



# CPAS: the UK's national machine learning-based hospital capacity planning system for COVID-19

Zhaozhi Qian<sup>1</sup> · Ahmed M. Alaa<sup>2</sup> · Mihaela van der Schaar<sup>1,2,3</sup>

Received: 13 July 2020 / Revised: 5 September 2020 / Accepted: 26 September 2020 /  
Published online: 24 November 2020  
© The Author(s) 2020

## Abstract

The coronavirus disease 2019 (COVID-19) global pandemic poses the threat of overwhelming healthcare systems with unprecedented demands for intensive care resources. Managing these demands cannot be effectively conducted without a nationwide collective effort that relies on data to forecast hospital demands on the national, regional, hospital and individual levels. To this end, we developed the *COVID-19 Capacity Planning and Analysis System* (CPAS)—a machine learning-based system for hospital resource planning that we have successfully deployed at individual hospitals and across regions in the UK in coordination with NHS Digital. In this paper, we discuss the main challenges of deploying a machine learning-based decision support system at national scale, and explain how CPAS addresses these challenges by (1) defining the appropriate learning problem, (2) combining bottom-up and top-down analytical approaches, (3) using state-of-the-art machine learning algorithms, (4) integrating heterogeneous data sources, and (5) presenting the result with an interactive and transparent interface. CPAS is one of the first machine learning-based systems to be deployed in hospitals on a national scale to address the COVID-19 pandemic—we conclude the paper with a summary of the lessons learned from this experience.

**Keywords** Automated machine learning · Gaussian processes · Compartmental models · Resource planning · Healthcare · COVID-19

---

Editor: Hendrik Blockeel.

---

✉ Zhaozhi Qian  
zhaozhi.qian@maths.cam.ac.uk

Ahmed M. Alaa  
ahmedmalaa@ucla.edu

Mihaela van der Schaar  
mv472@cam.ac.uk

<sup>1</sup> University of Cambridge, Cambridge, UK

<sup>2</sup> University of California, Los Angeles, USA

<sup>3</sup> The Alan Turing Institute, London, UK

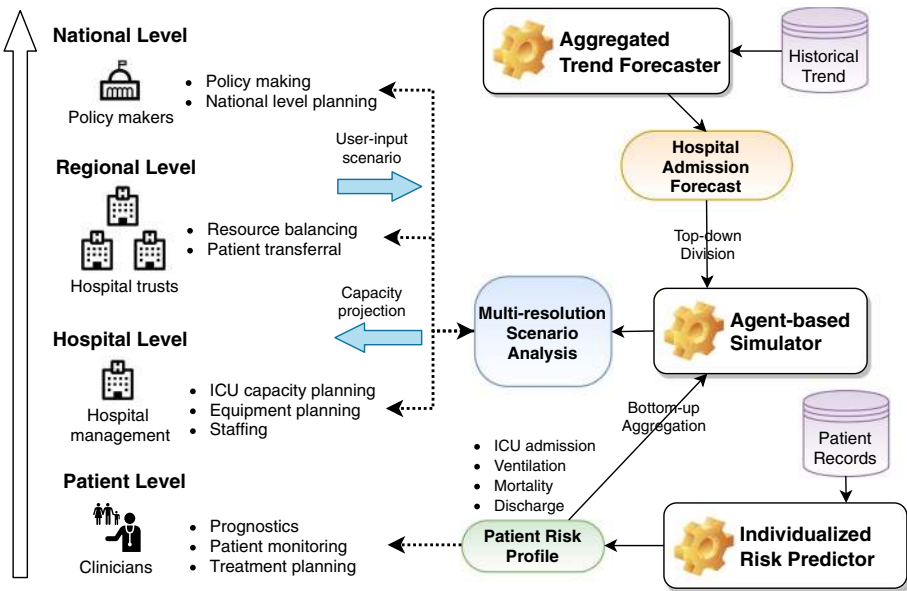
## 1 Introduction

The coronavirus disease 2019 (COVID-19) pandemic poses immense challenges to health-care systems across the globe—a major issue faced by both policy makers and front-line clinicians is the planning and allocation of scarce medical resources such as Intensive Care Unit (ICU) beds (Bedford et al. 2020). In order to manage the unprecedented ICU demands caused by the pandemic, we need nationwide collective efforts that hinge on data to forecast hospital demands across various levels of regional resolution. To this end, we developed the *COVID-19 Capacity Planning and Analysis System* (CPAS), a machine learning-based tool that has been deployed to hospitals across the UK to assist the planning of ICU beds, equipment and staff (NHS 2020d). CPAS is designed to provide actionable insights into the multifaceted problem of ICU capacity planning for various groups of stakeholders; it fulfills this goal by issuing accurate forecast for ICU demand over various time horizons and resolutions. It makes use of the state-of-the-art machine learning techniques to draw inference from a diverse repository of heterogeneous data sources. CPAS presents its predictions and insights via an intuitive and interactive interface and allows the user to explore scenarios under different assumptions.

Critical care resources—such as ICU beds, invasive mechanical ventilation and medical personnel—are scarce, with much of the available resources being already occupied by severely-ill patients diagnosed with other diseases (NHS 2020a, b). CPAS is meant to ensure a smooth operation of ICU by anticipating the required resources at multiple locations beforehand, enabling a timely management of these resources. While capacity planning has the greatest value at the peak of the pandemic, it also has important utility even *after* the peak because it can help the hospitals to manage the transition from the COVID-19 emergency back to the normal business. During the pandemic, the vast majority of the healthcare resources were devoted to treating COVID-19 patients, so the capacity for treating other diseases was much reduced. It is therefore necessary to review and re-assign the resources after the COVID-19 trend starts to decline. CPAS is one of the first machine learning decision support systems deployed nationwide to manage ICU resources at different stages of the pandemic.

CPAS is designed to serve the needs of various groups of stakeholders involved in capacity planning at different *levels* of geographical and administrative resolution. More specifically, as illustrated in Fig. 1, CPAS models ICU demands on the (a) patient, (b) hospital, (c) regional and (d) national levels. On the highest level, the system helps the *policy-makers* to make informed decisions by forecasting the national trends for ICU demand under different scenarios. Secondly, *regional* healthcare leadership can use CPAS as a “load balancing” tool as some hospitals may experience higher demand than others. Transfer of patients and resources can then be arranged accordingly. Next, hospital managers can use CPAS to plan ahead the *local* ICU space, equipment and staff. Lastly, the front-line clinicians can make use of the tool to understand the risk profile of *individual* patients. Importantly, these four groups of stakeholders require insights on different time horizons and levels of aggregation. CPAS addresses this challenge by combining the top-down projections from an “aggregated trend forecaster” and the bottom-up predictions from an “individualized risk predictor”. Using agent-based simulation, CPAS is able to provide multi-resolution scenario analysis for individual, local, regional and national ICU demand.

As illustrated in Fig. 1, CPAS comprises an aggregated trend forecaster (top right) that issues overall projections of hospital admission trend. The projections are made on the hospital level over a time horizon of thirty days, and they are further aggregated



**Fig. 1** Illustration of how the different components in CPAS address the diverse needs of stakeholders on various levels. On the regional level, “hospital trusts” refers to the NHS foundation trusts, organizations that manage several hospitals in a region

to form the regional and national forecasts. CPAS also makes use of an individualized risk predictor (bottom right) that contains a suite of machine learning pipelines predicting the patient-level risk profiles of ICU admission, ventilator usage, mortality and discharge in the thirty-day period after hospital admission. The patient level risk profiles can be directly used by the clinicians to monitor patient status and make a treatment plan. They can also be combined with the hospital admission forecast in the agent-based simulator (middle right) to perform scenario analysis for ICU capacity. CPAS casts the complex practical problem of ICU capacity planning into a set of sub-problems, each of which can be addressed by machine learning. This divide-and-conquer approach makes CPAS a transparent solution rather than a monolithic black-box. The machine learning models underlying CPAS are trained using data that reflects the entire patient journey: from hospital admission, to ICU admission, to ventilation treatment, and finally to discharge or death. We will discuss the problem formulation in detail in Sect. 2.1.

