Brain and Spine 4 (2024) 102858

Contents lists available at ScienceDirect

Brain and Spine

journal homepage: www.journals.elsevier.com/brain-and-spine

# From bed to bench and back again: Challenges facing deployment of intracranial pressure data analysis in clinical environments

Laura Moss [a,b,*], Martin Shaw [a,b], Ian Piper [b], Christopher Hawthorne [c]

[a] Dept. of Clinical Physics, NHS Greater Glasgow and Clyde, Glasgow, United Kingdom
[b] College of Medicine, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom
[c] Dept. of Neuroanaesthesia, Institute of Neurological Sciences, NHS Greater Glasgow and Clyde, Glasgow, United Kingdom

ARTICLE INFO

Handling Editor: Dr W Peul

*Keywords:*
Intracranial pressure
Machine learning
Big data

ABSTRACT

*Introduction:* Numerous complex physiological models derived from intracranial pressure (ICP) monitoring data have been developed. More recently, techniques such as machine learning are being used to develop increasingly sophisticated models to aid in clinical decision-making tasks such as diagnosis and prediction. Whilst their potential clinical impact may be significant, few models based on ICP data are routinely available at a patient's bedside. Further, the ability to refine models using ongoing patient data collection is rare. In this paper we identify and discuss the challenges faced when converting insight from ICP data analysis into deployable tools at the patient bedside.

*Research question:* To provide an overview of challenges facing implementation of sophisticated ICP models and analyses at the patient bedside.

*Material and methods:* A narrative review of the barriers facing implementation of sophisticated ICP models and analyses at the patient bedside in a neurocritical care unit combined with a descriptive case study (the CHART-ADAPT project) on the topic.

*Results:* Key barriers found were technical, analytical, and integrity related. Examples included: lack of interoperability of medical devices for data collection and/or model deployment; inadequate infrastructure, hindering analysis of large volumes of high frequency patient data; a lack of clinical confidence in a model; and ethical, trust, security and patient confidentiality considerations governing the secondary use of patient data.

*Discussion and conclusion:* To realise the benefits of ICP data analysis, the results need to be promptly delivered and meaningfully communicated. Multiple barriers to implementation remain and solutions which address real-world challenges are required.

## 1. Background

Intracranial pressure (ICP) monitoring can capture waveform quality data. The analysis of these ICP waveforms is of interest within neurocritical care settings and in the management of neurological disorders such as hydrocephalus. Sustained periods of raised ICP can be detrimental to a patient and prompt treatment aims to keep the mean ICP below around 20 mmHg (Carney et al., 2017), although this threshold will differ between patient populations and is likely to require individualisation.

To guide patient treatment or predict a patient's status or outcome, ICP has been analysed using a range of techniques or combined with other parameters to create numerous indices and couplings. One such clinically useful index is cerebral perfusion pressure (CPP). CPP is the pressure difference between the mean arterial pressure (MAP) and the venous outflow pressure, considered equivalent in pressure to the intracranial pressure (ICP). The RAP index is the correlation coefficient between the mean ICP and ICP wave amplitude and provides a measure of the pressure-volume reserve capacity (Balestreri et al., 2004). ICP values can also be used in models which represent a measure of the patient's cerebrovascular reactivity (i.e. reflecting a patient's autoregulatory state). For example, the pressure reactivity index (PRx) is a Pearson's correlation between arterial blood pressure (ABP) and ICP (Czosnyka et al., 1997). PRx is calculated with data every 6 seconds and

* Corresponding author. Room 2.41 Level 2, New Lister Building, Glasgow Royal Infirmary, 10-16 Alexandra Parade, Glasgow, G31 2ER, United Kingdom.
*E-mail addresses:* laura.moss@glasgow.ac.uk (L. Moss), martin.shaw@ggc.scot.nhs.uk (M. Shaw), Ian.piper@brainit.org (I. Piper), cwhawthorne@doctors.org.uk (C. Hawthorne).

https://doi.org/10.1016/j.bas.2024.102858
Received 14 June 2023; Received in revised form 29 May 2024; Accepted 3 July 2024
Available online 4 July 2024

for an overall duration of between 1 min and 3 min. Low resolution autoregulation index (LAx) is also a Pearson correlation between ICP and ABP (Depreitere et al., 2014). However, it is based on the more readily available data frequency of per minute, with the standard correlation window duration of between 10 and 30 min. CPPOpt (Aries et al., 2012) calculates the CPP required for a patient's optimal autoregulatory state. CPPOpt is achieved through a process of data thinning and collation, resulting in a quadratic polynomial linear regression model fit; the optimal value is found at the minimum turning point of the fit, this is calculated over a duration of 4 hours which is then moved through time and the estimate updated. The dynamic adaptive target of active cerebral autoregulation (DATACAR) methodology (Depreitere et al., 2014) was proposed to address the situation where no turning point can be found due to lack of data. DATACAR is a more computationally expensive methodology and repeatedly updates window lengths (for which autoregulation is to be calculated and collection of CPP values) to cover a spectrum of data possibilities to maximise the chance of finding the optimal point.

In addition to using the absolute ICP value, the morphology of the ICP waveform can be examined. The ICP waveform is a tripartite structure, comprising of a percussive wave (P1), a tidal wave (P2), and a dicrotic wave (P3) (Cardoso et al., 1983). The P1 wave is indicative of arterial pulsations, the P2 wave represents intracranial compliance, and the P3 wave corresponds to the closure of the aortic valve. Normally P1>P2>P3, however, deviations in these waves may be suggestive of pathological conditions, such as raised ICP or impaired cerebral compliance. When displayed against time, patterns can also be observed at a macroscopic level. These macroscopic patterns are known as Lundberg waves and defined as A-waves, B-waves, and C-waves (Lundberg, 1960). A-waves are considered as plateau waves occurring during very high ICP and associated with a poor prognosis, B-waves are short-duration ICP elevations with variable pressure levels, possibly representing poor compliance and C-waves have more frequent elevations up to around 30 mmHg which are associated with respiratory and cardiac cycles. The Morphological Clustering and Analysis of ICP (MOCAIP) algorithm performs automated analysis of the ICP waveform and can identify and measure non-artifactual ICP peaks, designating the three previously defined sub-components of the ICP pulse (Hu et al., 2009).

