COVIDCatcher: Developing a Low-Cost Machine-Learning Based Dashboard for Forecasting COVID-19 Cases

Michael Li, Amador Valley High School, Pleasanton, CA

Abstract

The objective of this research is to develop a low-cost, multimodal, data-driven dashboard powered by machine learning that provides daily forecasts of cumulative COVID-19 counts at the county level. Forecasting COVID-19 cases empowers health-officials to take proactive measures, mitigating potential outbreaks. We determined whether a feedforward artificial neural network (ANN), a temporal convolutional network (TCN), or a multimodal approach could forecast daily COVID-19- cases at the county-level based on prior cases. Model performance was evaluated using mean absolute error (MAE) and mean absolute percent error (MAPE). We found that TCN performed better than the multimodal model (19.08 vs. 26.45 average MAE). We also found that using dimensional-reduction and predicting percent change in cases do not improve model performance. A past-future forecasting method and county-county forecasting method were compared. Past-future had an MAE and MAPE of 19.72 and 0.014%, while county-county had an MAE and MAPE of 38.68 and 0.0076%, demonstrating the viability of such approaches. The novel contributions of our research are three-fold. First, the TCN model outperforms the MAE of models in existing literature (27.82 MAE vs. 49.26, 56.84 and 724.53 MAE) as well as the ensemble model used by the Centers for Disease Control and Prevention (CDC). Secondly, the feasibility and accurateness of the county-county and past-future forecasting methods are proven. Finally, we present our research in an accessible and easy-to-use dashboard which can assist health departments in taking preventive measures.

Objective

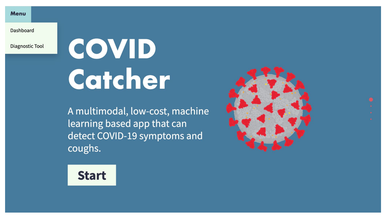
To develop a **low-cost, multimodal**, **data-driven** dashboard powered by machine learning that can forecast **daily cumulative COVID-19 cases** to inform health officials on whether to take preventive measures.

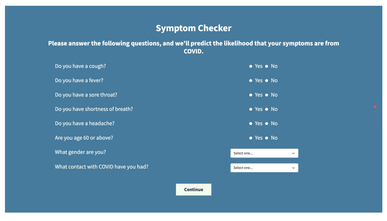
Background

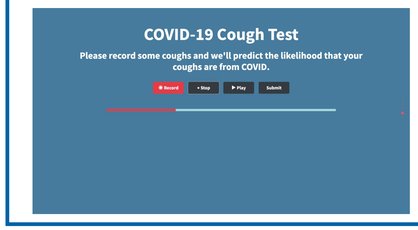
- More than **79 million COVID-19 cases** and **974k deaths** in the U.S.
  - The COVID-19 pandemic has proven to be **enduring** and **turbulent**, coming in **multiple waves** with the rise of **new** COVID-19 **variants**.
- **Essential** tool for public health officials is daily case **monitoring** through the use of COVID- 19 dashboards
- With **relaxed** preventive measures, local health departments must stay **vigilant** and watch for signs of **potential outbreaks**

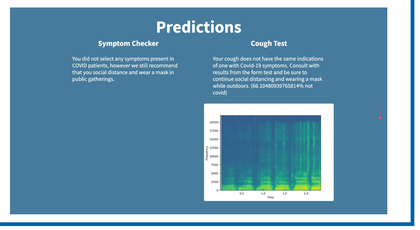
Last Year's Work

- To provide a **low-cost, multimodal** tool for **detecting** COVID-19 symptoms, I created **COVIDCatcher**, a free and accessible website
- COVIDCatcher uses a **health form** and **cough recording**, which are then inputted to an **XGBoost** and **SVC+VGG-19** model
- For this year's research, I am hosting the COVID-19 forecasting dashboard on the same website, accessible through a simple menu









Materials and Methods

Forecasting Techniques

- **County to County:** Forecast the entire COVID-19 case history of a single excluded county using the collective history of all the other counties.
- **Past to Future:** Forecast future COVID-19 cases using the past collective data of all counties.

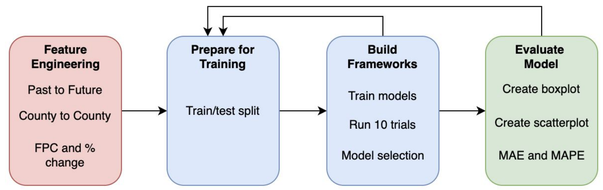
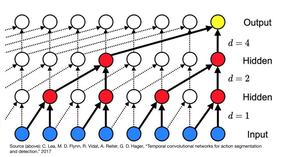
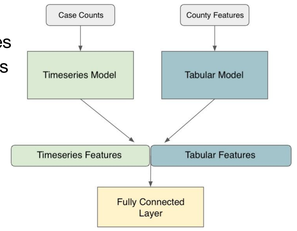


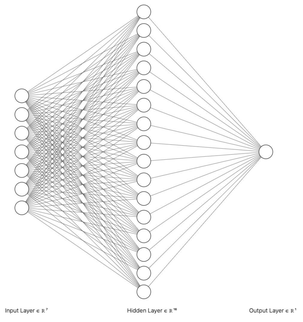
Figure 1. (above) Machine learning model development workflow.

Temporal Convolution Network (TCN). New class of temporal model that employs casual convolutions and dilations.



Multimodal Model. This model incorporates county-level featuristics (population size, gini index), with the TCN model. The concatenated result is outputted to a fully connected layer.





Artificial Neural Network (ANN). This served as our baseline model. It contains a dense layer, dropout layer and output layer.

