# Epidemic Model Guided Machine Learning for **COVID-19 Forecasts**

# Quanquan Gu

- **Department of Computer Science** University of California, Los Angeles
  - August 24, 2020
- KDD 2020 Workshop on Applied data science in Healthcare: Trustable and Actionable AI for Healthcare
  - https://covid19.uclaml.org

# We Launched "Using ML to Combat COVID-19" Project on 3/28/2020 https://covid19.uclaml.org

The novel coronavirus (COVID-19) has emerged as a global pandemic, and the global death toll has reached 100,000 as of April 10, 2020. Currently the data about COVID-19 are overwhelming, yet the use of these data for combating COVID-19 is still in its early stage. The overarching goal of this project is to make good use of these data by machine learning, to better understand the spread of COVID-19, to facilitate informed decisions by policy makers, and to better allocate the medical resources such as hospital/ICU beds, personal protective equipments (PPEs), ventilators, etc.

# World United States

+ 

> Presented by Statistical Machine Learning Lab at UCLA. All rights reserved ©. Contact Us Follow This website is constructed under GNU Affero General Public License v3.0.



# Confirmed Cases							
Dk	200k	400k	600k	800			

# Forecast the number of daily increased confirmed cases and deaths



The projections are based on the actual numbers up to 8/16/20

### https://covid19.uclaml.org

# **Mathematics of Epidemic Models**



Kermack, William Ogilvy, and Anderson G. McKendrick. "A contribution to the mathematical theory of epidemics." Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character 115, no. 772 (1927): 700-721.

Hethcote, Herbert W. "The mathematics of infectious diseases." SIAM review 42, no. 4 (2000): 599-653.

```
4
```



# Limitation of SIR and SEIR

- Do not take into account the unreported cases
  - testing capacity
  - asymptomatic individuals

# Our Model: SuEIR



# $\beta$ : Contact/Infectious rate

 $\mu$ : Discover rate of infected cases

**Epidemic Model Guided Machine Learning for COVID-19 Forecasts in the United States** Difan Zou, Lingxiao Wang, Pan Xu, Jinghui Chen, Weitong Zhang, Quanquan Gu, medRxiv, **doi:** <u>https://doi.org/10.1101/2020.05.24.20111989</u>

$$\frac{dS_t}{dt} = -\frac{\beta(I_t + E_t)S_t}{N}$$
$$\frac{dE_t}{dt} = \frac{\beta(I_t + E_t)S_t}{N} - \frac{dI_t}{N}$$
$$\frac{dI_t}{dt} = \mu\sigma E_t - \gamma I_t$$
$$\frac{dR_t}{dt} = \gamma I_t$$

 $\sigma$ : Incubation rate

## d cases $\gamma$ : Recover rate



# **Machine Learning Pipeline**





# Learning Parameters



Model parameters are learned by minimizing the following loss function

$$L(\beta, \sigma, \mu, \gamma; C_{1:T}, F_{1:T}) = \frac{1}{T} \sum_{t=1}^{T} \left[ \left( \log(\hat{C}_t + 1) - \log(C_t + 1) \right)^2 + \left( \log(\hat{F}_t + 1) - \log(F_t + 1) \right)^2 \right]$$

For state level and county level models, we take the daily reported confirmed and fatality cases from the New York Times (<u>https://github.com/nytimes/covid-19-data</u>) as the input.
For the country level model, i.e., the model for the US prediction, we take the daily reported confirmed and fatality cases from the JHU CSSE (<u>https://github.com/CSSEGISandData/COVID-19</u>).

# SuEIR with Hospitalization



For state level and country level models, hospitalization data are provided in (https://covidtracking.com/) For county level model, hospitalization data are provided in (https://data.ca.gov/dataset)



# Our Model Is Used by CDC Forecasts (Deaths)

CONC Centers for Disease Control and Prevention 2.24/7 Saving Lives, Protecting People"

	A-2.0	12 19 2 8
Search	COVID-19	۹

Advanced Search 🛞

Health Depts 🕞

### Coronavirus Disease 2019 (COVID-19)

Community, Work & School 🗠 Healthcare Workers & Labs 🕞 Your Health 🕑

Cases & Data 🗸

### More ~

🏦 Cases, Data & Surveillance

### CASES, DATA & SURVEILLANCE

### COVID-19 Forecasts: Deaths

Updated Aug. 21, Print 2020

6 🖸 🕲 🚱

Summary Cases & Deaths in

COVIDView Weekly

the US

Testing Data in the

US

Hospitalization Surveillance Network COVID-NET

Serology (Antibody) + Surveillance

Wastewater Surveillance

### Mathematical Modeling

COVID-19 Forecasts: Cases

Forecasts of Total

Deaths

Previous Forecasts of Total Deaths

Hospitalization Forecasts

Previous Hospitalization

Forecasts

COVID-19



Observed and forecasted new and total reported COVID-19 deaths as of August 17, 2020.

### Interpretation of Forecasts of New and Total Deaths

- This week CDC received forecasts of national COVID-19 deaths over the next 4 weeks from 33 modeling groups. Of the 33 groups, 31 provided Forecast Assumptions forecasts for both new and total deaths and two provided forecasts for total deaths only.
- On This Page National Forecast State Forecasts
- This week's national ensemble forecast predicts that 3,700 to 9,600 new COVID-19 deaths will be reported during the week ending September 12. and that 187,000 to 205,000 total COVID-19 deaths will be reported by that date.
- State- and territory-level ensemble forecasts predict that the number of reported new deaths per week will likely increase over the next four weeks in Minnesota and may decrease in 13 jurisdictions. Those with the greatest likelihood of a decrease over the next four weeks include Arizona, Florida, Mississippi, and South Carolina.