More recently, techniques such as machine learning are being used in ICP analysis. Machine learning is a subfield of artificial intelligence (AI). The aim of machine learning is the development of algorithms which extract patterns (or models) from datasets which can then be applied to other (unseen) data to aid in clinical tasks such as classification, prediction, and prognosis (Shillan et al., 2019). Whilst such approaches often produce models with high performance, many use 'black box' techniques which produce complex models for which a clinician has very little (or no) ability to understand the process applied or the resulting model (Arrieta et al., 2020).

Currently, the use of ICP monitoring data with machine learning approaches has mostly focused on the creation of models which predict ICP or use ICP as a variable in a prediction model (e.g. to predict mortality). An example of the former are recurrent neural networks, which have been used by both Ye at el (Ye et al., 2022) and Schweingruber at el (Schweingruber et al., 2022) to build models to predict ICP. In another example, Teplan et al. (2017) have used a hierarchical Gaussian Mixture Model (hGMM) for clustering extracted features which represent ICP time sub-sequences, the application of this work has been for the creation of optimum thresholds for ICP alarms. With respect to ICP as a variable in a prediction model, examples include the prediction of intracranial hypertension (Lee et al., 2021), identification of patients for a permanent cerebral spinal fluid shunt implantation (Mládek et al., 2022), and prediction of patient outcome (Rajagopalan et al., 2022), (Raj et al., 2019).

Whilst the potential clinical impact of indices and models derived from ICP measurements may be significant, few are routinely available for clinical use at the patient's bedside (Carra et al., 2020). Further, the ability to refine models using ongoing patient data collection is rare.

There are some examples which move ICP analysis closer to the patient bedside. The ICM+ software enables real-time multimodality monitoring, analysis, and data storage in neurological intensive care environments (ICM+). Several relevant indices, such as PRx, RAP and CPPOpt, can be calculated when using the software. However, the software is not available for routine clinical use (just for research purposes). The CHART-ADAPT platform (further details in Section 3) demonstrated the possibility of implementing complex models (such as those from machine learning), in real-time, at the patient bedside by providing sufficient computational infrastructure and integrating with existing patient monitoring equipment (Moss et al., 2021). This study implemented the following: PRx, LAx, HMF, CPPOpt, and DATACAR and machine learning models are planned. However, again, this platform is not widely available.

Whilst there has yet to be a formal review of the clinical implementation of ICP models resulting from machine learning approaches, Citerio has identified and discussed the lack of routine integration of AI and machine learning approaches within neurocritical care settings and identifies that whilst models concerning ICP exist, they have not been integrated into clinical reasoning and the selection of strategies is not guided by AI (Citerio, 2022). This gap between the known capabilities of AI/machine learning and its full integration into a real-world application is often referred to as the "AI chasm" (Aristidou et al., 2022) and is reflected more widely in medicine. Adegboro et al. reviewed the use of AI techniques to improve patient outcomes in the neonatal and paediatric critical care settings and showed that few models were in an implementation phase or deployed within an intensive care unit (ICU) environment (Adegboro et al., 2022). Further, van de Sande reviewed the use of AI in critical care and out of over 400 studies, the vast majority remained in testing and prototype stages; none reported outcome of an AI model integrated into routine clinical practice (van de Sande et al., 2021).

In this paper we identify and discuss some of the challenges faced when converting insight from data analysis (in particular, machine learning) into deployable knowledge at the patient bedside; Section 2 explores existing literature on the topic and Section 3 details a relevant case study in which particular issues regarding implementation of models were identified and subsequently resolved.

## 2. Barriers to implementation of complex ICP models

There are many different reasons why a promising clinical model may not successfully transfer into a clinical environment. These may be highly dependent on the individual hospital environment (e.g. local policies, finances), but most reasons are ubiquitous and can typically be classified into three broad reasons: technical, analytical and integrity. In the following sections these categories are explored from the perspective of complex ICP modelling and its implementation within ICP related clinical contexts.

### 2.1. Technical barriers to model implementation

#### 2.1.1. Data collection & integration

A lack of ICP data harmonization can be both an impediment to offline model creation and online model deployment.

Many models or indices derived from, or relevant to ICP, include parameters which are collected from different monitoring devices. These devices often use assorted platforms and output data in various formats, diverse devices can also have different internal time clocks which makes data synchronization very difficult (van de Sande et al., 2021), devices generally do not have standardized device interfaces which enable extraction of data (Alkhachroum et al., 2022).

Another data collection challenge is the availability of patient data to be processed in real time. In most units, patient data will be recorded in

real time and stored in a patient management system. Timely access to this data is required and there are multiple ways in which this could be achieved. For example, live patient data could be copied to a secondary dataset for models to process the data, or devices, such as ICM+ can be installed to enable data to be extracted from these systems and be provided for analysis.

With regards to data integration, there are no widely accepted standard data formats for ICP recordings and associated neurocritical care data (Alkhachroum et al., 2022). Integrating data from other specialities (e.g. radiology) is also a significant challenge, again due to differing data types and informatics architectures (van de Sande et al., 2021). Consequently, ICP data can be poorly integrated. Software such as Moberg's Component Neuromonitoring System (CNS Monitor) and ICM+ have improved the integration of some of the relevant data sources. However, each still has its own storage format (although ICM+ can output data in an open file format, HDF5). Agreement across the community on the data formats/standards to use or development of translations between standards would help with integration of data.

