Big data and targeted machine learning in action to assist medical decision in the ICU: the past, the present and the future

Romain Pirracchio, Mittchell J. Cohen, Ivana Malenica, Jonathan Cohen, Antoine Chambaz, Maxime Canesson, Christine Lee, Matthieu Resche-Rigon, Alan Hubbard

BIG DATA AND TARGETED MACHINE LEARNING IN ACTION TO ASSIST MEDICAL DECISION IN THE ICU

THE PAST, THE PRESENT AND THE FUTURE

Romain PIRRACCHIO, MD PhD [1,2,3,4]; Mitchell J COHEN[5], MD; Ivana MALENICA[1]; Jonathan COHEN [1]; Antoine CHAMBAZ [6], PhD; Maxime Cannesson[7,8], MD PhD ; Christine Lee[8], PhD; Matthieu RESCHE-RIGON, MD PhD[4]; Alan HUBBARD[1], PhD and the ACTERREA Research Group

[1] Division of Biostatistics, School of Public Health, University of California Berkeley, CA, U.S.A.

[2] Department of Anesthesia and Perioperative Medicine, University of California San Francisco, CA, U.S.A.

[3] Service d'Anesthésie-réanimation, Hôpital Européen Georges Pompidou, Université Paris Descartes, Sorbonne Paris Cite, Paris, France.

[4] Service de Biostatistique et Informatique Médicale, Hôpital Saint Louis, INSERM UMR-1153, Université Paris Diderot, Sorbonne Paris Cite, Paris, France

[5] Department of Surgery, University of Colorado Denver, Colorado, U.S.A.

[6] MAP5 (UMR CNRS 8145), Université Paris Descartes, Paris, France

[7] Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles, CA, USA

[8] Department of Bioengineering, University of California Irvine, CA, USA

Corresponding author:

Prof. Romain PIRRACCHIO

Service d'Anesthésie-réanimation

Hôpital Européen Georges Pompidou

20 rue Leblanc

75015 PARIS, FRANCE

Tel: +33.1.56.09.25.13

Fax: +33.1.56.09.25.20

Email: romain.pirracchio@aphp.fr

BIG DATA AND TARGETED MACHINE LEARNING IN ACTION TO ASSIST MEDICAL DECISION IN THE ICU

THE PAST, THE PRESENT AND THE FUTURE

Word Count: 4,894

Abstract Word Count: 104

Table: 0; Figures: 4

Conflicts of interest: Maxime Cannesson is a consultant for Edwards Lifesciences (Irvine, CA), Medtronic (Boulder, CO), Masimo Corp. (Irvine, CA). Maxime Cannesson has received research support from Edwards Lifesciences through his Department and NIH R01 GM117622 - Machine learning of physiological variables to predict, diagnose and treat cardiorespiratory instability and NIH R01 NR013912 - Predicting Patient Instability Non-invasively for Nursing Care-Two (PPINNC-2). Maxime Cannesson is co-owner of US patent serial no. 61/432,081 for a closed-loop fluid administration system based on the dynamic predictors of fluid responsiveness which has been licensed to Edwards Lifesciences. Christine Lee is an Edwards Lifesciences employee.

ABSTRACT

Historically, personalised medicine has been synonymous with pharmacogenomics and oncology. We argue for a new framework for personalised medicine analytics that capitalises on more detailed patient-level data and leverages recent advances in causal inference and machine learning tailored towards decision support applicable to critically ill patients. We discuss how advances in data technology and statistics are providing new opportunities for asking more targeted questions regarding patient treatment, and how this can be applied in the intensive care unit to better predict patient-centred outcomes, help in the discovery of new treatment regimens associated with improved outcomes, and ultimately how these rules can be learned in real-time for the patient.

INTRODUCTION

In his 2015 State of the Union, President Obama announced a federal Precision Medicine (PM) Initiative and the National Institutes of Health (NIH) has committed $72M to support this endeavor [1]. This effort came along with other relevant federal initiatives for personalized care, such as the Patient-Centered Outcomes Research Institute (PCORI) and California's Initiative to Advance Precision Medicine [2].

The Food and Drug Administration (FDA) describes PM as providing "the right patient with the right drug at the right dose at the right time." [3]. Advances in PM were most conspicuously driven by molecular oncology and pharmacogenomics, a sub-field of genetics that aims at studying gene-based variability in response to treatment [4]. However, although it is estimated that variability of response to over 25% of common medications have some genetic component [4], most therapies are not refined based upon biomarkers. There are many situations, such as critical illnesses, where the only immediately accessible information involves clinical or lab measurements, with limited time for collecting genotyping data. Nonetheless, critically ill patients would probably benefit from more tailored therapies since i) they receive many invasive treatments, which carry potentially lethal side effects; ii) the complex underlying pathophysiology of organ dysfunction makes every patient almost unique.

In such situations, there is a potential for PM beyond genomics, leveraging massive data from patients' electronic medical systems and continuous monitoring. Indeed, one now has the possibility to characterize a critically ill patient on all his observed characteristics (biological, contextual, demographic, phenotypic, medical imaging) including high throughput waveform signals generated by continuous monitoring. However, such a new approach for personalized medicine requires i) the ability to stream and store large amount of information, ii) appropriate analytics, i.e. statistical algorithms, and iii) the computational power to combine these. Machine learning (ML) is a field of statistical learning that may be used for various applications such as prediction/classification, regression, clustering or dimension reduction for instance. Machine learning algorithms are typically separated into *supervised* and *unsupervised* techniques. In supervised learning, on uses a training sample to provide the algorithm with example inputs (e.g. patients' characteristics) and their desired outputs (e.g., patients' outcome), and the goal is to learn a general rule that maps inputs to outputs. After a training phase, the supervised learner can be provided with new inputs (e.g. patients' characteristics from a new sample) and will produce a predicted output (e.g. a predicted outcomes). In unsupervised learning, no labels

(e.g. outcomes) are given to the learning algorithm, leaving it on its own to find structure in its input, the goal being for instance to discover hidden patterns in the data. While more and higher performing machine learning algorithms are constantly being developed, big data initiative are becoming mature enough with the emergent ability to stream and store large amount of medical information. Hence, the convergence of machine learning and big data has emerged as a crucial development that is shaping the future of personalized medicine and medical decision assistance in critically ill patients.