### The CPAS machine learning model

ICU planning for COVID-19 is a novel problem with few examples to learn from. The spread of COVID-19 is modulated by the intrinsic characteristics of the novel virus as well as the unprecedented intervention policies by the government. The need for deploying the system rapidly also means we cannot wait longer to collect more observations. As a result, the aggregated trend forecaster (Fig. 1 top right) needs to learn the disease’s transmission and progression characteristics from very limited data. To address this issue, we use a compartmental epidemiological model as a strong domain-specific *prior* to drive the trend forecast. We integrated the prior in the proven framework of Bayesian hierarchical modelling and Gaussian processes to model the complex disease dynamics from few observations (Rasmussen 2003). We explain the aggregated trend forecaster in detail in Sect. 2.3.

The individualized risk predictor (Fig. 1 bottom right) is a production-level machine learning pipeline tasked to predict four distinct outcomes. For each outcome, it needs to address all stages of predictive modelling: missing data imputation, feature processing, prediction, and calibration. For every stage, there are many machine learning algorithms to choose from, and for each algorithm there are multiple hyperparameters to tune. However, using a naive approach such as grid search to select algorithms and hyperparameters is very time-consuming, and it would hamper the rapid deployment of the system (Kotthoff et al. 2017). Since the pipeline configuration significantly affects the system's overall performance (Hutter et al. 2019), we use a state-of-the-art AutoML tool designed for medical applications, AutoPrognosis (Alaa and van der Schaar 2018), to address the algorithm and hyperparameter selection challenge for the individualized risk predictor (see Sect. 2.2).

It is inherently hard for a data-driven forecasting tool to accurately factor in the impact of unseen events (e.g. new social distancing policies). CPAS uses the agent-based simulator (Fig. 1 middle right) to allow the users to explore scenarios under different assumptions about the policy impact. The simulator first constructs a patient cohort that matches the feature distribution in the region of interest. It then uses the aggregated trend forecaster to determine the number of patients admitted to the hospital on each day based on the user's assumption about policy impact. As is standard in agent-based simulation (Railsback et al. 2006), each patient's outcomes are then simulated based on the risk profiles given by the individualized risk predictor. Finally, the simulated outcomes are aggregated to the desired level to form the scenario analysis. We introduce the details of the agent-based simulator in Sect. 2.4

### **Challenges associated with building practical machine learning-based decision support systems**

Data integration poses a practical challenge for implementing a large-scale machine learning system. As illustrated in Fig. 1, CPAS uses both patient records and the aggregated trends. Typically, the patient level information is collected and stored by various hospital trusts in isolated databases with inconsistent storage formats and information schema. It is a labour-intensive and time-consuming task to link and harmonize these data sources. Furthermore, historical data that contains valuable information about the patients' pre-existing morbidities and medications, are archived on separated databases. Accessing, processing and linking such historical data also proves to be challenging. CPAS is trained using a data set constructed from four data sources. By breaking the data silos and linking the data, we draw diverse information covering the full spectrum of the patient health condition, which leads to informed and accurate prediction. We introduce the details of the datasets in Sect. 3.1.

Presenting the insights in a user-friendly way is often neglected but vital to a machine-learning tool's successful adoption. Ideally, the user interface should not only present the final conclusion but also the intermediate steps to reach the conclusion. The transparency of the system's internal working makes the system more trustworthy. CPAS contains well-designed dashboards to display the outputs of the aggregated trend forecaster and the individualized risk predictors. Thanks to the agent-based simulator, CPAS also allows the users to interactively explore the future scenarios by changing the underlying assumptions rather than presenting the forecast as the only possible truth. We present an illustrative use case in Sect. 5.

The rest of the article is organized as follows: After formulating the ICU planning problem into a set of learning tasks in Sect. 2.1, we introduce the individualized risk predictor in Sect. 2.2, the aggregated trend forecaster in Sect. 2.3, and the agent-based simulator in Sect. 2.4. We proceed to describe the data sources used in CPAS and the training procedure

in Sect. 3. After that, we present the offline evaluation results and discuss the need for online performance monitoring in Sect. 4. In Sect. 5, we demonstrate how CPAS works in action by going through an illustrative use case. We conclude with the lessons learned in Sect. 6.

## 2 CPAS: a system for ICU capacity planning

### 2.1 Problem formulation

To formulate the ICU capacity planning problem, we start by modelling the patients' arrival at each hospital. Let  $A_h(t) \in \mathbb{N}$  be the number of COVID-19 patients admitted to a given hospital  $h \in \{1, \dots, N\}$  on the  $t$ th day since the beginning of the outbreak. Since not all hospitalized patients will require ICU treatment, we need to model the patient-level ICU admission risk to translate hospital admissions into ICU demand. Let  $X_i \in \mathbb{R}^D$  be the  $D$ -dimensional feature vector for a patient  $i$ . We use one-hot encoding to convert categorical variables into a real vector. Binary variables are encoded as 0 or 1. Further denote the event of ICU admission on the  $\tau$ th days after hospital admission as  $Y_i(\tau) \in \{0, 1\}$ . Note that  $Y_i(\tau)$  will not be available for some  $\tau$  due to insufficient follow-up time, e.g. we don't observe the outcome at 10 days after admission if the patient was only admitted 7 days ago. Formally speaking, We only observe  $Y_i(\tau)$  for  $\tau \in (0, \tau_i^*]$ , where  $\tau_i^*$  is the *censoring time* for patient  $i$ . The ICU admission event  $Y_i(\tau)$  directly translates into the ICU demand. To see this, consider a cohort of  $A_h(t)$  patients admitted to hospital  $h$  at time  $t$ . On a future date  $t' > t$ , the new ICU admission contributed by this cohort is given as:

$$\text{ICU-inflow}(t \rightarrow t') = \sum_{i=1}^{A_h(t)} Y_i(t' - t), \quad (1)$$

where we explicitly use  $t \rightarrow t'$  to emphasize the date of hospital admission  $t$  and the date of ICU demand  $t'$ . We can obtain the total ICU inflow on day  $t'$  by summing over all historical patient cohorts with different hospital admission dates:

$$\text{ICU-inflow}(t') = \sum_{t < t'} \text{ICU-inflow}(t \rightarrow t') = \sum_{t < t'} \sum_{i=1}^{A_h(t)} Y_i(t' - t). \quad (2)$$

It is apparent from the equation above that the ICU demand depends on two quantities: (1) the patient ICU risk profile  $Y_i(\tau)$  and (2) the number of hospital admissions  $A_h(t)$  in the range of summation. Therefore, CPAS uses the **individualized risk predictor** to model  $Y_i(\tau)$  and the **aggregated trend forecaster** to model  $A_h(t)$ . In addition to the ICU admission event, CPAS also contain models for three other clinical events: ventilator usage, mortality and discharge. We use  $Y_i(\tau)$  to conceptually represent any outcome of interest when the context is clear.

To build the CPAS individualized risk predictor, we consider a *patient-level* data set  $\mathcal{D}_{N,t}^p$  consisting patients from  $N$  hospitals over a period of  $t$  days, i.e.,

$$\mathcal{D}_{N,t}^p := \{X_i, Y_i[1 : \tau_i^*]\}_{i=1}^{N_p}, \quad (3)$$

where we use the square brackets to denote a sequence over a period of time i.e.  $Y_i[1 : \tau_i^*] := \{Y_i(1), \dots, Y_i(\tau_i^*)\}$ , and  $N_p := \sum_{h=1}^N \sum_{j=1}^t A_h(j)$ . CPAS learns the *hazard function* for any patient with feature  $X$  over a time horizon of  $\tau \in (0, H]$ :

$$\hat{h}(\tau) = \mathbb{P}(Y(\tau) = 1 \mid X, \mathcal{D}_{N,t}^P, Y(\tau') = 0, \forall \tau' < \tau) \tag{4}$$

Conceptually, the hazard function  $\hat{h}(\tau)$  represents the likelihood of ICU admission on day  $\tau$  given the patient has not been admitted to the ICU in the first  $\tau - 1$  days. Therefore, for any patient who is already admitted to the hospital and registered in the system, CPAS is able to issue individual-level risk prediction based on the observed patient features  $X_i$ . Those individualized predictions can directly help the clinicians to monitor the patient status and design personalized treatment plan. The predicted risk profiles can also be aggregated to the hospital level to measure the ICU demand driven by the *existing* patients.