Additionally, there is not always standardisation of what data to collect. For example, annotations of clinically relevant events which may affect interpretation of ICP values recorded in the data (e.g. turning a patient) (van de Sande et al., 2021). If agreement can be found with respect to the structure and content of the data, then it would make development and implementation of models at the bedside easier. However, this is not an easy task, nor a new endeavour. Common Data Elements (CDE) which define useful information to collect have been specified as part of some initiatives such as BrainIT (BrainIT) and CENTER-TBI (CENTER-TBI) but are not necessarily routinely applied outside of these initiatives. Further, over time and with differing research priorities, even established CDEs will require updating.

Data harmonization is not only hindered by differences in storage formats and structure, it can also be affected by differences in the meaning of the variables in the data (i.e. data semantics). For example, how to identify the same variables when different terms have been used, e.g. MAP vs. Mean Arterial Pressure. Semantic data integration enables the integration of different data based on the meaning of the data (Cheatham et al., 2017). This is often done by using information about the variables and the relationships between them which can be represented in ontologies. Ontologies provide a computational representation of a domain describing the elements in that domain and the relationships between them.

In the approaches discussed so far, the aim has been to pull together data from different sources to create one dataset for analysis or for the application of a model to. However, another approach is that the data is instead kept local, and analysis is performed in a distributed manner with results aggregated (can also be considered as federated learning). Whilst this approach may resolve some issues (e.g. permissions to transfer data), it can still be affected by differences in semantic meaning, structure of the data and differing data formats.

#### 2.1.2. Infrastructure

Many clinical environments do not have the technical infrastructure for the storage and real-time processing of large volumes of ICP data required by complex models (such as those involving waveform analysis). This makes it difficult to deploy new complex models on existing equipment at the patient bedside, as analysis of live data in clinically meaningful timescales requires sufficient bandwidth and processing capacities. Instead, high-resolution data (e.g. waveform data) is often integrated and visualised through third-party enterprise solutions, which are increasingly being located off-site. These platforms may provide technical solutions, but issues surrounding privacy, security, and access, can be controversial and need resolving (van de Sande et al., 2021), (Sanchez-Pinto et al., 2018). High performance infrastructure can also be expensive and therefore prohibitive in resource constrained settings.

Models are often created using large datasets consisting of data purposely combined offline from multiple sources. However, if the data in the clinical application of the model is not collated or available in the same way, then the model cannot be deployed or will not perform as well (due to reduced availability of parameters). Additionally, it is not desirable to have to re-code a model each time it is deployed in a new clinical environment because different equipment is used.

One solution to this problem is to centralise the model and decouple the model from the data (Richards, 2022). Decoupling the model from the data is a concept in which the layers of a software architecture, i.e. in this simple example, the model layer and the data layer, are made independent of each other. Often this requires the creation of a translation 'layer' which sits between the model and the data. For example, this translation layer could map between a variable in a new dataset (e.g. Mean Arterial Pressure) and the variable defined in the model (e.g. MAP). Decoupling means that any layer of the architecture can be changed and/or expanded without having to significantly change the other layers. For example, an improved model could be implemented in the model layer, but it would require no changes to the way it interacts with the data layer. This means that models only have to be implemented once and multiple healthcare settings, equipment, and providers could apply that model simultaneously to their data stream. This also makes it easier for clinical trials of models. Refinement of models can take place in real-time as the data flows through the platform and each user has instant access to the most up-to-date model. Although a unit would have access to a centrally deployed model, they would still be able to make their own decisions.

### 2.2. Analytical barriers to model implementation

#### 2.2.1. Clinical utility

It has been suggested that one of the reasons why models may not transfer well to clinical environments is the lack of clinical utility of the developed model (Carra et al., 2020). A plethora of machine learning models are published, but without their real clinical impact studied, their advancement will only be minimal (Bellini et al., 2022). Where the clinical impact has been explored, the impact of AI has been shown to be limited; in a review by Zhou et al., 40% of randomised controlled trials (RCT) evaluating AI-based clinical interventions found there was no clinical benefit of using AI prediction tools compared to routine standard of care; this was often despite a model having good area under the receiver operator characters (AUROC) in model development and evaluation (Zhou et al., 2021).

Another aspect of a model's suitability for clinical application is whether interaction with the clinicians using the model has been considered. Clinicians should be considered in the development and evaluation of AI-driven models, otherwise there is a risk that models do not reflect the required needs of clinicians (and hence there is then little motivation from clinicians to ensure they are implemented). As argued by Citerio (2022), neurointensivists and neurointensive care unit nurses are central to the delivery of patient care and for models to be successfully adopted in the neurocritical care environment, the models should reflect this by augmenting their role rather than aiming to replace it. Future human factors research into the interaction between clinicians and models is required.

#### 2.2.2. Generalizability

For a model to be accepted and routinely used in local clinical practice, the model will need to demonstrate adequate generalizability across different patient populations and settings. Therefore, it is important to understand the data used to train and evaluate a model. Performance of a machine learning model is likely to worsen when the data to which it is to be applied to differs from the data used to train it (Adegboro et al., 2022). Models are generally trained on a specific patient population and do not necessarily capture heterogeneity in real-world patient data (Carra et al., 2020). To ensure a model performs as well on unseen data, external validation of the model should be

performed. Alkhachroum et al. (2020) make several points on this topic: external evaluation should be executed on prospective datasets and not retrospective clinical data, this point is also made by Adegboro et al. (2022); using data from RCTs could introduce selection bias due to inclusion and exclusion criteria and therefore carefully curated datasets have less value for machine learning validation; and testing using diverse datasets, ideally from multi-centre cohorts, is essential to external validation studies, examples of these within ICP research include; BrainIT (BrainIT), The Collaborative European Neuro Trauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) and Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI).