We discuss below important developments in these areas. Although it seems too early at this stage to perform a systematic literature review on the topic, we will illustrate the potential of these developments for more refined patient care with selected concrete examples of the possibilities offered in the near future in terms of personalizing clinical decision in the intensive care unit (ICU) with the most recent advances in machine learning applied to granular patient-level characteristics.

THE PAST AND THE PRESENT: EXAMPLES OF BIG DATA AND MACHINE LEARNING IN BIOMEDICINE

THE DATA: BIG DATA AT THE BEDSIDE IN THE ICU

Because the questions asked to help clinical decision in the ICU are often ambitious, e.g., predicting which treatment will better serve a patient in the next time interval, one typically expects that the amount of data required to produce a robust predictor can be quite large. Recent advances in medical informatics have enabled the creation of, and access to big observational medical data. For instance, i2b2 (Informatics for Integrating Biology and the Bedside [5]), the NIH-funded National Center for Biomedical Computing, develops a scalable informatics framework that facilitates the use of existing clinical data by clinical researchers. As an illustration, the MIMIC ('Medical Information Mart for Intensive Care') project [6] started in 2001 in Boston, MA, as a single-centre database comprising information on ICU patients at a large tertiary care hospital. In MIMIC, databases are dumped off-line while bedside waveform data and derived trends are collected by an archiving agent over TCP/IP. As described in Figure , source data consists of i) bedside monitor waveforms, ii) clinical data including treatment extracted from the electric medical system, iii) data from hospital electronic archives, and iv) mortality data from the Social Security Death Index. These data are then assembled in a protected and encrypted central repository. The data are de-identified to produce a final set of data for public access and use (https://physionet.org/mimic2/).

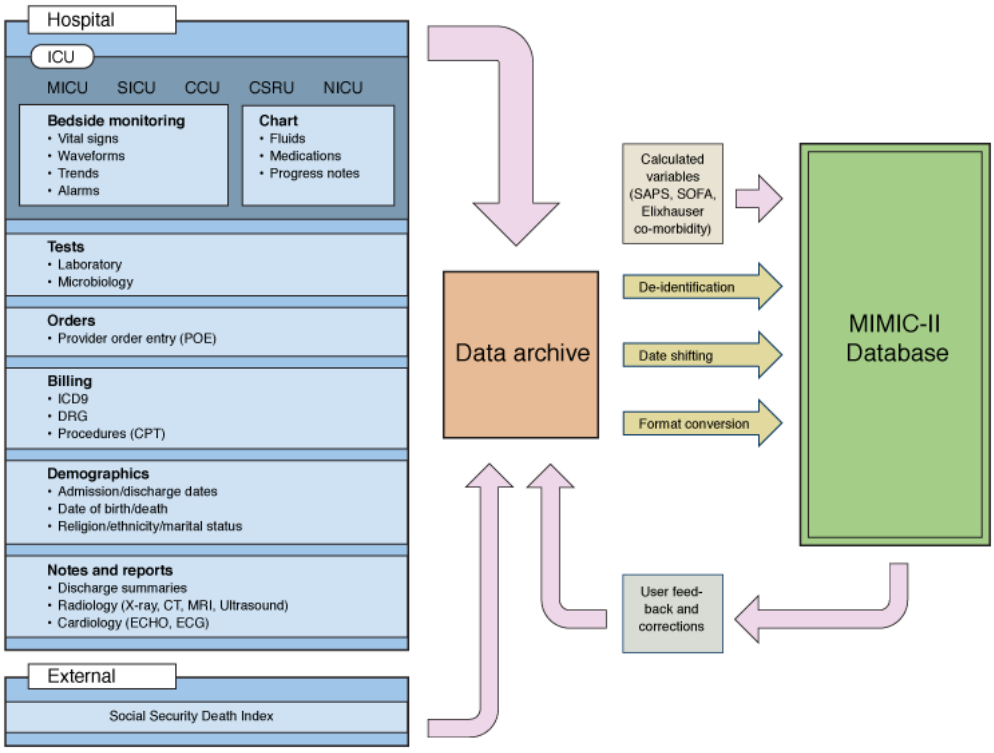

FIGURE 1: SCHEMATIC DESCRIPTION OF DATA COLLECTION AND MIMIC-II DATABASE CONSTRUCTION.

The Mayo Clinic has also developed a data warehouse from ICU patients' data, called the METRIC (Multidisciplinary Epidemiology and Translational Research in Intensive Care Data Mart) Data Mart [7] with the goal to develop and validate an informatics infrastructure for syndrome surveillance, decision support, reporting, and modelling of critical illness. This open database encompasses physiologic monitoring, medication orders, laboratory and radiologic investigations, and physician and nursing notes. The University of California, Irvine Machine Learning Repository (http://archive.ics.uci.edu/ml/index.php) maintains datasets for the ML research community that includes various physiological and biological data. In addition to ICU data repositories developed at the hospital level, some countries have decided to develop system-wide databases. In the Netherlands, almost 100 ICUs are sharing their clinical date through a common registry [8], while the Australia and New Zealand Intensive Care Society (ANZICS) registry is capturing clinical data from more than 40 ICUs [9]. This list is not exhaustive and many other institutions, hospital groups or countries have launched their own data acquisition program.

USE OF BIG DATA FOR PERSONALIZED MEDICINE IN THE ICU

Recently, there has been a worldwide effort to promote treatment personalization in the ICU. PM in the ICU is based upon accounting for patient heterogeneity, including accounting for variation in disease risk factors, and/or heterogeneous responses to treatment. Identification of homogeneous subgroups of patients in terms of outcome or response to treatment has so far relied on a relatively small subset of clinical features or biomarkers. For instance, in attempts to optimize hemodynamic resuscitation, the use of fluid responsiveness indices including diverse functional hemodynamic parameters has been advocated to allow for a better tailoring of fluid administration and avoid fluid overloading in patients. However, as underlined by Saugel et al. [10] goal-directed hemodynamic treatment strategies often use predefined fixed population-based 'normal' values as hemodynamic targets while hemodynamic variables have large inter-individual variability. Hence, in addition to functional hemodynamic parameters, these authors advocate for the use of more personalized hemodynamic targets identified through adaptive multiparametric hemodynamic.