To build the CPAS aggregated trend forecaster, we consider a *hospital-level* dataset  $\mathcal{D}_{N,t}^A$  for  $N$  hospitals covering a period of  $t$  days, i.e.,

$$\mathcal{D}_{N,t}^A := \{A_h[1 : t], M_h[1 : t]\}_{h=1}^N, \tag{5}$$

where the square brackets denote a sequence over time as before. The dataset contains the number of hospital admissions  $A_h(t)$ , which we have defined previously, and the community mobility  $M_h[1 : t] := \{m_h(1) \dots m_h(t)\}$  in the catchment of hospital  $h$ . The  $m_h(t) \in \mathbb{R}^K$  is a  $K$ -dimensional real vector, with each dimension  $k = 1 \dots K$  reflecting the relative decrease of mobility in one category of places (e.g. workplaces, parks, etc) due to the COVID-19 containment and social distancing measures. We used data for  $N = 94$  hospitals, each with  $K = 6$  categories of places. The details of the community mobility dataset is described in Sect. 3.1.4. For each hospital  $h$ , CPAS probabilistically forecasts the trajectory of the number of COVID-19 admissions within a future time horizon of  $T$  days with a given community mobility, i.e.,

$$\hat{A}_h[t : t + T] = \mathbb{P} \left[ A_h[t : t + T] \mid \underbrace{m_h(t), m_h(t + 1), \dots, m_h(t + T)}_{\text{Future mobility } M_h[t:t+T]}, \mathcal{D}_{N,t}^A \right]. \tag{6}$$

The future mobility  $M_h[t : t + T]$  depends on the intervention policies to be implemented in the future (e.g. schools to be closed in a week’s time) and therefore may not be learnable from the historical data. For this reason, CPAS only models the *conditional* distribution as in Eq. 6 and it allows the users to supply their own forecast of  $M_h[t : t + T]$  based on their knowledge and expectation of the future policies. By supplying different values of  $M_h[t : t + T]$ , CPAS is able to project the hospital admission trend under these different scenarios. By default, CPAS extrapolates the community mobility using the average value of the last seven days i.e.  $M_h(j) = \text{Avg}(M_h[t - 8 : t - 1])$  for  $j \in [t, t + T]$ . In the following three sub-sections, we will introduce the individualized risk predictor, the aggregated trend forecaster, and the agent-based simulator. When multiple modelling approaches are possible, we will discuss their pros and cons and the reason why we choose a particular one in CPAS.

**Table 1** The algorithms considered in each stage of the pipeline, which includes MICE (Buuren and Groothuis-Oudshoorn 2010), MissForest (Stekhoven and Bühlmann 2012), GAIN (Yoon et al. 2018), PCA, Fast ICA (Hyvarinen 1999), Recursive elimination (Guyon et al. 2002), Elastic net (Zou and Hastie 2005), Random forest (Liaw and Wiener 2002), Xgboost (Chen and Guestrin 2016), Multi-layer Perceptron (MLP) (Hinton 1990), Isotonic regression (De Leeuw 1977), Bootstrap (Chernick et al. 2011), Platt scaling (Platt et al. 1999)

Imputation	Feature selection	Prediction	Calibration
Median	<b>No selection</b>	Elastic net	<b>Isotonic regression</b>
MICE	PCA	Random forest	Bootstrap
MissForest	Fast ICA	<b>Xgboost</b>	Platt scaling
<b>GAIN</b>	Recursive elimination	MLP	

Algorithms in bold are the most frequently selected in each stage

### 2.2 Individualized risk prediction using automated machine learning

In order to learn the hazard function defined over a period of  $\tau \in (0, H]$  days (Eq. 4), the individualized risk predictor contains  $H$  calibrated binary classification pipelines, trained independently and each focusing on a single time step, as illustrated in Fig. 2. We will discuss the concept of pipelines later in this section. For now, the readers can assume  $P_\theta^*(X)$  to be a binary classifier with hyperparameters  $\theta$ . By training separate pipelines for each time step, we do not assume that the hazard function follows any specific functional form, which adds to the flexibility to model complex disease progression. The training data for each time step  $\tau$ , denoted as  $\mathcal{D}_{N,i}^P(\tau)$  is derived from the full patient-level dataset  $\mathcal{D}_{N,i}^P$  as follows:

$$\mathcal{D}_{N,i}^P(\tau) = \{(X_i, Y_i(\tau)) \mid (X_i, Y_i[1 : \tau_i^*]) \in \mathcal{D}_{N,i}^P, \tau_i^* \geq \tau, Y_i(\tau') = 0, \forall \tau' < \tau\} \tag{7}$$

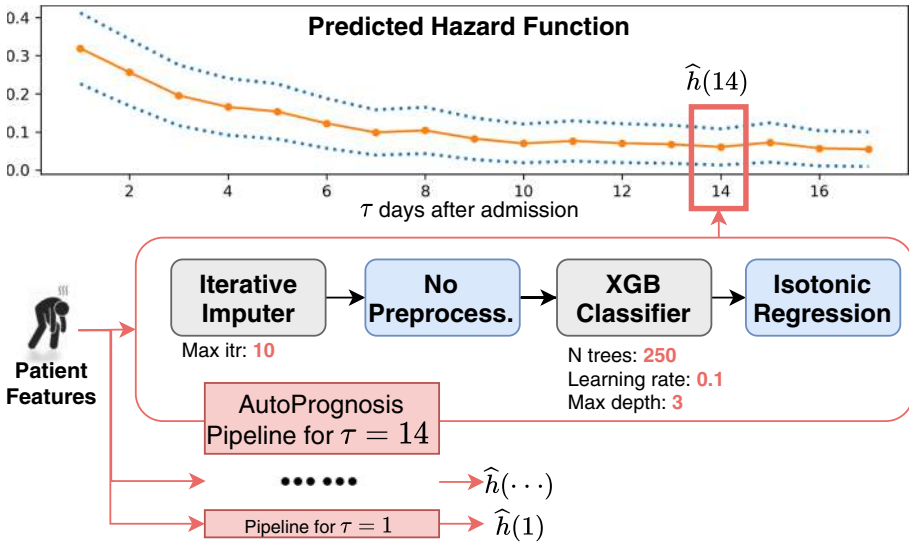
The first condition  $(X_i, Y_i[1 : \tau_i^*]) \in \mathcal{D}_{N,i}^P$  simply states that the patient  $i$  is in the full dataset. The second condition  $\tau_i^* \geq \tau$  ensures that the  $Y_i(\tau)$  is observed. In other words, the status of patient  $i$  at time  $\tau$  is not censored. The last condition  $Y_i(\tau') = 0, \forall \tau' < \tau$  arises from the definition of the hazard function in Eq. 4. Jointly these three conditions ensure that the binary classifier trained on  $\mathcal{D}_{N,i}^P(\tau)$  predicts the hazard function  $\hat{h}_i(\tau)$  at time  $\tau$ .