#### 2.2.3. Data quality

Models created in research settings generally use high quality data, curated specifically for a research study. However, when such models are transferred to clinical practice, this level of data quality does not always reflect reality and real-world, unfiltered patient data, can be incomplete, inaccurate, and unlabelled (Orphanidou, 2019); this can make model deployment difficult or impossible. Data quality can vary depending on monitoring equipment and its placement and maintenance. Artifacts are a particular challenge when using ICP data. On average between 5% and 20% of ICP data contains artifacts (Feng et al., 2011). Removal of artifacts in ICP data is a significant challenge. Manual removal is impractical, can introduce variability through different analysts and is not appropriate during the live deployment of models. Automated, online, ICP artifact removal has been studied (Feng et al., 2011), (Lee et al., 2019), (Feng et al., 2012), (Martinez-Tejada et al., 2021), (Megjhani et al., 2019) but, as yet, is not routinely implemented in clinical settings.

### 2.3. Integrity barriers to model implementation

An absence of consideration for model transparency, reproducibility, ethics, and effectiveness can result in a lack of translation of clinically promising models into clinical practice (Vollmer et al., 2020). Some of these topics are reflected upon in more detail in the following sections.

#### 2.3.1. Ethics

Ethical challenges and considerations regarding the deployment of models may, if not resolved, hinder implementation into a real-world clinical setting. Some ethical challenges are straightforward to resolve, whereas others, such as the reliance on AI as the source of medical knowledge raise potentially broad ethical concerns (Char et al., 2018). Further, the use of machine learning in complicated care practices will require ongoing consideration, since the correct diagnosis in a particular case and what constitutes best practice can be controversial (Char et al., 2018).

The impact of machine learning or automated decision making on the patient-clinician relationship (or family-clinician relationship) is an important factor to consider. Currently, this relationship is grounded in the principles of medical ethics. However, the introduction of automated decision making will extend it from a two-party relationship to a three-party relationship and introducing this new moral actor will challenge this existing grounding (Čartolovni et al., 2022). Furthermore, the introduction of automated decision making may exclude the patient from the decision-making process and lead to a more distant patient-doctor relationship (Al-Mufti et al., 2019).

#### 2.3.2. Trust

Lack of trust in a model can be a limiting factor for its widespread clinical acceptance. Generally, it has been suggested that to achieve an optimal level of trust in a machine learning model (or AI system) for use in healthcare environments, there are three factors which should be considered: fairness, transparency, and robustness (Asan et al., 2020). Whilst these are most likely universal concepts, further research is required into the factors specific to increasing trust between machine learning and AI in ICP settings and its impact on workflow (Adegboro et al., 2022).

#### 2.3.3. Fairness

Bias minimization is an important element of gaining trust in a machine learning model (Asan et al., 2020) (and consequently will help its adoption into a clinical environment). Bias can lead to unfair model outcomes, therefore, to ensure fair modelling, biases should be identified and resolved. There is concern that biases contained in training data will be reflected in the resulting model. The biases include those related to missing data and patients not identified by algorithms, sample size and underestimation, and misclassification and measurement error (Gianfrancesco et al., 2018). Bias can be introduced at all stages of model development and deployment. Vokinger et al. (2021) identify several ways in which bias can be introduced, for example, data used to train a model may be subject to sampling bias and may not reflect the patient population in which it is to be applied; modelling may propagate existing bias in the dataset as performance metrics concentrate on how good its predictions are on the average population, possibly at the cost of lower performance on underrepresented groups; model evaluation may be performed inadequately whereas it should examine how well the model performs across all groups of patients with errors examined; finally, when a model is deployed, it may not be applied to the cohort of patients for which it was designed for (known as domain shift).

#### 2.3.4. Transparency

A lack of transparency can lead to reduced trust in a model. As suggested by Dhar et al., "it is unlikely that we will accept a machine's decision to perform an invasive intervention on a patient without understanding what factors are driving this alert" (Dhar and Meyfroidt, 2022).

One route to improving transparency is for effective evaluation and reporting of machine learning studies (Carra et al., 2020). Recent initiatives to improve such reporting include the CONSORT-AI and SPIRIT-AI extensions; both make recommendations of the items which should be routinely reported for studies involving AI (Liu et al., 2020), (Rivera et al., 2020).

Models derived from machine learning (or other complex analyses) often perform well but the techniques used lead to models which tend to be less intuitive and cannot be fully interpreted by clinicians. For example, a neural network may consist of layers of artificial neurons with many thousands of parameters. With an understanding of neural networks, it may be possible to conceptually explain what is happening in the layers, however it is impossible to understand the computation and explain how the parameters worked together to generate the prediction. These so called 'black box' models may be problematic for implementation in clinical environments and such tools cannot be questioned for their medicolegal implications (Nicholson Price Glenn Cohen et al., 2018). Further, legislation such as the European Union directive General Data Protection Regulation (GDPR) (EU Regulation (EU), 2016), generate a duty of transparency or explainability from decision making using personal data and hence models used in a clinical context, in particular for fully automated decisions, need to provide an explanation to enable a patient to obtain meaningful information about the logic used, express their point of view, and contest a decision; this can be difficult to achieve using black box models.

It is suggested that to increase clinical acceptance of decision-making tools based on such black box algorithms, access should be increased to the information driving the algorithm (Alkhachroum et al., 2020). Consequently, there is growing interest in the field of interpretable machine learning. There are three main approaches to developing interpretable models: firstly, models, such as decision trees, which are intrinsically interpretable by a clinician can be used, secondly, interpretation methods can be applied after a model has been created, and thirdly the behaviour of a model could be explained using

example-based methods. For a discussion on interpretable machine learning in neurocritical care see Moss et al. (2022). The level of interpretability required for a model to be considered as trustworthy will be highly dependent on the context in which it is being applied and on centre-specific policies (Carra et al., 2020).

