

## Harnessing Big Data and Artificial Intelligence to Predict Pandemic Spread

2020-10-03

Texas A&M civil engineering researchers have developed a powerful deep learning computational model to help predict the future spread of COVID-19 cases at a county level.

Artificial intelligence has been used to enhance diagnostic efforts, deliver medical supplies and even assess risk factors from blood tests during the COVID-19 pandemic. Now, artificial intelligence is being used to forecast future COVID-19 cases.

[Texas A&M University](#) researchers, led by Ali Mostafavi, have developed a powerful deep learning computational model that uses artificial intelligence and existing big data related to population activities and mobility to help predict the future spread of COVID-19 cases at a county level. The researchers published their results in [IEEE Access](#).

The spread of pandemics is influenced by complex relationships related to factors including mobility, population activities and sociodemographic characteristics. However, typical mathematical epidemiological models only account for a small subset of relevant features. In contrast, the deep learning model developed by Mostafavi, associate professor in the Zachry Department of Civil and Environmental Engineering, and his [UrbanResilience.AI lab](#), can explain the complex relationship between a larger number of features to forecast the range of increase in COVID-19 infections in future days.

“We immediately realized the potential for employing artificial intelligence to complement the existing mathematical epidemiological models,” Mostafavi said. “We are living in the era of big data and leveraging these big data during crises is providing great opportunities for the development of models and data-driven tools to inform policies.”

A deep learning model is a subset of machine learning, a type of artificial intelligence, where computing systems, called neural networks, learn from large amounts of data. By training the

deep learning model with data from a certain time period, in this case from March through May 2019, the model identified features to predict the trajectories of another time period — June 2019.

The researchers' deep learning model accounts for features such as the movement of people within a community, census data, social-distancing data, past case count growth and social demographics to predict the growth of COVID-19 cases for each county with 64% accuracy, which is twice the accuracy of an untrained model. The model's greatest accuracy was for seven days into the future. The accuracy decreased the further into the future the model predicted.

"One aspect of modeling that is helpful is not the accuracy, but evaluating what factors drive the outcomes," Mostafavi said. "This model does not identify specific mitigation and response strategies, but it can help at different points in time to see which strategies could be effective based on various county-level features."

Knowing which features of the model have the most significant effect on the increase of cases, officials can develop policies that target those factors. If the most critical feature for a county is mobility, for example, officials can implement policies like stay-at-home orders.

The model can also offer insight into the effectiveness of policies after they have already been in place. Mostafavi discovered that overall, the initial travel reduction orders were effective — people from less populated counties traveled less to higher-populated cities, but the extent of travel in densely populated counties did not change drastically.

He said the influence of features can change over time for one county and vary from county to county. At the onset of the pandemic, the researchers saw travel-related and mobility-related factors were important predictors of cases, but as time went on, they saw other features, such as travel to points of interests and social demographic characteristics, were more important.

The outcome is that pandemic mitigation is complicated, and policies are not one size fits all.

In the future, Mostafavi's lab will use new data sets to develop different types of models. In addition to the current national-scale predictive surveillance model, the team is currently

working on an artificial intelligence-based model for city-scale surveillance to predict cases at the zip-code level. More importantly, they want to predict the factors that influence each zip code so that officials can explore location-specific policies. Mostafavi said instead of closing restaurants in an entire county, officials may close restaurants in only high-risk zip codes.

His research shows big data and artificial intelligence have the potential to play a key role in improving pandemic surveillance, prediction and policy development.

“Significant opportunities exist using these big data and AI to contain the existing pandemic and also better prepare and mitigate the future pandemics,” Mostafavi said.

Read the [original article](#) on Texas A&M University.