A machine learning pipeline consists of multiple stages of predictive modelling. Let  $\mathcal{F}_d, \mathcal{F}_f, \mathcal{F}_p, \mathcal{F}_c$  be the sets of all missing data imputation, feature processing, prediction, and calibration algorithms we consider (Table 1) respectively. A pipeline  $P$  is a tuple of the form:

$$P = (F_d, F_f, F_p, F_c), \tag{8}$$

where  $F_i \in \mathcal{F}_i, \forall i \in \{d, f, p, c\}$ . The space of all possible pipelines is given by  $\mathcal{P} = \mathcal{F}_d \times \mathcal{F}_f \times \mathcal{F}_p \times \mathcal{F}_c$ . Thus, a pipeline consists of a selection of algorithms from each column of Table 1. For example,  $P = (\text{MICE}, \text{PCA}, \text{Random Forest}, \text{Platt scaling})$ . The total number of pipelines we consider is  $|\mathcal{P}| = 192$ . While  $P$  specifies the “skeleton” of the pipeline, we also need to decide the hyperparameter configuration of its constituent algorithms. Let  $\Theta = \Theta_d \times \Theta_f \times \Theta_p \times \Theta_c$  be the space of all hyperparameter configurations. Here  $\Theta_v = \bigcup_a \Theta_v^a$  for  $v \in \{d, f, p, c\}$  with  $\Theta_v^a$  being the space of hyperparameters associated with the  $a^{\text{th}}$  algorithm in  $\mathcal{F}_v$ . Therefore, a fully specified pipeline configuration  $P_\theta \in \mathcal{P}_\theta$  determines the selection of algorithms  $P \in \mathcal{P}$  and their corresponding hyperparameters  $\theta \in \Theta$ .





**Fig. 2** Schematic depiction for the individualized risk predictor. A patient’s features are fed into multiple pipelines in parallel. Each pipeline estimates the hazard function  $\hat{h}(\tau)$  at a different time step  $\tau \in [1, H]$ . The pipeline for  $\tau = 14$  is illustrated in more details in the figure. The pipeline configuration specifies the algorithms and the associated hyperparameters. The configurations are determined by AutoPrognosis using Eq. 9 and may vary across  $\tau$

*Why not use survival analysis?* Some readers might have noticed that our problem formalism is similar to *survival analysis*, which also deals with time-varying outcome (survival) and considers censored data. When developing CPAS, we have considered this class of models. In fact, we have used the Cox proportional hazard model as a baseline benchmark (see Sect. 4). However, two factors discouraged us from further pursuing survival models. Firstly, the available implementations of survival models are not as abundant as classifiers and they are often immature for industrial scale applications. Secondly, many modern machine-learning powered survival models do not make the proportional hazards assumption (Pölsterl et al. 2016; Van Belle et al. 2011; Hothorn et al. 2006). The expense of relaxing assumption is that these models are often not able to estimate the full survival function and measure the absolute risk at a given time. There are recent works in survival analysis trying to address this issue (Lee et al. 2019), but it is still an open research area.

*Training the individualized risk predictor* In training, we need to find the best pipeline configuration  $P_{\theta^*} \in \mathcal{P}_{\theta}$  that empirically minimizes the  $J$ -fold cross-validation loss:

$$P_{\theta^*} = \operatorname{argmin}_{P_{\theta} \in \mathcal{P}_{\theta}} \sum_{j=1}^J \mathcal{L}(P_{\theta}, \mathcal{D}_{N,t}^P[\text{Train}(j)], \mathcal{D}_{N,t}^P[\text{Val}(j)]), \tag{9}$$

where  $\mathcal{L}$  is the loss function (e.g. Brier score), and  $\mathcal{D}_{N,t}^P[\text{Train}(j)]$  and  $\mathcal{D}_{N,t}^P[\text{Val}(j)]$  are the training and validation splits of the patient-level dataset  $\mathcal{D}_{N,t}^P$ . Note that this is a very hard optimization problem due to three facts (1) the space of all pipeline configurations  $\mathcal{P}_{\theta}$  has very high dimension, (2) the pipeline stages interact with each other and prevents the problem from being easily decomposable, and (3) evaluating the loss function via cross validation is a time-consuming operation.



Instead of relying on heuristics, we apply the state-of-the-art automated machine learning tool AutoPrognosis to configure all stages of the pipeline jointly (Alaa and van der Schaar 2018). AutoPrognosis is based on Bayesian optimization (BO), an optimization framework that has achieved remarkable success in optimizing black-box functions with costly evaluations as compared to simpler approaches such as grid and random search (Snoek et al. 2012). The BO algorithm used by AutoPrognosis implements a sequential *exploration-exploitation* scheme in which balance is achieved between exploring the predictive power of new pipelines and re-examining the utility of previously explored ones. To deal with high-dimensionality, AutoPrognosis models the “similarities” between the pipelines’ constituent algorithms via a sparse additive kernel with a Dirichlet prior. When applied to related prediction tasks, AutoPrognosis can also be warm-started by calibrating the priors using an meta-learning algorithm that mimics the empirical Bayes method, further improving the speed. For more technical details, we refer the readers to Alaa and van der Schaar (2018).

The AutoPrognosis framework has been successfully applied to building prognostic models for Cystic Fibrosis and Cardiovascular Disease (Alaa and van der Schaar 2018; Alaa et al. 2019). Implemented as a Python module, it supports 7 imputation algorithms, 14 feature processing algorithms, 20 classification algorithms, and 3 calibration methods; a design space which corresponds to a total of 5,880 pipelines. We selected a subset of most commonly used and well-understood algorithms for CPAS (Table 1), all of which have achieved considerable success in machine learning applications.

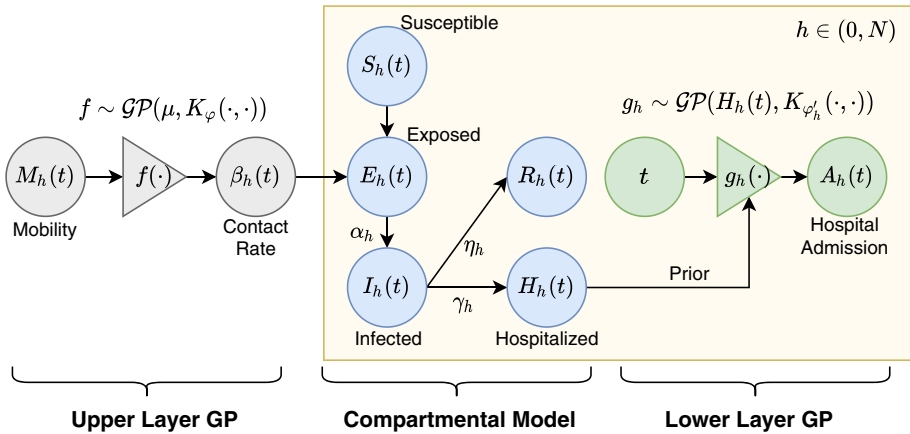
### 2.3 Trend forecast using hierarchical Gaussian process with compartmental prior

CPAS uses a hierarchical Gaussian process with compartmental prior (HGPCP) to forecast the trend of hospital admission. HGPCP is a Bayesian model that combines the data-driven Gaussian processes (GP) and the domain-specific compartmental models (Li and Muldowney 1995; Rasmussen 2003).

The compartmental model is a family of time-honoured mathematical models designed by domain experts to model epidemics (Kermack and McKendrick 1927). We use a specific version to model hospital admission (Hethcote 2000; Osemwinyen and Diakhaby 2015). As illustrated in Fig. 3, the compartmental model partitions the whole population containing  $C_h$  individuals into disjoint *compartments*: Susceptible  $S_h(t)$ , Exposed  $E_h(t)$ , Infectious  $I_h(t)$ , Hospitalized  $H_h(t)$ , and Recovered or died outside hospital  $R_h(t)$ , for  $\forall h \leq N$ . At any moment in time, the sum of all compartments is equal to the size of the population i.e.  $S_h(t) + E_h(t) + I_h(t) + H_h(t) + R_h(t) = C_h$  for  $\forall t > 0$ . As the pandemic unfolds, the sizes of the compartments change according to the following differential equations:

$$\begin{aligned} \frac{dS_h}{dt} &= -\beta_h(t)S_hI_h, & \frac{dE_h}{dt} &= \beta_h(t)S_hI_h - \alpha_hE_h, & \frac{dI_h}{dt} &= \alpha_hE_h - \gamma_hI_h, \\ \frac{dH_h}{dt} &= \eta_h\gamma_hI_h, & \frac{dR_h}{dt} &= (1 - \eta_h)\gamma_hI_h, \end{aligned} \quad (10)$$

where  $\alpha_h$ ,  $\gamma_h$ ,  $\eta_h$  are hospital-specific parameters describing various aspects of the pandemic, and  $\beta_h(t)$  is the *contact rate* parameter that varies across hospitals and time. Among all the parameters, the contact rate  $\beta_h(t)$  is of special importance for two reasons: (1)  $\beta_h(t)$  is the coefficient of the only non-linear term in the equation:  $S_h \cdot I_h$ . Therefore the nonlinear dynamics of the system heavily depends on  $\beta_h(t)$ . (2) The contact rate changes over time according to how much people travel and communicate. It is therefore heavily influenced by the various intervention policies.