#### 2.3.5. Regulatory requirements

Absence of sufficient organisational and regulatory frameworks may also be a contributing factor to a lack of model deployment at a patient's bedside. For countries in which regulatory frameworks do exist, the procedures may be perceived as overly bureaucratic, time consuming, and expensive. It is unlikely that those involved in the creation of models in a research capacity have the resources to complete a regulatory process to enable the model to be widely used as a routine clinical decision support tool. Therefore, the onus will be on the relatively few ICP medical device and software companies to shoulder that financial burden and to develop a model commercially.

A further consideration is that regulatory standards will need to be updated to consider the use of machine learning/artificial intelligence which can independently learn and evolve.

In addition, even if a model is made clinically available, there are still important legal considerations which require discussion. One of the main identified limitations to the application of machine learning into clinical environments is the accountability if a machine learning model makes an error (Chaudhry et al., 2020). For example, Al-Mufti et al. (2019), questions who is accountable when machine learning models (or AI driven systems) decide to intervene on a patient and cause an adverse event? Furthermore, if an error is made, it would currently be difficult to identify the source of the error due to the black box nature of the models.

## 3. CHART-ADAPT case study

The CHART-ADAPT (Connecting Healthcare And Research Through A Data Analysis Provisioning plaTform) project was the first to demonstrate a realistic approach to actionable analytics at the bedside of a single Scottish adult neurointensive care unit. The successfully implemented platform facilitated, as a proof of concept, the online, bedside, deployment of models and analysis of data. In this section we describe experiences from the research project and how some of the challenges faced in converting promising clinical models into meaningful clinical impact were addressed in the hope that it may be useful for those about to encounter similar issues. The aim of the CHART-ADAPT project was to demonstrate proof of concept and was considered a research project. The CHART-ADAPT platform is not a product and is not commercially available.

CHART-ADAPT was implemented in the Neurocritical Care Unit, Institute of Neurological Sciences, Glasgow, UK. 831 patients, admitted to the unit between 1/4/2015 and 31/3/2017, were transferred through CHART-ADAPT, of which 34 included waveform data to demonstrate the platform's ability to handle waveform quality data. Clinical research studies were conducted as part of the platform's evaluation including a comparison of several models of cerebral autoregulation and calculation of optimal cerebral perfusion pressure (CPPOpt); some of these indices were previously restricted to use within research settings or not clinically applied due to their computational requirements.

As part of the platform, patient data was automatically integrated, de-identified and transferred to a securely hosted, cloud-based platform providing data storage, an analysis engine, and a specifically designed app to control applied analyses. Complex physiological models and algorithms (as specified by the control app), were applied to live patient data and the results passed back into the clinical environment, re-identified, and integrated with the existing patient management system for display at the patient's bedside.

Regulatory approval was required for CHART-ADAPT and consisted of Caldicott Guardian approval, NHS Research Ethics, and NHS Greater Glasgow and Clyde I.T approval from both technical and security perspectives. Acquiring these approvals took a considerable amount of the project's timeline and was due to the nature of approval processes and the novelty of the project at the time.

As part of these approval processes and to ensure patient confidentiality, the project developed fully configurable software (Automated Neurointensive Care Anonymisation (ANCA)) to process patient data in accordance with the relevant policy and legal requirements and automatically anonymise the patient data before it left the neurocritical care unit. A testing plan to identify lapses in confidentiality was followed and repeated at regular intervals. Additionally, an Information Governance Strategy was developed, making explicit the data handling and security procedures put in place and personnel responsible for information governance in both organisations (healthcare and commercial) worked closely together.

To overcome data harmonization issues, a solution adopted in the CHART-ADAPT platform was to decouple the model from the data. To enable this separation of concern in the CHART-ADAPT platform, established data standards were used for data input to the models; Medical waveform Format Encoding Rules (MFER) waveform format (ISO/TS11073-92001) (Takeuchi et al., 2009) and HL7 (Health Level 7) Version 2 (HL7). In the unit in which CHART-ADAPT was being demonstrated, this required the Philips Rhapsody system (Philips medical IB SC200 interoperability engine) to be configured to output HL7, and a custom piece of software written to convert the recorded waveform data into MFER. The entire flow of patient data was then automatically transformed into HL7. Although HL7 is already an accepted standard in clinical environments, it is acknowledged that this is not a perfect solution as it still puts the responsibility on the manufacturer to output data in an established format, however it does at least make the implementation of the model itself easier.

The CHART-ADAPT application domain also faced not having access to adequate technical infrastructure to run complex models. Using existing, local equipment, some of the ICP models implemented would not have run in clinically meaningful timescales. The choice made to resolve this issue was to transfer the anonymous patient data to a high-performance computing platform provided by an external commercial provider. The cloud-based service provision included Spark (Apache Software Foundationa), Hadoop (Apache Software Foundationb) and Greenplum (Greenplum). To highlight the possible performance gains from this approach, a comparison was made between implementing PRx in R versus Scala (Scala) (the language used for implementation on Spark). Experiments showed the runtime of R and Scala increasing linearly with data volume, in which Scala performed in the range of 50–200 times faster than the R code (Moss et al., 2016). It is acknowledged that the transfer of patient data from a clinical environment to an external provider can be challenging and not without confidentiality issues, but at the time it was the only feasible way to achieve the required processing power.

The aims of the CHART-ADAPT project required technical and analytical knowledge alongside specialist clinical knowledge of ICP monitoring and neurocritical care. Without input from all these perspectives, there was a potential for the platform (and models developed to run on it) to not be fit for purpose. To mitigate this concern, a multidisciplinary team was created for CHART-ADAPT consisting of neurocritical care clinicians, data scientists, clinical scientists (with specialisms in computational sciences) and a medical device manufacturer (Philips). Having access to people with different skills and viewpoints ensured that the platform was suitable for the clinical environment in which it was to be implemented, met the clinical requirements, incorporated into existing clinical pathways, and integrated properly with existing technology in the unit, enabling it to be used at the bedside without further devices having to be installed.