Biomarkers have been also studied in critical illnesses such as sepsis and acute respiratory distress syndrome. Simple and routinely available biomarkers, such as HbA1C, can be extremely useful in tailoring glycaemic control in critically ill patients [11]. Other serum biomarkers have been proposed to better differentiate septic patients with acute respiratory distress syndrome (ARDS) from those without [12], as well as ARDS from direct or indirect lung injury [13]. In the future, more complicated multi-parameter panels may prove helpful in characterizing critical illnesses, as well as optimizing subject-selection for clinical trials. As an example, based on the combination of vital signs, ventilator settings, and laboratory data including serum biomarkers, sub-phenotypes of ARDS were recently identified in terms of prognosis and response to specific treatments [14,15].

Genomics was also proposed to better characterize complex ICU syndromes such as sepsis. Gene expression profiling has been proposed as a way to differentiate inflammation status from septic and non-septic origin [16], or to separate different subtypes of sepsis or septic shocks [17–19]. Genotyping has also proven useful to predict favourable or unfavourable outcomes in sepsis [20] or ARDS [21], and the response to a variety of therapeutics [22,23]. Although genomics and other omics technologies (proteomics, metabolomics, etc.) may open new paths towards better understanding of the heterogeneity across critically ill patients, their use in daily clinical practice remains limited. This is essentially due to the

fact that diagnostic tests in the ICU need to be easily deployable at the point-of-care, when genotyping technologies are still too slow.

Beyond omics, there is yet another promising approach to PM in the ICU, based on leveraging massive data from electronic medical systems (including biology and imaging) and continuous waveform monitoring. As stated by Levin et al. [24], 5-lead EKG waveform sampling generates in 2 hours 37MB of data. Since the human genome contains approximately 1.5 GB, 54 hours sampling of a 5-lead EKG waveform generates as much raw data as contained in the human genome. With the cost related to data storage falling, the opportunity to harvest these data for research and PM is becoming a practical reality. As recently emphasized in both the Johns Hopkins magazine [25] and Harvard Business Review [26], big data and ML will be useful in the near future to make wiser medical decisions at the individual level because i) large amount of data is necessary for each patient to precisely characterize each of them; ii) large sample sizes are needed to observe a sufficiently large number of patients with similar characteristics, and therefore to allow estimation of treatment effects within more narrowly defined clusters, iii) the human brain cannot quickly process high throughput signals such as continuous EKG, whereas modern ML algorithms can. Thus, ML may be used in this context to process in real-time not only a large amount of data but also high-fidelity signals and to extract an information of reduced dimension that may be integrated and interpreted by doctors to make better medical decisions. How this can be done using a statistical framework is discussed and illustrated using a non-exhaustive list of examples in next section.

MACHINE LEARNING FOR PREDICTIVE ANALYTICS AND DECISION SUPPORT IN THE ICU

PREDICTIVE ANALYTICS

Predictive analytics is usually defined as the branch of advanced analytics, which is devoted to making predictions about future events. In the medical setting, the goal is usually to answer the question "will my patient experience the event of interest in the future?". Predictive analytics is often opposed to *prescriptive analytics* which goal is to answer the question of "what should I do in a given patient?". Prescriptive analytics will be discussed in more details in a specific section on *how to assist medical decision*.

Predictive analytics uses many modelling techniques including ML to analyse current data and make predictions about future patient health given current and past states. Over the past decades, there have been multiple attempts to predict diverse outcomes in critically ill patients, such as ICU mortality, 30-day readmission [27,28], or the risk cardiovascular decompensation using ML [29]. Based on MIMIC-II, recent studies have shown that new

generation ML algorithms can substantially improve the prediction of hospital mortality in critically ill patients over current commonly used instruments [30,31]. Large scale repositories such as MIMIC-II were also used to predict more specific outcomes such as i) the optimal duration for a trial of intensive care in patients with active cancer and septic shock [32], ii) the need to perform an endoscopy or a surgical treatment in patients admitted for gastrointestinal bleeding or iii) the likelihood that a new laboratory test would substantially differ from the last determination with the goal to reduce unnecessary blood draws [33]. Most of these studies did not include continuous physiological signals such as heart rate or blood pressure for prediction.

PREDICTION FROM CONTINUOUS MONITORING

Recently, ML algorithms were also proposed to analyse continuous signals provided by monitoring systems. Moss et al. [34] used continuous EKG to better predict patient deterioration defined as unanticipated death or ICU transfer. They showed that adding continuous EKG to discrete vital signs and laboratory results substantially improves the prediction performance. Accordingly, the data from monitoring systems were also used to predict cardiorespiratory insufficiency in different settings such as the operating room, the ICU or step-down units [29,35–37]. In addition, the use of continuous monitoring information as well as clinical data has been successful in predicting diagnoses in the emergency department (ED) and ICU [38–40].

Continuous monitoring signals were also used to develop automated prediction tools of acute events such as hypotensive episode in critically ill patients. Jiang et al. [41] used a Probability Distribution Patterns Analysis (PDPA) method to extract relevant information from time series of continuous blood pressure. Subsequently, they used a ML approach combining Genetic Algorithm (GA) and Support Vector Machine (SVM) to select the vital features for effective classification. When applied to a large validation cohort, the obtained accuracy for hypotension classification and forecasting was of 80.8%, sensitivity of 78.2%, specificity of 81.5%. Edwards LifeSciences (Irvine, California, USA) has developed, embedded into a monitor and commercialized a hypotension prediction algorithm called Acumen Hypotension Probability Index ® (HPI). The underlying machine algorithm was trained once on a population of patients including 13,000 hypotensive events (defined as a mean arterial pressure < 65mmHg) and 12,000 non-hypotensive events. The signal used to train the algorithm was the continuous invasive blood pressure waveform. Although this algorithm can be described as a static learner since it does not retrain using new batches of data, it actualizes its prediction every 20 seconds based on the evolution of the blood

pressure waveform signal. Clinical evaluation of the Acumen Hypotension Probability Index ® are currently ongoing in the operating room and in critically patients.