Results and Interpretations

TCN showed top performance for COVID-19 past-to-future case forecasting

Models	MAE	MAPE
TCN	19.72	0.014%
Multimodal	20.23	0.018%

Table 1. MAE and MAPE for TCN and Multimodal trained on the County to County dataset

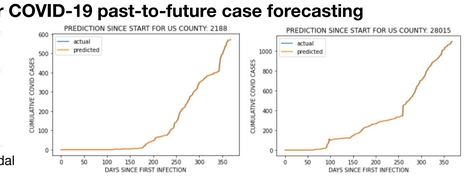


Figure 2. TCN Model Graph Actual Cases v. Predicted Cases **(County to County)**

TCN also achieved high performance for COVID-19 county-to-county case forecasting

Model	MAE	MAPE
TCN	38.68	0.0076%
Multimodal Model	35.55	0.0078%

Table 2. MAE and MAPE for TCN and Multimodal trained on the Past to Future dataset

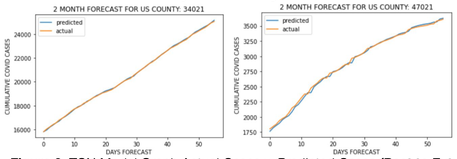


Figure 3. TCN Model Graph Actual Cases v. Predicted Cases **(Past to Future)**

Model comparison between existing literature and CDC ensemble model for forecasts

Model	MAE	MAPE
TCN	55.62	0.016%
ARIMA	629.26	0.099%

Table 3. (left) Performance of TCN in terms of MAE and MAPE errors versus ARIMA.

Model	MAE	MAPE
LSTM	56.84	0.047%
CNN	724.53	0.349%
ANN	49.26	0.035%
TCN	27.82	0.019%
Multimodal	74.23	0.036%

Table 4. (left) Performance of TCN and multimodal model in terms of MAE and MAPE errors versus LSTM, CNN, and ANN.

Model	MAE	MAPE
TCN	94.54	5.02e11%
CDC Ensemble	109.45	1.63e15%

Table 5. (above) MAE and MAPE comparison between TCN and the United States CDC ensemble model, an average of their best performing research models. The TCN has lower MAE and MAPE.

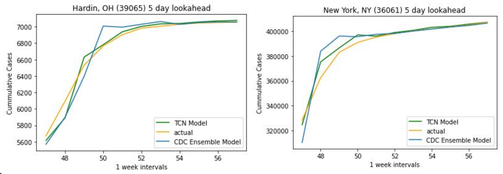


Figure 4. (above) Number of 1 week intervals vs. Cumulative Cases for Hardin County and New York County. TCN shows a finer ability to adapt to the curve than the CDC ensemble model

Results and Interpretations (cont.)

- **COVIDCatcher**'s website and backend were designed to incorporate the top performing model for past-to-future prediction: **TCN** performed the **best** and was successfully deployed
- A survey was conducted to beta-testers to better understand limitations and iterate
  - "This is something that I would ***use every week*** or if I'm ***feeling anxious***"
  - "COVID-Catcher is ***creative*** and ***intuitive*** to use. Saves me money and time, and ***reduces transmission risk*** of me going outside"
  - "I have ***peace of mind*** in checking on my elderly parents with a ***few simple clicks***, without even leaving the house"

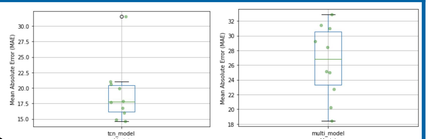
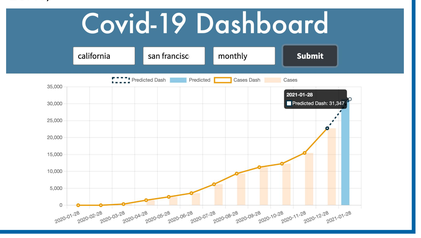


Figure 5. TCN and Multimodal Model Boxplot of MAE across 10 trials. TCN has lower average MAE (19.08 vs. 26.45)



Conclusion and Significance

1. In order to aid local health departments in public health management I developed a **low-cost data-driven** machine learning based dashboard for forecasting COVID-19 cases.
2. COVIDCatcher employs **TCN** to forecast daily cumulative COVID-19 cases at the county-level.
3. County to County and Past to Future are **viable** and **accurate** methods of prediction.
4. **TCN** outperforms both existing models in literature as well as the ensemble model used by the **Centers for Disease Control and Prevention**

Relevant Applications to Biotechnology

1. **An accessible dashboard that is free and scalable for local health departments across the United States:** Due to its low-cost and scalability as a software solution, COVIDCatcher can assist local health officials, who lack access to federal resources, with *no user costs* to manage their public health policies.

2. **A quick and easy-to-use supplement for personal health decision making:** COVIDCatcher is easy to use and can provide ease-of-mind to those worried about potential outbreaks in their area, without any hassle. A simple user interface and quick results in <1 minute ensures anyone can use it efficiently to plan anything from back-to-work policies to family vacations.

3. **Assist doctors and health-policy experts in triaging COVID-19 outbreaks:** As more privacy-approved COVID-19 datasets are collected and released to the public, COVIDCatcher can continue to improve and become useful as a tool to assist doctors and health-policy experts to quickly triage COVID-19 outbreaks.

Acknowledgements

I would like to thank my mentor Professor Zhang for guiding me through this project, Yushan Hu, who preprocessed the datasets, and the following organizations for providing access to their datasets: United States Census Bureau and The Center for Systems Science and Engineering at Johns Hopkins University