**Fig. 3** Pictorial illustration of HGPCP. Left to right: The upper-layer GP  $f(\cdot)$  models the contact rate  $\beta_h(t)$  based on community mobility  $M_h(t)$ . The compartmental model gives the deterministic trajectory of the five compartments based on  $\beta_h(t)$ . The lower layer GP uses the hospitalized compartment  $H_h(t)$  as prior and predicts the hospital admission

To model the time-varying contact rate, the **upper layer** of HGPCP utilizes a function  $f$  drawn from a GP prior with mean  $\mu$  and covariance kernel  $K_\varphi(\cdot, \cdot)$ . The function  $f$  maps community mobility  $M_h(t)$  to the time-varying contact rate  $\beta_h(t)$  as follows:

$$f \sim \mathcal{GP}(\mu, K_\varphi(\cdot, \cdot)), \quad \beta_h(t) = f(M_h(t)), \tag{11}$$

where  $\mu$  and  $\varphi$  are hyperparameters of the GP. It is worth noticing that the upper layer of HGPCP is shared across hospitals. With the contact rate given by Eq. 11, the differential Eq. (10) can be solved to obtain the trajectory of all compartments. We used Euler’s method (Hutzenthaler et al. 2011) to solve these equations but other solvers are possible. The **lower layer** of HGPCP consists of  $N$  independent GPs that use the hospitalized compartment  $H_h(t)$  as a *prior* mean function and predicts hospital admissions over time:

$$g_h \sim \mathcal{GP}(H_h(t), K_{\varphi'_h}(\cdot, \cdot)), \quad A_h(t) = g_h(t), \quad \forall h \leq N \tag{12}$$

where  $\varphi'$  denotes the kernel hyperparameters. We use the radial basis function as the kernel for both upper and lower layers. Combining Eqs. 10, 11, and 12, the prediction problem (6) can be formulated in the following posterior predictive distribution:

$$\hat{A}_h[t : t + T] = \int \underbrace{\mathbb{P}(\mathcal{A}_h[t : t + T] \mid \mathcal{A}_h[1 : t], H_h)}_{\text{Lower layer GP}} \cdot \underbrace{\mathbb{P}(H_h \mid \beta_h, \theta_h)}_{\text{Compartmental model}} \underbrace{d\mathbb{P}(\beta_h \mid \mathcal{D}_{N,t}^A, \mathcal{M}_h[t : t + T])}_{\text{Upper layer GP}} \mathbb{P}(\theta_h), \tag{13}$$

where  $H_h := \{H_h(1), \dots, H_h(t + T)\}$ ,  $\beta_h := \{\beta_h(1), \dots, \beta_h(t + T)\}$ , and  $\theta_h = (\alpha_h, \gamma_h, \eta_h, \varphi'_h)$ . In this equation, the GP posterior terms can be computed analytically (Rasmussen 2003) while the compartmental model term has no closed form solution due to its nonlinearity. Therefore, we evaluate the integral via Monte Carlo approximation and derive the mean and quantiles from the Monte Carlo samples.

*What is the advantage over a standard GP?* Compared with a standard zero-mean GP, HGPCP is able to use the domain knowledge encoded in the compartmental models to inform the prediction. Since the spread of a pandemic is a highly non-stationary process — after the initial phase of exponential growth, the trend flattens and gradually reaches saturation, extrapolating observed data without considering the dynamics of the pandemic spread is likely to be misleading. The injection of domain knowledge is especially helpful at the early stage of the pandemic where little data are available.

*What is the advantage over a compartmental model?* Compared with the compartmental models, HGPCP uses the GP posterior to make prediction in a data-driven way. Since HGPCP only uses the compartmental trajectories as a prior, it is less prone to model misspecifications. Moreover, HGPCP is able to quantify prediction uncertainties in a principled way by computing the predictive posterior, whereas the compartmental models can only produce trajectories following deterministic equations. In capacity planning applications, the ability to quantify uncertainty is especially important.

*Training HGPCP* So far, we have assumed that the hyperparameters  $\alpha = (\varphi, \mu)$  of the upper layer GP are given. In practice, we optimize these hyperparameters by maximizing the log-likelihood function on the training data  $\mathcal{D}_{N,t}^A$ :

$$\mathcal{L}(\mathcal{D}_{N,t}^A | \alpha) := \log \int \prod_{h=1}^N \mathbb{P}(\mathcal{A}_h[1 : t] | H_h) \cdot \mathbb{P}(H_h | \mathcal{M}_h[1 : t], \alpha) dH_h, \quad (14)$$

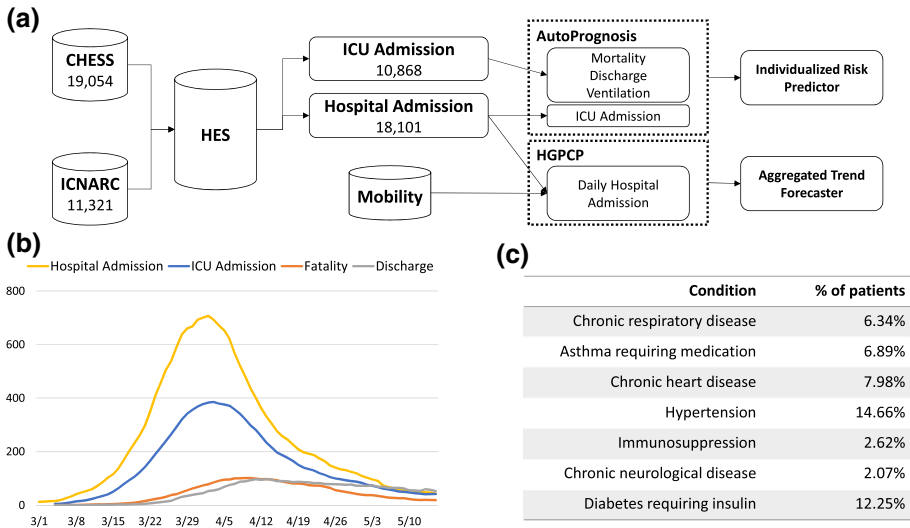
and  $\alpha^* = \operatorname{argmax}_{\alpha} \mathcal{L}(\mathcal{D}_{N,t}^A | \alpha)$ . Since the integral in (14) is intractable, we resort to a variational inference approach for optimizing the model's likelihood (Ranganath et al. 2014; Wingate and Weber 2013). That is, we maximize the evidence lower bound (ELBO) on (14) given by:

$$\text{ELBO}_i(\alpha, \phi) = \mathbb{E}_{\mathbb{Q}} \left[ \log \mathbb{P}(\mathcal{A}_h[1 : t], H_h | \alpha) - \log \mathbb{Q}(H_h | \mathcal{A}_h[1 : t], \phi) \right],$$

where  $\mathbb{Q}(\cdot)$  is the variational distribution parameterized by  $\phi$  with conditioning on  $\mathcal{M}_h[1 : t]$  omitted for notational brevity. We choose a Gaussian distribution for  $\mathbb{Q}(\cdot)$ , which simplifies the evaluation the ELBO objective and its gradients. We use stochastic gradient descent via the ADAM algorithm to optimize the ELBO objective (Kingma and Ba 2014).