Outside the ICU, ML applications have also gained interest to classify medical images such as mammograms and MRIs [42]. In the aforementioned applications, and beyond, ML algorithms are not interpretable in the sense that one cannot fathom how a trained ML algorithm derives a result such as a score, an index [29,35–37] or a more complex output. In addition, even when the model itself produces interpretable results, e.g., the coefficients produced by a logistic regression model as used to derive the HPI [43], the features themselves included as predictors in the model (e.g. interactions or high-order terms) might not be clinically interpretable. In a recent effort to circumvent this weakness and translate trained ML algorithms into clinically meaningful models, van Poucke et al.[44] have developed a specific open and code-free environment to provide visual tools and scalable predictive analytics for clinical research. In the same vein, Che et al. [45] proposed deep learning algorithms, which have shown a wide range of success across many fields, that are more interpretable for wide clinical adoption. Variable importance measures [46] can also provide complementary information that may help end users better understand the relative importance of a particular variable (or group of variables) and thus mitigate the limitation in terms of direct interpretability of flexible learning tools that yield black-box predictions.

THE FUTURE : GOING FROM USING FIXED DATA TO LEARNING "ON THE FLY"

In the former section we have discussed how ML algorithms can exploit big data to assist medical decision-making (e.g., by predicting a wide range of outcomes in critically ill patients). However, as recently emphasized by Lee et al. [47], it should be feasible, in the era of digital healthcare, to dynamically personalize decision support by identifying and analysing similar past patients, in a way that is analogous to personalized product recommendation in e-commerce. Nevertheless, most of the existing clinical applications of ML were obtained using static population-wide prediction models, i.e. ML algorithms first trained on a large population of patients and subsequently used to predict the outcome in a different sample. Recent advances in ML methods will allow for i) *realtime time processing and learning* through stream data analytics platform, such as the one proposed by Bai et al. [48]; ii) *continuous learning* from time series at the patient level, thus providing more personalized predictions; iii) *prescriptive* analytics. Such an automated technology deployable at the bedside is the path for the ultimate goal, i.e. personalized point-of-care decision support tools. We illustrate here how new statistical tools capitalizing on recent

advances in statistics and ML may be used to further leverage the data generated in the ICU and achieve the above three-fold ambitious objective.

STATISTICAL METHODS FOR ONLINE LEARNING

Reinforcement *learning and reinforcement deep learning have showed great performance in situations where the machine can rely on essentially infinitely many experiments (e.g., in chess or go, where the machine plays against itself) and/or massive amounts of data (e.g., billions of e-purchases, huge corpuses of text)*[49]. However, theoretical arguments suggest that despite its extreme versatility, deep learning might not do as well in finite sample and for data and problems related to precision medicine [50]. Because of i) the complex and heterogeneous nature of healthcare data in electronic medical records, ii) the potentially massive number of variables and iii) the relatively limited sample size, learning procedures that balance variance and bias more parsimoniously are needed. Thus, more general approaches are necessary that do not rely on single algorithms/learners to develop a prediction or prescriptive function.

SUPER LEARNING

As an alternative, a recent ensemble ML algorithm called Super Learning (SL) [51] guarantees prediction optimality. SL uses cross-validation to build an optimal meta-algorithm from a user-supplied library consisting of individual ML algorithms. Thus, one does not need to choose a specific ML algorithm. Instead, one has to provide a library of candidate ML algorithms and to let the Super Learner (SL) identify the optimal algorithm data-adaptively. Theoretical results show that the meta-algorithm resulting from this stacking procedure performs as well as the so-called oracle algorithm, defined as the problem-specific best ML algorithm in the library.

In practice, the user provides SL with a library encompassing a wide range of ML algorithms. The library typically includes both rigid algorithms based on low-dimensional parametric models (such as a linear regression with few predictors) on one side of the spectrum and, on the other side, very flexible ML algorithms (such as neural networks). Not only will the SL identify the optimal algorithm data-adaptively, but it will also build a unique meta-algorithm defined as a weighted combination of the individual algorithms in the library. Learned from the data, each weight reflects the empirical performance of the corresponding algorithm in such a way that, overall, algorithms that perform best are given larger weights. As mentioned earlier, studies have shown for instance that SL can substantially improve the prediction of hospital mortality in critically ill patients [30,31] by learning from independent data. But, it is often the case that one wants to learn from data

which are not independent, like for instance from a single trajectory/time series obtained by monitoring a single patient. By doing so, the algorithm should i) constantly improve its prediction performance by integrating new pieces of information and ii) progressively tailor a single patient since it is learning for its own history, thereby achieving one of the goals of acute care personalized medicine.

Online Super Learning

Online ML, sometimes referred to as stream analytics, is a method of ML where data accrue and are used sequentially to update the current trained algorithm. Online ML is usually opposed to batch learning which generates the best predictor by learning on the entire training data set at once and then, this predictor is used to predict from new batches of data. It is important to note that batch learning algorithms are sometimes misleadingly referred to as *online predictors* (as opposed to online learners) since the initial predictor can be used several times to update the prediction, but without any further learning sequence. The Acumen Hypotension Probability Index from Edwards (Irvine, California, USA) referenced earlier in the paper, is an example of online predictor (but not online learner since the prediction algorithm is not automatically updated when plugged in to a new patient).

The concept of SL naturally lends itself to that of Online Super Learning (OSL). The theoretical, and practical, performances are essentially preserved. Carrying out OSL consists in i) providing a library of individual ML algorithms; ii) training each of them on an initial batch of data; iii) evaluating the performances of all the algorithms based on a new batch of data, and deriving from them either the current best algorithm or the current best combination of the algorithms; iv) sequentially updating the initial SL by exploiting successive, mutually exclusive new batches of data as in (ii) (substituting "updating" for "training") and (iii). Separating training data from testing data (here, sequentially) is crucial. It is as an instantiation of the cross-validation principle for time (and/or space) structured data.

We illustrate the concept of OSL with a schematic example. The goal is to predict future diastolic blood pressure (DBP) 1 minute in the future based on the history of DBP of the patient up to that point, using the MIMIC-II dataset [6]. Though of limited clinical significance, this illustration provides a simple example of how algorithms can be developed to learn from the history of a patient to develop accurate clinical prediction for anticipatory clinical care.

