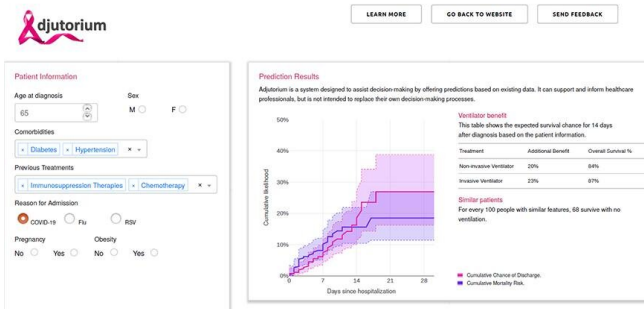


# Using COVID-19 patient data to train machine learning models for healthcare

6 April 2020, by Professor Mihaela Van Der Schaar



Credit: University of Cambridge

One short week ago, I called on governments to use existing data and proven machine learning and AI techniques to help healthcare systems combat the COVID-19 pandemic.

The response was amazing. My team and I received encouragement, ideas, and proposals for collaboration.

We also received, courtesy of Public Health England, a set of (depersonalized) data on existing COVID-19 cases. Along with my team at the Cambridge Centre for AI in Medicine, I've spent the last few days training our models on this data. The results so far are extremely encouraging.

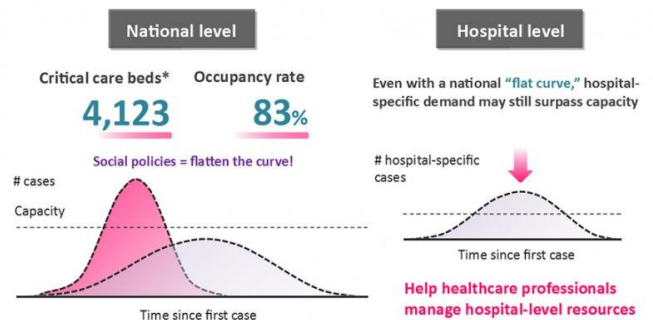
Among other things, we wanted to demonstrate that machine learning techniques can accurately predict how COVID-19 will impact resource needs (ventilators, ICU beds, etc.) at the individual patient level and the hospital level, thereby giving a reliable picture of future resource usage and enabling healthcare professionals to make well-informed decisions about how these scarce resources can be used to achieve the maximum benefit.

Based on the data we received from Public Health

England, we now have a proof-of-concept demonstrator showing that this can be done, in the form of a new system we call Adjutorium.

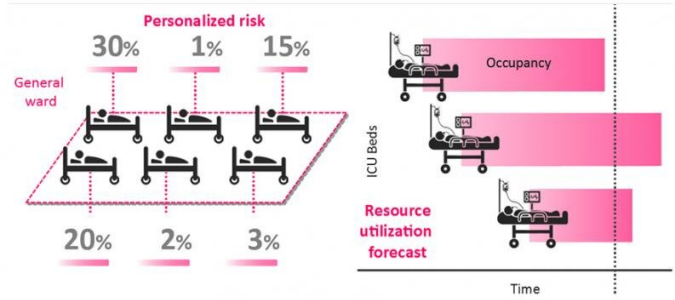
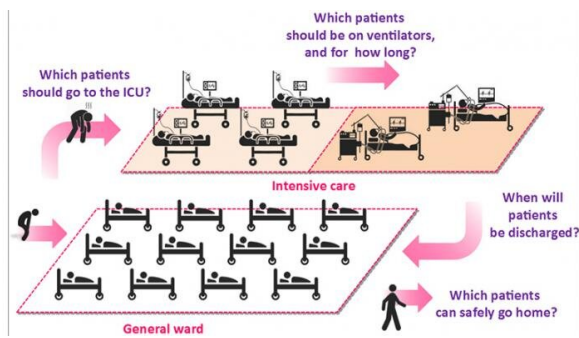
## Isn't flattening the curve enough?

Social policies can certainly help take the strain off healthcare systems around the world. But there's no guarantee that certain individual hospitals won't still be stretched well beyond capacity. Additionally, these measures may not be properly observed by everyone, or may be relaxed slowly over time. It's important to ensure that hospitals remain armed with information that will help them manage peaks in demand for resources like ICU beds or ventilators.



As I touched upon last week, life-or-death choices will be made regarding the use of scarce resources like ventilators and ICU beds. If you are managing or working in a hospital, it would be incredibly helpful (but it's currently not possible) to have a highly reliable picture of the likely usage status of these resources over time.