Another concern faced during the implementation of CHART-ADAPT was the lack of data science knowledge of the potential users in the neurocritical care unit. The creation and deployment of models requires knowledge of data science and of the computing infrastructure for which

it will be deployed to. Most clinical staff do not have these skillsets. Further, access to in-house expertise in these areas is sporadic. Therefore, to ensure clinicians were not excluded from the project (and hence increase adoption of the models/platform), a control app was developed which allowed clinicians to easily: select which patients should have their data passed to the platform, which models should be applied to each included patient, view the platform's technical performance, and define different user roles.

The role of commercial companies in the analysis of sensitive patient data is controversial and can be a barrier to model deployment and use. Due to the involvement of commercial partners in the CHART-ADAPT project, a survey was conducted to gain feedback on the topic (Kinsella et al., 2017). From this survey, most respondents felt that their medical data should be used for research purposes and would be happy to share their data. Additionally, most respondents trusted clinicians with this data, but when it came to the role of private companies in such studies, the response was mixed. To mitigate this possible barrier in the CHART-ADAPT project, engagement initiatives were implemented. For example, a public event was hosted to discuss patient data sharing within critical care, posters and leaflets about CHART-ADAPT were made available in the neurocritical care unit and staff were briefed and updated on project progress. Further, attendance at relevant academic and healthcare events provided opportunities to discuss the platform and gather feedback which was then fed back into the development of the project.

The types of challenges encountered in the CHART-ADAPT project mostly focused on technical and analytical barriers. Integrity barriers were not considered in the design of the platform. For example, interpretability and bias were not topics considered in the design of the CHART-ADAPT infrastructure. However, providing tools which help to counteract these problems would be important enhancements to the platform. For example, it would be possible to implement algorithms such as LIME (local interpretable model-agnostic explanations) (Ribeiro et al., 2016) on the CHART-ADAPT platform to automatically provide a local interpretable model to explain the individual predictions of a black box model. Further regulatory approvals in the project largely focused on the transfer of data. The platform was not used in clinical practice and therefore issues regarding the regulation of models running on the platform were not considered.

## 4. Conclusion

ICP research has the potential to significantly benefit from techniques such as machine learning. However, complex models, for example those resulting from machine learning, rarely transfer into clinical practice. As demonstrated during the CHART-ADAPT project, hurdles to model implementation in a clinical environment can be significant but ultimately overcome. Barriers to model implementation largely fall under the categories of technical, analytical, and integrity. Access to sufficient technical infrastructure in clinical settings needs to be improved, however some of the barriers to access to relevant technology are starting to be removed and increasingly processing power is becoming available in-house or through formal partnerships and agreements which make the transfer of patient data easier to external providers. As suggested by Moberg et al., issues regarding data integration require a consensus within the community to mitigate these challenges and develop a path forward for data 'readiness' for machine learning and artificial intelligence approaches (Moberg et al., 2022). Device manufacturers and other stakeholders will need to work closely together to agree upon standards which will, alongside the development of new techniques for artifact detection and removal, accelerate model development and deployment. Clinical end users should be put at the heart of model development and close attention paid to ensure the clinical utility of these models. Vollmer et al. propose a series of questions which should be asked of machine learning studies to ensure their transferability to a clinical environment. These range from questions to ensure a suitable clinical hypothesis is being explored, the correct data is being examined, appropriate techniques are being applied and the implementation and clinical impact is considered (Vollmer et al., 2020). Finally, models are only likely to gain widespread adoption when they are understood and trusted by clinicians. Interpretability, bias reduction, and robustness have already been identified as relevant for increasing trust in a model, but trust is a complicated concept and in order to increase levels of trust and consequently routine use of complex ICP modelling, future research is required to understand clinicians' needs from machine learning and AI and its role within existing clinical workflows.

## Author contributions

LM: conceptualisation of the work.
LM: wrote the manuscript.
MS: contributed to the discussion of several points.
All authors critically revised the paper for intellectual content and approved the final version of the manuscript.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Laura Moss reports financial support was provided by Innovate UK. Martin Shaw reports financial support was provided by Innovate UK. Ian Piper reports financial support was provided by Innovate UK. Christopher Hawthorne reports financial support was provided by Innovate UK.

## Acknowledgements

The CHART-ADAPT project mentioned in the manuscript was co-funded by Innovate UK (project number: 102113) and authors LM, MS, IP and CW were supported at the time by this grant.

This preparation of this article did not directly receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

Adegboro, C.O., Choudhury, A., Asan, O., Kelly, M.M., 2022. Artificial intelligence to improve health outcomes in the NICU and PICU: a systematic review. Hosp. Pediatr. 12 (1), 93–110. https://doi.org/10.1542/hpeds.2021-006094.

Al-Mufti, F., Kim, M., Dodson, V., et al., 2019. Machine learning and artificial intelligence in neurocritical care: a specialty-wide disruptive transformation or a strategy for success. Curr. Neurol. Neurosci. Rep. 19, 89. https://doi.org/10.1007/s11910-019-0998-8.

Alkhachroum, A., Terilli, K., Megjhani, M., 2020. Harnessing big data in neurocritical care in the era of precision medicine. Curr. Treat. Options Neurol. 22, 15. https://doi.org/10.1007/s11940-020-00622-8.

Alkhachroum, A., Kromm, J., De Georgia, M.A., 2022. Big data and predictive analytics in neurocritical care. Curr. Neurol. Neurosci. Rep. 22, 19–32. https://doi.org/10.1007/s11910-022-01167-w.

Apache Software Foundation. Apache Spark. http://spark.apache.org/. (Accessed 12 June 2023).