Formally, let consider a time series data structure $X_t \equiv (W_t, Y_t)P \in M$ where Y_t is a response variable and W_t is the vector of covariates, with observations indexed by time t. The parameter of interest for a forecasting method with horizon δ is:

$$\theta(\bar{W}(t), \bar{Y}(t)) \equiv E\{Y(t+\delta)|\bar{W}(t), \bar{Y}(t)\}$$

where E is the conditional expectation of the outcome Y, $(\bar{W}(t), \bar{Y}(t))$ represents the history of both a set of covariates $\bar{W}(t) = (W(0), W(1), \dots, W(t))$ and the outcome process $\bar{Y}(t) = (Y(0), Y(1), \dots, Y(t))$.

In words, this parameter is the mean of the outcome, δ units of time after time *t*, considering the history of the covariates W and the outcome Y up to time t.

One could either fit this based only upon the data available on a single patient (the *longitudinal* source of information) or using data across previous patients (the *transversal* source of information). The specific cross-validation procedure used for online will then depends on what one is trying to optimize (prediction for a single patient or average performance across many patients for instance).

Though one could use a wide variety of online algorithms, for this illustration the library specifically contained only four versions of the Long Short Term Recurrent Neural Network (LSTRNN) [52] algorithm, each controlled by different tuning parameters. In the present example, the optimal combination of the four versions of LSTRNN included in the SL library is updated in time with new batches of validation data.

Figures 2 and 3 illustrate how, in order to predict the outcome $Y_t$in some time interval based on current history of observations $\bar{W}(t), \bar{Y}(t)$, one can use OSL. This can be expended by creating many intervals on the same patient so that the model is moving forward in time and thus constantly refining from an initial start.

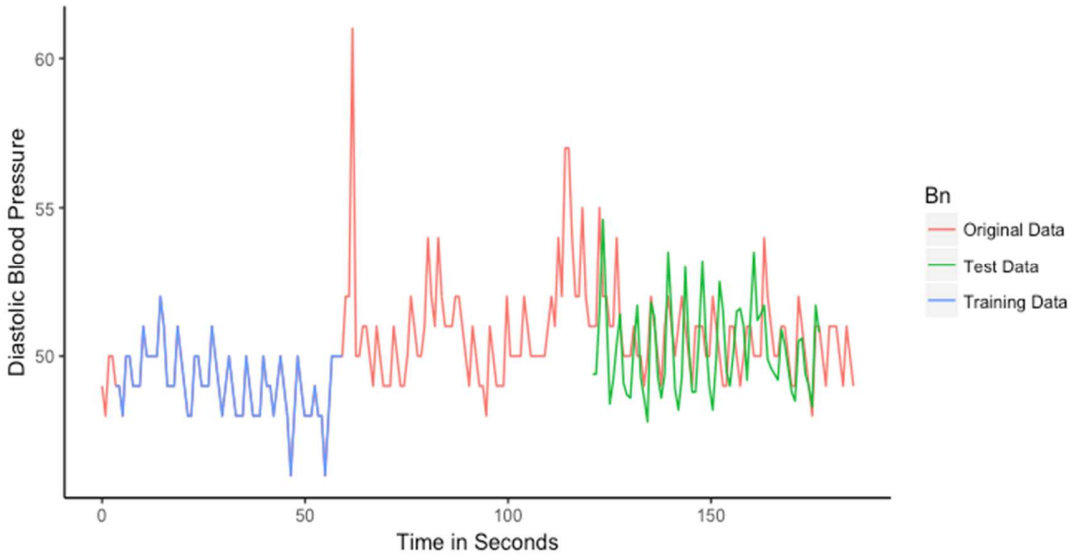

FIGURE 2: ILLUSTRATION OF THE DIFFERENT PHASES OF THE PROCEDURE. THE BLUE LINE INDICATES THE TRAINING SET, THE RED THE ACTUAL OUT-COME SERIES AND THE GREEN (AFTER THE LAG) THE PREDICTION OF DBP BASED ON THE HISTORY PRECEDING THE GAP.

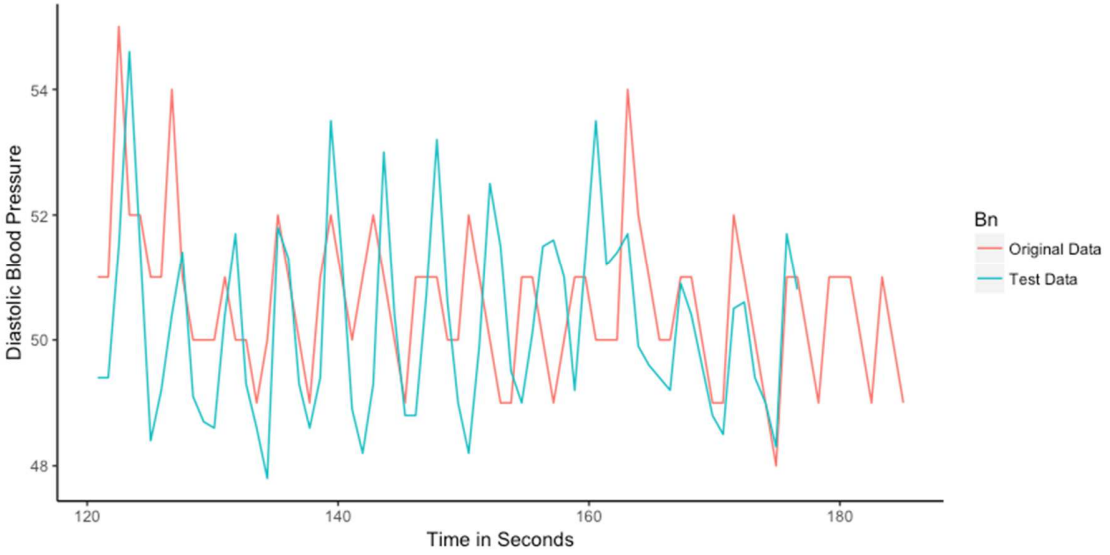

FIGURE 3: ZOOM ON THE SIMILARITY BETWEEN PREDICTED AND ACTUAL DIASTOLIC BLOOD PRESSURE.

METHODS FOR ASSISTING CLINICAL DECISION

The ability to better predict an outcome at the patient level, as described in the former section can be useful for clinicians, but even more useful would be to predict future scenarios under different treatment regimens. Indeed, such information would help the

clinician to decide which treatment option is most relevant at the patient level. As an illustration, in the situation where a hypotensive episode is predicted as described above, prescriptive analytics should help the clinician to answer the question "will the patient be responsive to fluid?" To do so, the algorithm needs to be previously trained on observations under the 2 treatment options (e.g. fluid versus no fluid).