This is what too many healthcare professionals around the world are currently worrying about:



We can help answer these questions by being smart about how we use existing data on hospital admissions, ICU admissions, use of ventilators, [patient outcomes](#) (e.g. discharge, mortality), and more. If we have access to high-quality datasets containing such information, we can use machine learning to answer questions such as:

- Which patients are most likely to need ventilators within a week?
- How many free ICU beds is this hospital likely to have in a week?
- Which of these two patients will get the most benefit from going on a ventilator today?

While these questions can reliably be answered using the machine learning techniques we've developed, I cannot emphasise enough that the decisions themselves will, of course, still be made by healthcare professionals on the basis of their organisation's priorities and policies.

Here's how a machine learning model can help answer questions in a way that's useful to healthcare professions:

As you can see, patients are given risk scores based on their likelihood of ICU admission or ventilator usage. These are then aggregated across the hospital to give a picture of future demand on resources.

### Using Public Health England data

Last week, I shared a firm belief that existing and proven machine learning techniques can already tackle these kinds of challenges and can deliver essential insights, even using existing (possibly quite noisy) data sources. Thanks to the data we received from Public Health England, I feel more confident than ever.

We received data for nearly 1,700 patients, and that number continues to increase because the dataset is updated daily. While the data was depersonalized, it includes basic information, lab results, hospitalization details, risk factors and outcomes.

We fed this data to AutoPrognosis, a state-of-the-art automated machine learning framework that our team developed in 2018 (initially for cardiovascular issues, but subsequently also for cystic fibrosis and breast cancer, among others).

To predict mortality, we used data from 850 patients to train our model, and then verified the accuracy of the model using results from 197 other patients from the same dataset. For ICU admission prediction, we trained with data from 950 patients and verified with data from 285 patients. To predict

need for ventilation, we trained with 810 patients and verified with 276 patients.

We called the new system we created "Adjutorium," meaning help, assistance or support.

### What we learned

So how did Adjutorium perform? Simply put: it did really, really well.

Once trained with patient data, Adjutorium was able to make highly accurate predictions about the patients whose data we used for verification. Crucially, we managed to do so much more accurately than existing and widely-used survival analysis techniques such as Cox regression or well-known indexes such as the Charlson comorbidity index.

Event	Adjutorium accuracy	Cox regression accuracy	Charlson index accuracy
Mortality	0.871 ± 0.002	0.773 ± 0.003	0.596 ± 0.002
ICU admission	0.835 ± 0.001	0.771 ± 0.002	0.556 ± 0.013
Ventilation	0.771 ± 0.002	0.690 ± 0.002	0.618 ± 0.002

Accuracy is measured using AUC-ROC. Higher is better.

It's also worth bearing in mind that Adjutorium achieved these results with a relatively small proportion of the data that could be gathered from COVID-19 cases globally. The more data we have access to, the better we can train our models and improve their accuracy, and the more useful Adjutorium becomes.

### Next steps

The progress we've made so far is extremely encouraging: we now have a functioning proof of concept that demonstrates the potential use of [machine learning](#) in helping to manage scarce resources like ICU beds and ventilators. There's still work to be done, though, and much of this will rely on continuing to receive new and high-quality

data.

Our immediate priority is to continue to validate the models we've developed. Doing so will bring us closer to finalizing the system for usage by healthcare professionals.

We also need to get our hands on new types of data that will make our existing models even more accurate. Specifically, we require longitudinal data that enables us to gain a deeper understanding of the progression of patients while they're hospitalized (rather than irregularly-recorded "snapshots" that show the state of affairs at specific times). Given how little is known about COVID-19, such data would provide valuable insights. Additionally, we're hoping for clearer [data](#) regarding the timing and effects of ventilators when used to treat patients. This would let us tell, for example, how long individual patients could or should have waited before ventilation in order to achieve the best possible outcomes.

We will also be working with the NHS and Public Health England to transform our tools into a system that can easily be used and understood by [healthcare professionals](#). In this sense, interpretability is key: we want to ensure that decision-makers can debug and analyze the information generated by our system.

Last week, I gave a presentation summarizing our progress at a COVID-19 workshop hosted by ELLIS (European Laboratory for Learning and Intelligent Systems). You can find the slide deck [here](#) and a video of my presentation embedded below.

Provided by University of Cambridge

APA citation: Using COVID-19 patient data to train machine learning models for healthcare (2020, April 6) retrieved 11 April 2021 from <https://medicalxpress.com/news/2020-04-covid-patient-machine-healthcare.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*