The trained HGPCP model can issue forecasts on hospital level. To obtain regional or national level forecast, denote the set of hospitals in the region  $r$  as  $\mathcal{H}_r$ , and the regional forecast is obtained by taking summation of the constituent hospitals i.e.  $\hat{A}_r(t) = \sum_{h \in \mathcal{H}_r} \hat{A}_h(t)$  for  $\forall t \in (0, T]$ .





**Fig. 4** Illustration of the CPAS datasets and the training set up. **a** CHES and ICNARC data are joined and linked to HES to form the hospital patient data (18,101 cases) and the ICU patient data (10,868 cases). AutoPrognosis uses these two patient level datasets to train the various predictive pipelines in the individualized risk predictor. The aggregated hospital admission data together with the community mobility data empowers HGPCP to forecast the trend of admission. **b** The daily hospital admission, ICU admission, fatalities and discharges as recorded in the CPAS data set. **c** The prevalence of comorbidities and complications of hospitalized COVID-19 patients

ICU inflow. The ICU outflow due to discharge or death as well as ventilator usage can be derived in a similar fashion, and is omitted for brevity.

### 3 Training and deploying CPAS

#### 3.1 Dataset

CPAS relies on three distinct sources of patient-level data, each covering a unique aspect of patient health condition (Fig. 4). CPAS also makes use of community mobility trend data to issue aggregated trend forecast. The details of these data sources are described in the following sub-sections. The summary statistics of the data sets as of May 20<sup>th</sup> are shown in Fig. 4.

##### 3.1.1 COVID-19 hospitalizations in England surveillance system (CHES)

COVID-19 Hospitalizations in England surveillance system (CHES) is a surveillance scheme for monitoring hospitalized COVID-19 patients. The scheme has been created in response to the rapidly evolving COVID-19 outbreak and has been developed by Public Health England (PHE). It has been designed to monitor and estimate the impact of COVID-19 on the population in a timely fashion, to identify those who are most at risk and evaluate the effectiveness of countermeasures.

CPAS uses the de-identified CHES data updated daily from 8th February (data collection start), which records COVID-19 related hospital and ICU admissions from 94 NHS trusts across England. The data set features comprehensive information on patients' general health condition, COVID-19 specific risk factors (e.g. comorbidities), basic demographic information (age, sex, etc.), and tracks the entire patient treatment journey: when each patient was hospitalized, whether they were admitted to the ICU, what treatment (e.g. ventilation) they received, and their mortality or discharge outcome.

### 3.1.2 Intensive care national audit and research centre database (ICNARC)

Intensive care national audit & research centre (ICNARC) maintains a database on patients critically ill with confirmed COVID-19. The data are collected from ICUs participating in the ICNARC Case Mix Programme (covering all NHS adult, general intensive care and combined intensive care/high dependency units in England, Wales and Northern Ireland, plus some additional specialist and non-NHS critical care units). CPAS uses the de-identified ICNARC data which contains detailed measurements of ICU patients' physiological status (PaO<sub>2</sub>/FiO<sub>2</sub> ratio, blood pH, vital signals, etc.) in the first 24 hours of ICU admission. It also records each patient's organ support status (respiratory support, etc). ICNARC therefore provides valuable information about the severity of patient condition.

### 3.1.3 Hospital episode statistics (HES)

Hospital Episode Statistics (HES) is a database containing details of all admissions, A and E attendances and outpatient appointments at NHS hospitals in England. We retrieve HES records for patients admitted to hospital due to COVID-19. While the HES record contains a wide range of information about an individual patient, CPAS only makes use of the clinical information about disease diagnosis. All other information is discarded during the data linking process to maximally protect privacy. HES is a valuable data source because it provides comprehensive and accurate information about patients' pre-existing medical conditions, which are known to influence COVID-19 mortality risk.

### 3.1.4 Community mobility reports

In addition to the above patient-level data, we also used the COVID-19 Community Mobility Reports produced by Google (Google 2020). The dataset tracks the movement trends over time by geography, across six different categories of places including (1) retail and recreation, (2) groceries and pharmacies, (3) parks, (4) transit stations, (5) workplaces, and (6) residential areas, resulting in a  $K = 6$  dimensional time series of community mobility. It reflects the change in people's behaviour in response to the social distancing policies. We use this dataset to inform the prediction of contact rates over time. The dataset is updated daily starting from the onset of the pandemic.

## 3.2 Training procedure

By linking the three patient-level data sources described in the last section, we create two data sets containing hospitalized and ICU patients respectively (Fig. 4). For the ICU patients, we use AutoPrognosis to train three sets of models for mortality prediction, discharge prediction and ventilation prediction over a maximum time horizon of 30 days. For

**Table 2** Performance in forecasting individualized risk profile using different feature sets and algorithms measured by AUC-ROC

Model	Feature	ICU admission	Mortality	Ventilation
AutoPrognosis	All features	<b>0.835 ± 0.001</b>	<b>0.871 ± 0.002</b>	<b>0.771 ± 0.002</b>
AutoPrognosis	CHESS only	0.781 ± 0.002	0.836 ± 0.002	0.754 ± 0.003
AutoPrognosis	Demographics	0.770 ± 0.002	0.799 ± 0.003	0.702 ± 0.003
Cox PH Model	All features	0.771 ± 0.002	0.773 ± 0.003	0.690 ± 0.003
Charlson index	–	0.556 ± 0.013	0.596 ± 0.002	0.530 ± 0.006

The results in bold are significantly better than the rest

hospitalized patients, we carry out a similar routine to train predictors for ICU admission. All the AutoPrognosis pipelines are then deployed to form the individualized risk predictor in CPAS. Furthermore, we aggregate the patient-level data to get the daily hospital admission on hospital level. The data set is then combined with the community mobility report data to train the aggregated trend forecaster.

To make sure CPAS uses the most up-to-date information, we retrain all models in CPAS daily. The re-training process is automatically triggered whenever a new daily batch of CHESS or ICNARC arrives. Data linking and pre-processing is carried out on a Spark cluster with 64 nodes, and usually takes less than an hour to complete. After that, model training is performed on a HPC cluster with 116 CPU cores and 348 GB memory, and typically finishes within three hours. Finally, the trained models are deployed to the production server and the older model files are archived.

## 4 Evaluation and performance monitoring

### 4.1 Offline evaluation

In the offline evaluation, we first validated two hypothesis about the individualized risk predictor: (1) using additional patient features improves risk prediction and (2) the AutoPrognosis pipeline significantly outperforms the baseline algorithms. We performed 10-fold cross validation on the data available as of March 30 (with 1200 patients in total) and evaluated the AUC-ROC score on the  $\tau = 7$  day risk prediction. We compared AutoPrognosis to two benchmarks that are widely used in clinical research and Epidemiology: the Cox proportional hazard model (Cox 1972) and the Charlson comorbidity index (Charlson et al. 1994). The results are shown in Table 2. We can clearly see that there is a consistent gain in predictive performance when more features are included in the AutoPrognosis model and AutoPrognosis significantly outperforms the two benchmarks.

Next, we turned to the aggregated trend forecaster and validated the following two hypothesis: (1) HGPCP outperforms the zero-mean GP due to a more sensible prior (the compartmental model), and (2) HGPCP outperforms the compartmental model because GP reduces the risk of model mis-specification. We evaluate the accuracy of the 7-day projections issued at three stages of the pandemic: before the peak of infections (March 23), in the midst of the peak (March 30), and in the “plateauing” stage (April 23). Accuracy was evaluated by computing the mean absolute error between true and predicted daily hospital admission throughout the forecasting horizon, i.e.,  $\sum_{t=1}^7 |A_h(t) - \hat{A}_h(t)|/7$ . In Table 3 we



**Table 3** Performance in forecasting hospital admission

	Mar. 23 before peak			Mar. 30 at peak			Apr. 23 after peak		
	CPAS	GP	CM	CPAS	GP	CM	CPAS	GP	CM
STH	<b>3.01</b>	5.32	7.39	<b>11.49</b>	13.79	14.72	<b>3.18</b>	6.45	6.71
SGH	<b>1.37</b>	1.60	2.46	<b>6.41</b>	11.11	16.33	4.05	5.58	<b>3.20</b>
NPH	<b>3.90</b>	5.62	8.62	5.37	4.00	<b>3.97</b>	1.40	<b>1.22</b>	2.15
KCH	5.03	4.68	<b>3.84</b>	<b>1.74</b>	3.59	7.64	<b>2.31</b>	3.21	3.91
RLH	<b>3.24</b>	4.88	7.43	<b>2.59</b>	5.29	8.86	1.39	<b>1.06</b>	1.36
National	<b>14.35</b>	43.51	63.25	<b>46.47</b>	120.59	324.59	<b>25.19</b>	39.35	123.57

The candidate models are CPAS (HGPCP), GP (zero-mean GP) and CM (compartmental models). The first five rows refer to the performance in the five hospitals with most admitted patients. The last row refers to the national total admission. The lowest error for each task is bolded

report the performance for the five hospitals with most COVID-19 patients as well as the national level projection produced by aggregating all hospital level forecasts. We observe that HGPCP consistently outperforms the benchmarks on the national level across different stages of the pandemic. HGPCP also performs well on hospital level where the day-to-day fluctuation of ICU admission is bigger.