The causal inference framework is very useful to identify the difference of outcome under two potential treatments or exposure and thus help the clinician to choose the one associated with the most favourable outcome. To do so, one trains an algorithm to predict the outcome based on the actual treatment and adjusting for all potential confounders. Then, for each individual, the prediction can be updated considering 2 potential situations, i.e., the patient receives the treatment and the patient does not receive it. The difference in outcome between these 2 potential situations can be used to identify the best treatment option. This "substitution-estimator" approach, which relies on the definition of counterfactual treatments and outcomes, is derived from the causal inference theory [53]. At the population level, the typical causal quantity of interest is often the average treatment effect of one treatment against the other, which can be simply obtained by averaging the difference in potential outcomes at the individual level across all subjects in the population. Another interesting quantity is the variability of the difference in potential outcomes at the individual level across all subjects in the population, which reflects the heterogeneity in response to the treatment among patients. If the latter quantity is non-null, one might be interested in identifying which characteristics are associated with response to the treatment, in order for the clinician to better tailor his/her therapeutic decision. Whatever the causal quantity, the use of appropriate causal inference techniques can be very useful to ensure that conclusions drawn can indeed answer the scientific question actually of interest. In some sense, one can argue that data science, to go from pure prediction to explanation, must be informed by the causal inference framework.

Similar to predicting patient outcomes via an algorithm being updated by patient historical data, recent efforts have also lead to generalizing such online approaches to predicting treatment effect in real-time. Online prescriptive analytics is an extreme form of these efforts, where the goal is to learn causal quantities and high-dimensional parameters at the patient's level based on streaming monitoring data. We have already given a sense of how ML algorithms can be adapted for online and real-time learning to specific patients. Likewise, it appears that an online version of targeted minimum loss estimators [54,55] can be elaborated to estimate and infer in real-time causal quantities such as treatment effect at the patient level. Targeted minimum loss estimation (TMLE) refers to a broad estimation framework that facilitates the construction of robust, asymptotically efficient estimators

with desirable finite-sample properties translating into lower bias and variance in finite samples [55] [56]. TMLE is a multi-step procedure and relies on estimates of different portions of the data distribution. In the simplest case of a binary outcome and exposure, in order to estimate the average treatment effect (ATE), TMLE uses information in the estimated treatment prediction model, g(a|W) = *P(A=a|W)* to update the initial estimator of the outcome prediction model: *Q(A,W) = E (Y|A,W)*. The updating step is used to achieve a targeted bias reduction for the parameter of interest, i.e. the ATE. Another virtue is that TMLE will be consistent if either the outcome prediction model or the exposure mechanism is estimated consistently. For TMLE, the word online refers to the facts that online ML is carried out and that the additional targeting step is online too. This extension comes at a price; the theoretical requirements are stronger when inferring from a single time series at the individual level than from independent data at the population level. The conditional independence, and existence of common underlying model over time are minimal requirements for identifiability of causal effects.

CONCLUSION AND PERSPECTIVES

Unlike pharmacogenomics, the vision of personalized medicine proposed in this article leverages the emergence of big medical data to appropriately estimate treatment effect at the patient level. This paradigm relies on several components, including big data acquisition and storage, online ML and robust statistical estimation that were made available thanks to the recent innovation in both data science, ML and statistics. Moving from *predictive* to *prescriptive* analytics, i.e., from predicting the future to accurately predicting the impact of competing medical decision, implies additional difficult challenges. The first one is to improve data quality and standardization. Indeed, issues involving data inaccuracy, including the problem of noise for continuous physiological signal monitoring, and lack of data remain substantial concerns. Therefore, it is critical to pursue our efforts in terms of signal processing, while at the same time insisting on the education of healthcare professionals to systematically check signal quality at the bedside. A second substantial challenge is to quantify how uncertain predictions are. Indeed, providing the clinician with a measure of uncertainty is important for him/her to integrate this piece of information into his/her decision process.

Mirroring the extraordinary penetration of artificial intelligence in every facet of our daily life, expectations are very high regarding the potential of ML and big data to improve patient care. However, because of poor interoperability between platforms, legal barriers and questions of data reliability, only a small fraction of the gigabytes of data generated in the ICUs are accessible for research. Thus, the true promise of data analytics in critically ill patients will come when big data generated from bedside monitors and electronic medical

records will be made available and when ML algorithms will be placed into a statistical framework able to produce trustworthy estimations and confidence regions. In addition, cloud computing as well as innovative techniques implemented for distributed computing now enable hospitals to access such technology without the need for local complex IT systems. This should ultimately make possible the deployment of online learners, possibly embedded into the clinical monitoring devices at the bedside (such as the architecture described in Figure 4). This is the true promise of acute care personalized medicine, and the time is right to reach for this promise.

