogenesis of Coronaviruses: A Comparative Genomics Approach to Planetary Health to Prevent Zoonotic Outbreaks in the 21st Century*. Omics : a journal of integrative biology. <https://pubmed.ncbi.nlm.nih.gov/32940573/>

-*This is the third coronavirus outbreak so far in the 21st century. The reference to comparative genomics sounds to me to be something like multiple regression analysis of human sub-populations.*

Bernstein, B. (2021, April 17). *COVID-19 Death Toll Surpasses 3 Million Worldwide*. National Review. <https://www.nationalreview.com/news/covid-19-death-toll-surpasses-3-million-worldwide/>

D'Agostino, A., Demartini, B., Cavallotti, S., & Gambini, O. (2020, May). *Mental health services in Italy during the COVID-19 outbreak*. The Lancet: Psychiatry. <https://doi.org/10.1016/S2215-0366(20)30133-4>

Dong, E., Du, H., & Gardner, L. (2020, February 19). *An interactive web-based dashboard to track COVID-19 in real time*. Redirecting. <https://doi.org/10.1016/S1473-3099(20)30120-1>

Elrashdy, F., Redwan, E. M., & Uversky, V. N. (2020, September 11). *Why COVID-19 Transmission Is More Efficient and Aggressive Than Viral Transmission in Previous Coronavirus Epidemics?* Biomolecules. <https://pubmed.ncbi.nlm.nih.gov/32933047/>

-*This is why wearing face masks is so important when trying to reduce the effects of this virus.*

El-Sayed, A., Abdel-Daim, M. M., & Kamel, M. (2021, May 5). *Causes of respiratory failure in COVID-19 patients*. Environmental science and pollution research international. <https://pubmed.ncbi.nlm.nih.gov/33954917/>

Liu, Z., Xiao, X., Wei, X., Li, J., Tan, H., Zhu, J., Zhang, Q., Wu, J., & Liu, L. (2020, March 11). *Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2*. Journal of medical virology. <https://pubmed.ncbi.nlm.nih.gov/32100877/>

Mujtaba, H. (2020, November 30). *How to detect face recognition Using Viola Jones Algorithm*. GreatLearning Blog: Free Resources what Matters to shape your Career! <https://www.mygreatlearning.com/blog/viola-jones-algorithm>

*-“This algorithm is painfully slow to train but can detect faces in real-time with impressive speed.”*

Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020, March 28). *COVID-19, SARS and MERS: are they closely related?* Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. <https://pubmed.ncbi.nlm.nih.gov/32234451/>

Singh, R., Kang, A., Luo, X., Jeyanathan, M., Gillgrass, A., Afkhami, S., & Xing, Z. (2021, March). *COVID-19: Current knowledge in clinical features, immunological responses, and vaccine development*. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. <https://pubmed.ncbi.nlm.nih.gov/33577115/>

Yang, Y., Rao, X., & Shang, W. (2020, June). *Facing the COVID-19 outbreak: What should we know and what could we do?* Journal of medical virology. <https://pubmed.ncbi.nlm.nih.gov/32091134/>

-*Eighteen other articles cited this refereed journal article, so I thought I should, too.*