## 4.2 Online monitoring

It is vital to continuously monitor the performance after a machine-learning system is deployed. This is to prevent two possible scenarios (1) the gradual change in feature distribution worsens the predictive performance and (2) the breaking change in the upstream data pipelines causes data quality issues. Therefore, we always validate the model using a held-out set and record the performance whenever a model is re-trained. We have automated this process as part of the training procedure and developed a dashboard to visually track the performance over time.

## 5 Illustrative use case

Here we present an illustrative use case to demonstrate how CPAS works in real life. In this example, we show how the management of a particular hospital located in central London can use CPAS to plan ICU surge capacity for future patients before the peak of COVID-19. On March 23th, the hospital's ICU capacity had almost been fully utilized, but the pandemic had not yet reached the peak (refer to Fig. 4). The hospital management was desperate to know how many more patients would be admitted to the hospital and to the ICU in the coming weeks. They were planning to convert some of the existing general wards into ICU wards and, if necessary, to send some of the patients to the Nightingale hospital, a hospital specially constructed to support all NHS London hospitals in the surge of COVID-19 (NHS 2020c). CPAS could help the management to estimate how many general wards to convert and how many patients to transfer.

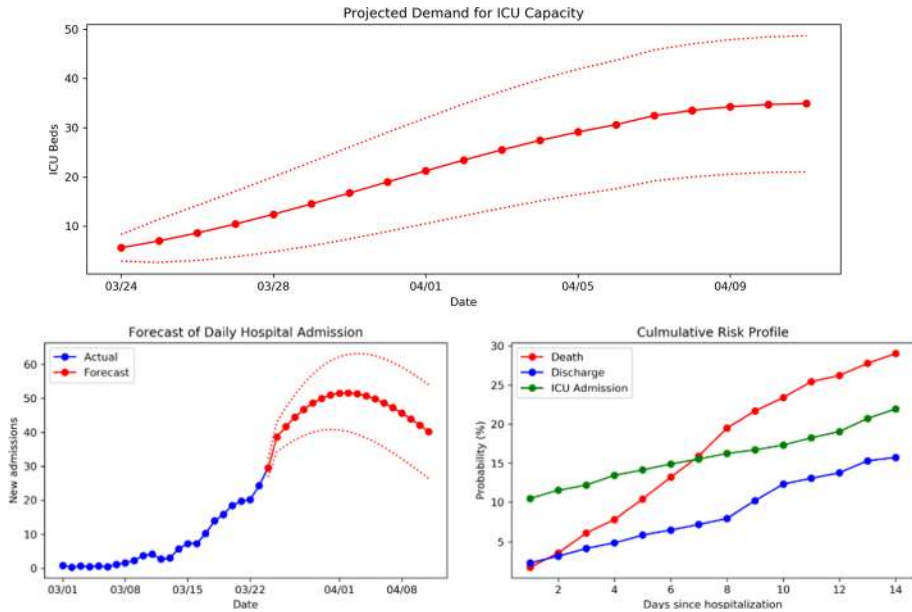
**Fig. 5** The configuration interface of CPAS. The user enters the desired level of resolution and the region of interest. The user then inputs the assumed trend for future community mobility. The empirical feature distribution in the region of interest is displayed below for reference

Figure 5 presents the input interface of CPAS<sup>1</sup>. The user first informed the system that the simulation should be performed on *hospital* level. Since the peak of the pandemic had not occurred yet on March 23th, it is reasonable to assume that the lockdown would continue and the community mobility would maintain at the lockdown level. Therefore, the user selected the *Extrapolation - Constant* option for the community mobility trend. Alternatively, the user can specify their own community mobility projections by selecting *User-defined* in the drop down and upload a file with the forecast numbers. CPAS displays the estimated feature distribution in the region of interest in the “Patient Cohort” section for reference.

The result of the simulation is shown in Fig. 6. The panel on the top projects that the ICU demand in the hospital would increase over the next two weeks as the pandemic progresses. The ICU capacity for thirty-five additional patients will be needed by mid April with the best scenario of 20 patients and the worst scenario of 50 patients. In reality, the hospital admitted 42 patients to the ICU in this period, which falls within the 95% confidence region given by CPAS. With this estimation, the management decided that the hospital could cope with the surge by converting existing general wards into ICU wards, and there was probably no need to transfer patients to the Nightingale hospital. This also turned out to be the case in reality.

The ICU demand projection above is driven by the aggregated hospital admission forecast (Fig. 6 bottom left), which predicts a steady increase of hospital admissions (around 50 per day) until early April when the trend starts to decline. The decline in admissions is caused by the social distancing policy and the lowered community mobility, which are assumed to persist throughout the simulation. The projection is also driven by the predictions of individual risk profiles. The plot on bottom right shows the average risk profile of the patient cohort. There is significant risk for ICU admission from the first day of hospitalization, whereas the risk for death and discharge increases over time starting from a small value. By presenting these intermediate results, CPAS shows more transparency in the overall ICU demand projection.

<sup>1</sup> All figures presented in this section are based on developmental versions of CPAS. They are for illustrative purposes only and do not represent the actual “look-and-feel” of CPAS in production.



**Fig. 6** The output interface of CPAS. CPAS displays the projected ICU demand with confidence intervals on the top. It then shows the intermediate prediction that leads to the projections. On the bottom-left, it shows the output of the aggregated trend forecaster. On the bottom-right, it shows the average risk profile for various outcomes given by the individualized risk predictor

## 6 Lessons learned

AI and machine learning have certainly not moved slowly in bringing seismic change to countless areas including retail, logistics, advertising, and software development. But in healthcare, there is still great unexploited potential for systematic change and fundamental innovation. As in the CPAS project, we can use AI and machine learning to empower medical professionals by enhancing the guidance and information available to them.

Collaboration is one of the most important aspects of straddling the divide between machine learning research and healthcare applications. In the CPAS project, we work closely with clinicians and stakeholders because they bring in domain expertise to inform the formulation of the problem and the design of the system. Effective collaboration is a challenge as we are all highly specialized in our respective areas, with different ways of thinking and different professional languages and approaches. As a result, we must each make extra effort to reach the middle ground. But it's a fascinating and invigorating way to work. Listening to clinicians and stakeholders can guide us to where problems and challenges actually lie, and then we can start being creative in trying to solve them. We found it is particularly helpful to build prototypes rapidly and get timely feedback from the collaborators. Each prototype should clearly demonstrate what can be achieved and what assumptions are made. We can then iteratively find out what functionalities we should focus on most and what we can assume about the problem.

Linking and accessing data is another challenge in healthcare applications. In the CPAS project, we first surveyed the potential data sources and understand the associated cost, which includes the financial cost, the waiting time for approval, the engineering cost, and

so on. After a thorough discussion with the collaborators, we prioritized what data to acquire first. It is often practical to start out with a single easy-to-access data source, and then expand the data sources as the system gets more adoption. The success of CPAS is greatly facilitated by the solid data infrastructure in the UK's healthcare system. Initiatives such as CHESS play a vital role in the data-driven response to the pandemic.