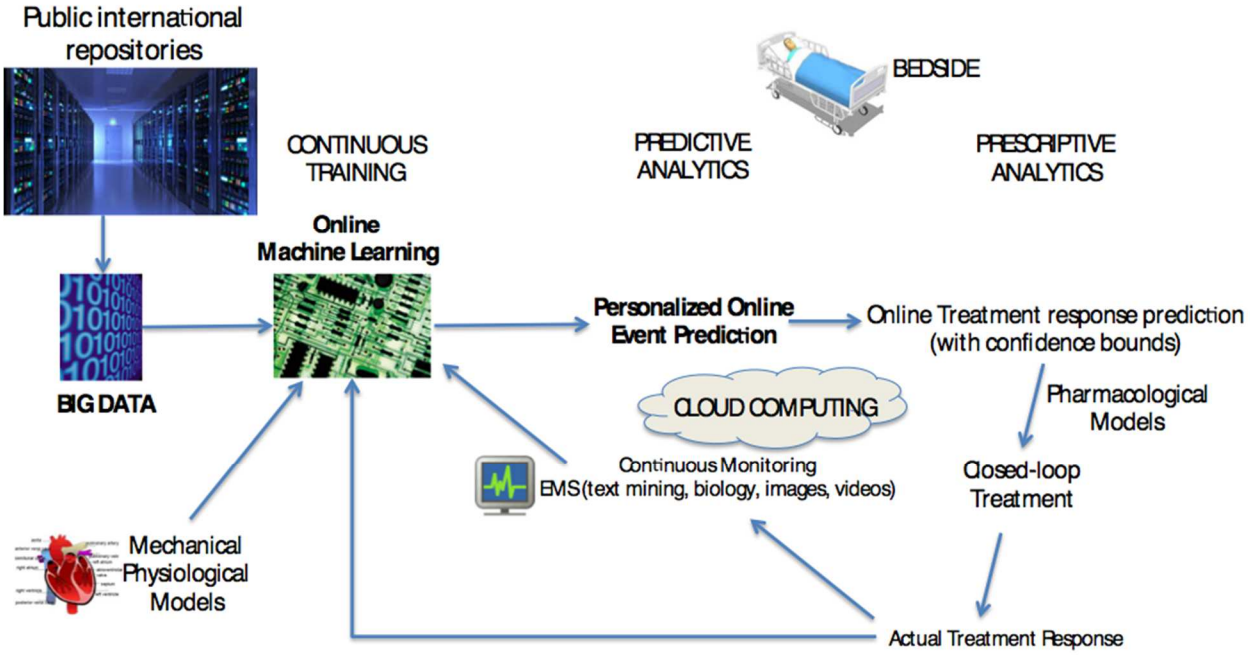

FIGURE 4: FORTHCOMING ARCHITECTURE FOR PERSONALIZED ONLINE PREDICTIVE AND PRESCRIPTIVE ANALYTICS IN ACUTE CARE.

REFERENCES

[1] Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015;372:793–5.

[2] Frank L, Ethan Basch, Selby JV. The PCORI perspective on patient-centered outcomes research. Jama 2014;312:1513–4.

[3] Food and Drug Administration. Paving the way for personalized medicine: FDA's role in a new era of medical product development. 2013. Wash DC US Dep Health Hum Serv 2015.

[4] Johnson JA. Pharmacogenetics: potential for individualized drug therapy through genetics. TRENDS Genet 2003;19:660–6.

[5] Murphy SN, Mendis ME, Berkowitz DA, Kohane I, Chueh HC. Integration of clinical and genetic data in the i2b2 architecture. AMIA Annu Symp Proc AMIA Symp 2006:1040.

[6] Lee J, Scott DJ, Villarroel M, Clifford GD, Saeed M, Mark RG. Open-access MIMIC-II database for intensive care research. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Conf 2011;2011:8315–8. doi:10.1109/IEMBS.2011.6092050.

[7] Herasevich V, Pickering BW, Dong Y, Peters SG, Gajic O. Informatics infrastructure for syndrome surveillance, decision support, reporting, and modeling of critical illness. Mayo Clin Proc 2010;85:247–54. doi:10.4065/mcp.2009.0479.

[8] van de Klundert N, Holman R, Dongelmans DA, de Keizer NF. Data Resource Profile: the Dutch National Intensive Care Evaluation (NICE) Registry of Admissions to Adult Intensive Care Units. Int J Epidemiol 2015;44:1850-1850h. doi:10.1093/ije/dyv291.

[9] Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D, et al. Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database. J Crit Care 2006;21:133–41. doi:10.1016/j.jcrc.2005.11.010.

[10] Saugel B, Vincent J-L, Wagner JY. Personalized hemodynamic management. Curr Opin Crit Care 2017;23:334–41. doi:10.1097/MCC.0000000000000422.

[11] Marik PE. Precision Glycemic Control in the ICU. Crit Care Med 2016;44:1433–4. doi:10.1097/CCM.0000000000001683.

[12] Ware LB, Koyama T, Zhao Z, Janz DR, Wickersham N, Bernard GR, et al. Biomarkers of lung epithelial injury and inflammation distinguish severe sepsis patients with acute respiratory distress syndrome. Crit Care Lond Engl 2013;17:R253. doi:10.1186/cc13080.

[13] Calfee CS, Janz DR, Bernard GR, May AK, Kangelaris KN, Matthay MA, et al. Distinct molecular phenotypes of direct vs indirect ARDS in single-center and multicenter studies. Chest 2015;147:1539–48. doi:10.1378/chest.14-2454.

[14] Wilson JG, Liu KD, Zhuo H, Caballero L, McMillan M, Fang X, et al. Mesenchymal stem (stromal) cells for treatment of ARDS: a phase 1 clinical trial. Lancet Respir Med 2015;3:24–32. doi:10.1016/S2213-2600(14)70291-7.

[15] Famous KR, Delucchi K, Ware LB, Kangelaris KN, Liu KD, Thompson BT, et al. Acute Respiratory Distress Syndrome Subphenotypes Respond Differently to Randomized Fluid Management Strategy. Am J Respir Crit Care Med 2017;195:331–8. doi:10.1164/rccm.201603-0645OC.

[16] Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci Transl Med 2015;7:287ra71. doi:10.1126/scitranslmed.aaa5993.

[17] Wong HR, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Willson DF, et al. Identification of pediatric septic shock subclasses based on genome-wide expression profiling. BMC Med 2009;7:34. doi:10.1186/1741-7015-7-34.

[18] Maslove DM, Tang BM, McLean AS. Identification of sepsis subtypes in critically ill adults using gene expression profiling. Crit Care Lond Engl 2012;16:R183. doi:10.1186/cc11667.

[19] Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, et al. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. Lancet Respir Med 2016;4:259–71. doi:10.1016/S2213-2600(16)00046-1.

[20] Walley KR, Thain KR, Russell JA, Reilly MP, Meyer NJ, Ferguson JF, et al. PCSK9 is a critical regulator of the innate immune response and septic shock outcome. Sci Transl Med 2014;6:258ra143. doi:10.1126/scitranslmed.3008782.

[21] Sapru A, Liu KD, Wiemels J, Hansen H, Pawlikowska L, Poon A, et al. Association of common genetic variation in the protein C pathway genes with clinical outcomes in acute respiratory distress syndrome. Crit Care Lond Engl 2016;20:151. doi:10.1186/s13054-016-1330-5.

[22] Wong HR, Atkinson SJ, Cvijanovich NZ, Anas N, Allen GL, Thomas NJ, et al. Combining Prognostic and Predictive Enrichment Strategies to Identify Children With Septic Shock Responsive to Corticosteroids. Crit Care Med 2016;44:e1000-1003. doi:10.1097/CCM.0000000000001833.