Apache Software Foundation. Apache Hadoop. http://hadoop.apache.org/. (Accessed 12 June 2023).

Aries, M.J.H., Czosnyka, M., Budohoski, K.P., Steiner, L.A., Lavinio, A., Kolias, A.G., et al., 2012. Continuous determination of optimal cerebral perfusion pressure in traumatic brain injury. Crit. Care Med. 40 (8), 2456–2463. https://doi.org/10.1097/CCM.0b013e3182514eb6.

Aristidou, A., Jena, R., Topol, E.J., 2022. Bridging the chasm between AI and clinical implementation. Lancet 399 (10325), 620. https://doi.org/10.1016/S0140-6736(22)00235-5.

Arrieta, A.B., Diaz-Rodriguez, N., Del Set, J., et al., 2020. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Asan, O., Bayrak, A.E., Choudhury, A., 2020. Artificial intelligence and human trust in healthcare: focus on clinicians. J. Med. Internet Res. 22 (6), e15154 https://doi.org/10.2196/15154.

Balestreri, M., Czosnyka, M., Steiner, L.A., Schmidt, E., Smielewski, P., Matta, B., Pickard, J.D., 2004. Intracranial hypertension: what additional information can be

derived from ICP waveform after head injury? Acta Neurochir. 146 (2), 131–141. https://doi.org/10.1007/s00701-003-0187-y. Feb.
Bellini, V., Valente, M., Pelosi, P., et al., 2022. Big data and artificial intelligence in intensive care unit: from "bla, bla, bla" to the incredible five v's. Neurocritical Care 37 (Suppl. 2), 170–172. https://doi.org/10.1007/s12028-022-01472-9.
BrainIT. http://www.brainit.org.uk/brainit/public/. (Accessed 12 June 2023).
Cardoso, E.R., Rowan, J.O., Galbraith, S., 1983. Analysis of the cerebrospinal fluid pulse wave in intracranial pressure. J. Neurosurg. 59, 817–821. https://doi.org/10.3171/jns.1983.59.5.0817.
Carney, N., Totten, A.M., O'Reilly, C., Ullman, J.S., Hawryluk, G.W.J., et al., 2017. Guidelines for the management of severe traumatic brain injury. Fourth Edition Neurosurgery 80 (1), 6–15. https://doi.org/10.1227/NEU.0000000000001432.
Carra, G., Salluh, J.I.F., da Silva Ramos, F.J., Meyfroidt, G., 2020. Data-driven ICU management: using Big Data and algorithms to improve outcomes. J. Crit. Care 60, 300–304. https://doi.org/10.1016/j.jcrc.2020.09.002.
Čartolovni, A., Tomičić, A., Mosler, E.L., 2022. Ethical, legal, and social considerations of AI-based medical decision-support tools: a scoping review. Int. J. Med. Inform 161, 104738. https://doi.org/10.1016/j.ijmedinf.2022.104738.
CENTER-TBI -/https://www.center-tbi.eu - Accessed 12th June 2023.
Char, D.S., Shah, N.H., Magnus, D., 2018. Implementing machine learning in health care - addressing ethical challenges. N. Engl. J. Med. 15 (11), 981–983. https://doi.org/10.1056/NEJMp1714229, 378.
Chaudhry, F., Hunt, R.J., Hariharan, P., Anand, S.K., Sanjay, S., Kjoller, E.E., Bartlett, C. M., Johnson, K.W., Levy, P.D., Noushmehr, H., Lee, I.Y., 2020. Machine learning applications in the neuro ICU: a solution to big data mayhem? Front. Neurol. 11 https://doi.org/10.3389/fneur.2020.554633.
Cheatham, M., Pesquita, C., 2017. Semantic data integration. In: Zomaya, A., Sakr, S. (Eds.), Handbook of Big Data Technologies. Springer, Cham. https://doi.org/10.1007/978-3-319-49340-4_8.
Citerio, G., 2022. Big data and artificial intelligence for precision medicine in the Neuro-ICU: Bla, Bla, Bla. Neurocritical Care 37 (Suppl. 2), 163–165. https://doi.org/10.1007/s12028-021-01427-6.
Czosnyka, M., Piechnik, S., Richards, H.K., Kirkpatrick, P., Smielewski, P., Pickard, J.D., 1997. Contribution of mathematical modelling to the interpretation of bedside tests of cerebrovascular autoregulation. J. Neurol. Neurosurg. Psychiatry 63 (6), 721–731. https://doi.org/10.1136/jnnp.63.6.721.
Depreitere, B., Giza, F., Van den Berghe, G., Schuhmann, M.U., Maier, G., Piper, I., Meyfroidt, G., 2014. Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. J. Neurosurg. 120 (6), 1451–1457. https://doi.org/10.3171/2014.3.JNS131500.
Dhar, R., Meyfroidt, G., 2022. Navigating the ocean of big data in neurocritical care. Neurocritical Care 37 (Suppl. 2), 157–159. https://doi.org/10.1007/s12028-022-01558-4.
EU Regulation (EU), 2016. 2. In: 016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) off J Eur Communities, vol. 59, pp. 1–88.
Feng, M., Loy, L.Y., Zhang, F., Guan, C., 2011. Artifact removal for intracranial pressure monitoring signals: a robust solution with signal decomposition. Annu Int. Conf. IEEE Eng. Med. Biol. Soc. 2011, 797–801. https://doi.org/10.1109/IEMBS.2011.6090182.
Feng, M., Loy, L.Y., Sim, K., Phua, C., Zhang, F., Guan, C., 2012. Artifact correction with robust statistics for non-stationary intracranial pressure signal monitoring. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, pp. 557–560.
Gianfrancesco, M.A., Tamang, S., Yazdany, J., Schmajuk, G., 2018. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern. Med. 178 (11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763.
Greenplum: Greenplum Database. https://greenplum.org/. Accessed 12th June 2023.