### National Forecast



https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html 10

# National Forecast



# Our Model Is Used by CDC Forecasts (Hospitalizations)



New Hospitalizations

https://www.cdc.gov/coronavirus/2019-ncov/covid-data/hospitalizations-forecasts

### National Forecast



# 11

# Our Model Is Used by California COVID Assessment Tool (CalCAT)

California COVID Assessment Tool Introduction Nowcasts Forecasts Scenarios

### Current R-effective in California

The effective reproduction number (R-effective) is the average number of people each infected person will pass the virus onto and represents the rate at which COVID-19 is spreading.

Statewide Estimates of R-effective



Latest Estimate of R-effective is:



### https://calcat.covid19.ca.gov/cacovidmodels/

Technical Notes

About CalCAT

# Our Model Is Used by California COVID Assessment Tool (Forecast)

California County Hospitalization Forecasts

Select a county to see how modeled number of hospitalizations compare with actual numbers to date and with the number of licensed hospital beds (black box).



### Current Total Deaths:



### Statewide Total Death Forecast

The black box (left) represents the current number of total COVID deaths in California. The blue box represents the COVID hub (Reich Lab) forecasted number of total deaths at the 30 day mark based on models for California.



https://calcat.covid19.ca.gov/cacovidmodels/

# Model Performance Comparison (before Resurgence)

Region: US

✓ Week: Jun. 16, 2020 - Jun. 22, 2020 ✓

Comparison on the predicted number of deaths by different models

	2020-06-16	2020-06-17	2020-06-18	2020-06-19	2020-06-20	2020-06-21	2020-06-22	Rank
Reported Number	116963	117717	118432	119124	119720	119979	120402	
Columbia	117716 (+753)	118478 (+761)	119247 (+815)	120030 (+906)	120840 (+1120)	121640 (+1661)	122485 (+2083)	6
IHME	NaN (NaN)	NaN (NaN)	NaN (NaN)	NaN (NaN)	119383 (-337)	NaN (NaN)	NaN (NaN)	8
LANL	118188 (+1225)	119069 (+1352)	119935 (+1503)	120804 (+1680)	121626 (+1906)	122427 (+2448)	123214 (+2812)	7
MIT	117357 (+394)	118154 (+437)	118937 (+505)	119703 (+579)	120461 (+741)	121207 (+1228)	121943 (+1541)	5
MOBS	117471 (+508)	118126 (+409)	118769 (+337)	119401 (+277)	120023 (+303)	120635 (+656)	121237 (+835)	4
UCLA	116952 (-11)	117696 (-21)	118414 (-18)	119024 (-100)	119527 (-193)	120003 (+24)	120438 (+36)	1
UT	117077 (+114)	117803 (+86)	118506 (+74)	119208 (+84)	119914 (+194)	120594 (+615)	121265 (+863)	3
YYG	117133 (+170)	117826 (+109)	118514 (+82)	119197 (+73)	119877 (+157)	120555 (+576)	121230 (+828)	2

### https://covid19.uclaml.org/compare.html

# Model Performance Comparison (after Resurgence)

Region: US

✓ Week: Aug. 11, 2020 - Aug. 17, 2020 ✓

Comparison on the predicted number of deaths by different models

	2020-08-15(pred. 4 weeks ago)	2020-08-15(pred. 3 weeks ago)	2020-08-15(pred. 2 weeks ago)	2020-08-15(pred. 1 weeks ago)	Rank
Reported Number	169481	169481	169481	169481	
Columbia	158517 (-10964)	172209 (+2728)	170467 (+986)	168938 (-543)	5
IHME	165928 (-3553)	165205 (-4276)	166198 (-3283)	167771 (-1710)	6
LANL	156451 (-13030)	159814 (-9667)	164636 (-4845)	167959 (-1522)	7
MIT	166110 (-3371)	166701 (-2780)	166385 (-3096)	167947 (-1534)	2
MOBS	175867 (+6386)	168995 (-486)	173626 (+4145)	168990 (-491)	2
UCLA	160300 (-9181)	166420 (-3061)	170872 (+1391)	169279 (-202)	2
UT	178684 (+9203)	179600 (+10119)	177936 (+8455)	NaN (NaN)	8
YYG	163913 (-5568)	166145 (-3336)	168588 (-893)	169401 (-80)	1

https://covid19.uclaml.org/compare.html

# Acknowledgement

- Collaborators:
  - Prof. Nicolas Reich at the UMass Amherst
  - Prof. Wei Wang, Prof. Eran Halperin at UCLA
  - Dr. Susan Cheng, Cedars-Sinai Medical Center
  - Velma Lopez at CDC
  - John Pugliese, Ryan Mccorvie, David Crow at CDPH
- Team Members:
  - Pan Xu, Jinghui Chen, Lingxiao Wang, Difan Zou, Weitong Zhang, Jenny Zhan, Jeniffer Zhang, Nathan Midkiff, Emily Nguyen
- Google Cloud:
  - Google Maps Platform Crisis Response



azie mahalo danku kiitos danke thankyou hvala tak dékuji mer acies ありがとう хвала tack спа ဆြေးacias 감사합니다 grazie 潮湖 d Idies عنكراً dakujem obrigado ဆि s danku عنكراً ala tak 감사합니다 dēkuli merci iš kvou merci kiltos grācies хва асибо tack gracias THANKYOU dēkuji thankyou ihalo danku asante THANKYOU dēkuji thankyou

### 17