[23] Russell JA. Genomics and pharmacogenomics of sepsis: so close and yet so far. Crit Care Lond Engl 2016;20:185. doi:10.1186/s13054-016-1374-6.

[24] Levin MA, Wanderer JP, Ehrenfeld JM. Data, Big Data, and Metadata in Anesthesiology. Anesth Analg 2015;121:1661–7. doi:10.1213/ANE.0000000000000716.

[25] Duffy J. Personalizing health care through big data. Johns Hopkins Mag 2015.

[26] Halamka JD. Using Big Data to Make Wiser Medical Decisions. Harv Bus Rev 2015.

[27] Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, et al. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. JAMA Cardiol 2017;2:204–9. doi:10.1001/jamacardio.2016.3956.

[28] Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD. Predicting 30-day readmissions with preadmission electronic health record data. Med Care 2015;53:283–9.

[29] Pinsky MR, Clermont G, Hravnak M. Predicting cardiorespiratory instability. Crit Care Lond Engl 2016;20:70. doi:10.1186/s13054-016-1223-7.

[30] Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med 2015;3. doi:10.1016/S2213-2600(14)70239-5.

[31] Pirracchio R, Ranzani OT. Recalibrating our prediction models in the ICU: time to move from the abacus to the computer. Intensive Care Med 2014;40:438–41. doi:10.1007/s00134-014-3231-x.

[32] Lai PS, Shrime MG, Ferket BS, Scott DJ, Lee J, Celi LA, et al. Using Markov Models To Determine The Optimal Duration For A Trial Of Intensive Care In Patients With Active Cancer And Septic Shock. C24

THIS CAN BE DONE BETTER Qual. Saf. PROCESS Improv. Crit. CARE, American Thoracic Society; 2012, p. A3998–A3998. doi:10.1164/ajrccm-conference.2012.185.1_MeetingAbstracts.A3998.

[33] Cismondi F, Celi LA, Fialho AS, Vieira SM, Reti SR, Sousa JMC, et al. Reducing unnecessary lab testing in the ICU with artificial intelligence. Int J Med Inf 2013;82:345–58. doi:10.1016/j.ijmedinf.2012.11.017.

[34] Moss TJ, Clark MT, Calland JF, Enfield KB, Voss JD, Lake DE, et al. Cardiorespiratory dynamics measured from continuous ECG monitoring improves detection of deterioration in acute care patients: A retrospective cohort study. PloS One 2017;12:e0181448. doi:10.1371/journal.pone.0181448.

[35] Hravnak M, Devita MA, Clontz A, Edwards L, Valenta C, Pinsky MR. Cardiorespiratory instability before and after implementing an integrated monitoring system. Crit Care Med 2011;39:65–72. doi:10.1097/CCM.0b013e3181fb7b1c.

[36] Hravnak M, Edwards L, Clontz A, Valenta C, Devita MA, Pinsky MR. Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system. Arch Intern Med 2008;168:1300–8. doi:10.1001/archinte.168.12.1300.

[37] Tarassenko L, Hann A, Young D. Integrated monitoring and analysis for early warning of patient deterioration. Br J Anaesth 2006;97:64–8. doi:10.1093/bja/ael113.

[38] Lipton ZC, Kale DC, Elkan C, Wetzell R. Learning to diagnose with LSTM recurrent neural networks. ArXiv Prepr ArXiv151103677 2015.

[39] Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PloS One 2017;12:e0174708. doi:10.1371/journal.pone.0174708.

[40] Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. JMIR Med Inform 2016;4:e28. doi:10.2196/medinform.5909.

[41] Jiang D, Peng C, Chen Y, Fan Z, Garg A. Probability distribution pattern analysis and its application in the Acute Hypotensive Episodes prediction. Measurement 2017;104:180–91. doi:10.1016/j.measurement.2017.03.030.

[42] Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine learning in medical imaging. IEEE Signal Process Mag 2010;27:25–38.

[43] Hatib F, Jian Z, Buddi S, Lee C, Settels J, Sibert K, et al. Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis. Anesthesiology 2018. doi:10.1097/ALN.0000000000002300.

[44] Van Poucke S, Zhang Z, Schmitz M, Vukicevic M, Laenen MV, Celi LA, et al. Scalable Predictive Analysis in Critically Ill Patients Using a Visual Open Data Analysis Platform. PloS One 2016;11:e0145791. doi:10.1371/journal.pone.0145791.

[45] Che Z, Purushotham S, Khemani R, Liu Y. Interpretable Deep Models for ICU Outcome Prediction. AMIA Annu Symp Proc AMIA Symp 2016;2016:371–80.

[46] Pirracchio R, Yue JK, Manley GT, van der Laan MJ, Hubbard AE. Collaborative targeted maximum likelihood estimation for variable importance measure: Illustration for functional outcome prediction in mild traumatic brain injuries. Stat Methods Med Res 2018;27:286–97. doi:10.1177/0962280215627335.

[47] Lee J, Maslove DM, Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. PloS One 2015;10:e0127428. doi:10.1371/journal.pone.0127428.

[48] Bai Y, Sow D, Vespa P, Hu X. Real-Time Processing of Continuous Physiological Signals in a Neurocritical Care Unit on a Stream Data Analytics Platform. Acta Neurochir Suppl 2016;122:75–80. doi:10.1007/978-3-319-22533-3_15.

[49] van Otterlo M, Wiering M. Reinforcement learning and markov decision processes. Reinf. Learn., Berlin Heidelberg: Springer; 2012, p. 3–42.

[50] Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? ArXiv Prepr ArXiv151106348 2015.

[51] van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6:1–21. doi:10.2202/1544-6115.1309.

[52] Hochreiter S, Schmidhuber J. LSTM can solve hard long time lag problems, 1997, p. 473–9.

[53] Pearl J. Causality. Cambridge University Press; 2009.

[54] van der Laan MJ, Rubin D. Targeted maximum likelihood learning. Int J Biostat 2006;2.

[55] Van der Laan MJ, Rose S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; 2011.

[56] Bühlmann P, Drineas P, Kane M, van der Laan M. Handbook of Big Data. CRC Press; 2016.