eVision

George Kouretas, Ron Huang, and Prof. Navid Shaghaghi*
BioInnovation and Design Lab
Santa Clara University, Santa Clara, California, USA
{gkouretas, ryhuang, nshaghaghi}@scu.edu

Seasonal influenza (aka the flu) is a serious respiratory infection caused by influenza viruses. Influenza is extremely pervasive, with an estimated 3 to 5 million severe cases annually. During the 2018-19 flu season, influenza vaccines prevented between 3.4 and 7.1 million flu cases. Since vaccines are vital to combating influenza, an accurate prediction model is essential for pharmaceutical companies and healthcare providers to prepare for upcoming flu seasons by producing an adequate supply of vaccines.

Having adequate knowledge of the demand for vaccines is essential. Unfortunately, the major providers of data for influenza and other related diseases have a reporting lag. For example, the Center for Disease Control and Prevention (CDC) has a two-week reporting lag for their reports on national influenza cases. This impedes companies in the pharmaceutical industry from knowing when to increase and decrease the manufacturing of their influenza-related products, such as testing kits and vaccines.

eVision utilizes a long short term memory (LSTM) neural network and Google Trends, CDC, and World Health Organization (WHO) data as input layers to predict influenza outbreak. The model predicts influenza cases geographically, so it is able to predict influenza trends anywhere there is sufficient data for. The model is primarily tested using data from the United States, since it is the most accessible and numerous. eVision can provide influenza case predictions from three to fourteen weeks ahead, with error rates as low as 8.57%. Having access to a highly accurate predictor like eVision can provide the pharmaceutical industry essential knowledge on what to expect about the trend of the flu outbreak during the flu season so they can wramp up or down their production of vaccines, tester kits, and medicine.

eVision is built using a combination of MATLAB and Python. The LSTM model, data processing and formatting algorithms, and graphical user interface are implemented via MATLAB utilizing MATLAB's Deep Learning Toolkit and App Designer. Web scrapers for Google Trends, the CDC, and WHO are built via Python utilizing a variety of packages such as pytrends, selenium, pandas, and more.

There are many methods being utilized in order to optimize the model's accuracy. These include granulating the data to smaller regions in Google Trends, selecting data from more dense areas of regions, and by taking Google Trends data from states that contain ports of entry such as: California, Texas, and New York because ports of entry are places where outbreaks spread rapidly.

* Faculty Advisor