Transparency and interpretability are crucial for high-stake machine learning applications. The reality is that most machine learning models can't be used as-is by medical professionals because, on their own, they are black boxes that are hard for the intended users to apply, understand, and trust. While interpretable machine learning (Ahmad et al. 2018) is still an open research area, CPAS explored two practical approaches to make a machine learning system more interpretable. The first approach is to break down the problem into a set of sub-problems. This divide-and-conquer approach allows the users to understand how the final answer is derived from the smaller problems. The second approach is to let the users autonomously explore different scenarios using simulation rather than presenting the results as the only possible answer. Scenario analysis also allows the users to understand the level of uncertainty and sensitivity of the machine learning predictions.

Last but equally importantly, automated machine learning is a powerful tool for building large and complex machine learning systems. Most machine learning models cannot be easily used off-the-shelf with the default hyperparameters. Moreover, there are many machine learning algorithms to choose from, and selecting which one is best in a particular setting is non-trivial – the results depend on the characteristics of the data, including number of samples, interactions among features and among features and outcomes, as well as performance metrics used. In addition, in any practical application, we need entire processing pipelines which involve imputation, feature selection, prediction, and calibration. AutoML is essential in order to enable machine learning to be applied effectively and at scale given the complexities stated above. In CPAS, we used AutoML to generate a large number of machine learning models with minimum manual tweaking. AutoML not only helps the CPAS models to issue more accurate predictions but also saves the developers manual work so that we can focus on the design and modelling aspects. In the future, we will also explore the usage of AutoML to address the temporal shift in the feature distribution (Zhang et al. 2019), and to model time series collected in the clinical setting (van der Schaar et al. 2020).

**Acknowledgements** We thank Professor Jonathan Benger, Chief Medical Officer of NHS Digital, Dr. Jem Rashbass, Executive Director of Master Registries and Data at NHS Digital, Emma Summers, Programme Delivery Lead at NHS Digital, and the colleagues from NHS Digital and Public Health England for their tremendous support in providing datasets, computation environment, and domain expertise.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmad, M. A., Eckert, C., & Teredesai, A. (2018). Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics* (pp. 559–560).
- Alaa, A., & van der Schaar, M. (2018). Autoprognosis: Automated clinical prognostic modeling via bayesian optimization with structured kernel learning. In *International conference on machine learning* (pp. 139–148).
- Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H., & van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 uk biobank participants. *PLoS one*, *14*(5), e0213653.
- Alaa, A. M., & van der Schaar, M. (2018). Prognostication and risk factors for cystic fibrosis via automated machine learning. *Scientific Reports*, *8*(1), 1–19.
- Bedford, J., Enria, D., Giesecke, J., Heymann, D. L., Ihekweazu, C., Kobinger, G., et al. (2020). Covid-19: Towards controlling of a pandemic. *The Lancet*, *395*(10229), 1015–1018.
- Buuren, S. v., & Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in r. *Journal of statistical software* 1–68.
- Charlson, M., Slatkowski, T. P., Peterson, J., & Gold, J. (1994). Validation of a combined comorbidity index. *Journal of clinical epidemiology*, *47*(11), 1245–1251.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).
- Chernick, M. R., González-Manteiga, W., Crujeiras, R. M., & Barrios, E. B. (2011). Bootstrap methods
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.
- De Leeuw, J. (1977). Correctness of kruskal’s algorithms for monotone regression with ties. *Psychometrika*, *42*(1), 141–144.
- Google. (2020). Covid-19 community mobility. Retrieved July 4, 2020 from, <https://www.google.com/covid19/mobility/>.
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422.
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, *42*(4), 599–653.
- Hinton, G. E. (1990). Connectionist learning procedures. In *Machine Learning* (pp. 555–610). Elsevier.
- Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. J. (2006). Survival ensembles. *Biostatistics*, *7*(3), 355–373.
- Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). *Automated machine learning: Methods, systems, challenges*. Berlin: Springer.
- Hutzenthaler, M., Jentzen, A., & Kloeden, P. E. (2011). Strong and weak divergence in finite time of euler’s method for stochastic differential equations with non-globally lipschitz continuous coefficients. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *467*(2130), 1563–1576.
- Hyvarinen, A. (1999). Fast ica for noisy data using gaussian moments. In *1999 IEEE international symposium on circuits and systems (ISCAS)* (Vol. 5, pp. 57–61). IEEE.
- Kermack, W. O., & McKendrick, A. G. (1927). 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*(772), 700–721.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. Preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Kotthoff, L., Thornton, C., Hoos, H. H., Hutter, F., & Leyton-Brown, K. (2017). Auto-weka 2.0: Automatic model selection and hyperparameter optimization in weka. *The Journal of Machine Learning Research*, *18*(1), 826–830.
- Lee, C., Zame, W., Alaa, A., & Schaar, M. (2019). Temporal quilting for survival analysis. In *The 22nd international conference on artificial intelligence and statistics* (pp. 596–605).
- Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R News*, *2*(3), 18–22.
- Li, M. Y., & Muldowney, J. S. (1995). Global stability for the seir model in epidemiology. *Mathematical Biosciences*, *125*(2), 155–164.
- NHS. (2020a). Health careers in intensive care medicine. Retrieved July 4, 2020 from, <https://www.healthcareers.nhs.uk/explore-roles/doctors/roles-doctors/intensive-care-medicine>.
- NHS. (2020b). Intensive care. Retrieved July 4, 2020 from, <https://www.nhs.uk/conditions/Intensive-care/>.
- NHS. (2020c). Nhs nightingale london hospital. Retrieved July 4, 2020 from, <http://www.bartshealth.nhs.uk/nightingale>.

- NHS. (2020d). Trials begin of machine learning system to help hospitals plan and manage covid-19 treatment resources developed by nhs digital and university of cambridge. Retrieved June 28, 2020 from, <https://digital.nhs.uk/news-and-events/news/trials-begin-of-machine-learning-system-to-help-hospitals-plan-and-manage-covid-19-treatment-resources-developed-by-nhs-digital-and-university-of-cambridge>.
- Osemwinyen, A. C., & Diakhaby, A. (2015). Mathematical modelling of the transmission dynamics of ebola virus. *Applied and Computational Mathematics*, 4(4), 313–320.
- Platt, J., et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3), 61–74.
- Pölstner, S., Navab, N., & Katouzian, A. (2016). An efficient training algorithm for kernel survival support vector machines. Preprint [arXiv:161107054](https://arxiv.org/abs/161107054)
- Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9), 609–623.
- Ranganath, R., Gerrish, S., & Blei, D. (2014). Black box variational inference. In *Artificial Intelligence and Statistics* (pp. 814–822).
- Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer School on Machine Learning* (pp. 63–71). Springer.
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems* (pp. 2951–2959).
- Stekhoven, D. J., & Bühlmann, P. (2012). Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.
- Van Belle, V., Pelckmans, K., Suykens, J. A., & Van Huffel, S. (2011). Learning transformation models for ranking and survival analysis. *Journal of Machine Learning Research*, 12(3).
- van der Schaar, M., Yoon, J., Qian, Z., Jarrett, D., & Bica, I. (2020). clairvoyance alpha: the first unified end-to-end automl pipeline for time-series data. Retrieved July 4, 2020 from, <https://www.vanderschaar-lab.com/clairvoyance-alpha-the-first-unified-end-to-end-automl-pipeline-for-time-series-data/>.
- Wingate, D., & Weber, T. (2013). Automated variational inference in probabilistic programming. Preprint [arXiv:13011299](https://arxiv.org/abs/13011299)
- Yoon, J., Jordon, J., & Van Der Schaar, M. (2018). Gain: Missing data imputation using generative adversarial nets. In *International conference on machine learning (ICML)*.
- Zhang, Y., Jordon, J., Alaa, A. M., & van der Schaar, M. (2019). Lifelong bayesian optimization. Preprint [arXiv:190512280](https://arxiv.org/abs/190512280).
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.