HL7. Health level seven. http://www.hl7.org.uk. (Accessed 12 June 2023).
Hu, X., Xu, P., Scalzo, F., Vespa, P., Bergsneider, M., 2009. Morphological clustering and analysis of continuous intracranial pressure. IEEE Trans. Biomed. Eng. 56 (3), 696–705. https://doi.org/10.1109/TBME.2008.2008636.
Kinsella, J., Hawthorne, C., Shaw, M., Piper, I., Philips Healthcare, Aridhia, Moss, L., 2017. Public perception of the collection and use of critical care patient data beyond treatment: a pilot study. 2016. Crit. Care Med. 44 (12), 470. https://doi.org/10.1097/01.ccm.0000510251.02647.1f.
Lee, S.B., et al., 2019. Artifact removal from neurophysiological signals: impact on intracranial and arterial pressure monitoring in traumatic brain injury. J. Neurosurg. 132 (6), 1952–1960. https://doi.org/10.3171/2019.2.JNS182260.
Lee, H.J., Kim, H., Kim, Y.T., Won, K., Czosnyka, M., Kim, D.J., 2021. Prediction of life-threatening intracranial hypertension during the acute phase of traumatic brain injury using machine learning. IEEE J. Biomed. Health Inform. 25 (10), 3967–3976. https://doi.org/10.1109/JBHI.2021.3085881.
Liu, X., Cruz Rivera, S., Moher, D., et al., 2020. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat. Med. 26, 1364–1374. https://doi.org/10.1038/s41591-020-1034-x.
Lundberg, N., 1960. Continuous recording and control of ventricular fluid pressure in neurosurgical practice. Acta Psychiatr. Scand. Suppl. 36, 1–193.
Martinez-Tejada, I., Wilhjelm, J.E., Juhler, M., Andresen, M., 2021. Empirical mode decomposition-based method for artefact removal in raw intracranial pressure signals. Acta Neurochir. Suppl. 131, 201–205. https://doi.org/10.1007/978-3-030-59436-7_39.
Megjhani, M., Alkhachroum, A., Terilli, K., et al., 2019. An active learning framework for enhancing identification of non-artifactual intracranial pressure waveforms. Physiol. Meas. 40 (1), 015002 https://doi.org/10.1088/1361-6579/aaf979.
Mládek, A., Gerla, V., Skalický, P., Vlasák, A., Zazay, A., Lhotská, L., Beneš Snr, V., Beneš Jr, V., Bradáč, O., 2022. Prediction of shunt responsiveness in suspected patients with normal pressure hydrocephalus using the lumbar infusion test: a machine learning approach. Neurosurgery 90 (4), 407–418. https://doi.org/10.1227/NEU.0000000000001838.
Moberg, R., Moyer, E.J., Olson, D., et al., 2022. Harmonization of physiological data in neurocritical care: challenges and a path forward. Neurocritical Care 37 (Suppl. 2), 202–205. https://doi.org/10.1007/s12028-022-01524-0.
Moss, L., Shaw, M., Piper, I., Hawthorne, C., Kinsella, J., Aridhia, Philips Healthcare, 2016. Apache Spark for the analysis of high frequency neurointensive care unit data: Preliminary comparison of Scala vs. R. In: Proc. Of American Medical Informatics Association 2016 Annual Symposium (AMIA 2016), p. 1523.
Moss, L., Shaw, M., Piper, I., Kinsella, J., Hawthorne, C., 2021. CHART-ADAPT: enabling actionable analytics at the critical care unit bedside. In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), vol. 2021. Proceedings. Los Alamitos:IEEE Computer Society, pp. 301–306. https://doi.org/10.1109/CBMS52027.2021.00032.
Moss, L., Corsar, D., Shaw, M., Piper, I., Hawthorne, C., 2022. Demystifying the black box: the importance of interpretability of predictive models in neurocritical care. Neurocritical Care 37 (Suppl. 2), 185–191. https://doi.org/10.1007/s12028-022-01504-4.
Nicholson Price, W., 2018. Medical Malpractice and black-box medicine. In: Glenn Cohen, I., et al. (Eds.), Big Data, Health Law, and Bioethics. Cambridge University Press, Cambridge, pp. 295–306.
Orphanidou, C., 2019. A review of big data applications of physiological signal data. Biophys. Rev. 11 (1), 83–87. https://doi.org/10.1007/s12551-018-0495-3.
Raj, R., Luostarinen, T., Pursiainen, E., Posti, J.P., Takala, R.S.K., Bendel, S., Konttila, T., Korja, M., 2019. Machine learning-based dynamic mortality prediction after traumatic brain injury. Sci. Rep. 9 (1), 17672 https://doi.org/10.1038/s41598-019-53889-6.
Rajagopalan, S., Baker, W., Mahanna-Gabrielli, E., Kofke, A.W., Balu, R., 2022. Hierarchical cluster analysis identifies distinct physiological states after acute brain injury. Neurocritical Care 36 (2), 630–639. https://doi.org/10.1007/s12028-021-01362-6.
Ribeiro, M.T., Sameer, S., Guestrin, C., 2016. Why should I trust you?: explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). Proceedings. Association for Computing Machinery, New York, NY, USA, pp. 1135–1144. https://doi.org/10.48550/arXiv.1602.04938.
Richards, M., 2022. Software Architecture Patterns, second ed. O'Reilly Media Inc. ISBN: 9781098134273.
https://icmplus.neurosurg.cam.ac.uk. (Accessed 12 June 2023).
Rivera, S.C., Liu, X., Chan, A., Denniston, A.K., Calvert, M.J., 2020. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. BMJ 370, m3210. https://doi.org/10.1136/bmj.m3210.
Sanchez-Pinto, L.N., Luo, Y., Churpek, M.M., 2018. Big data and data science in critical care. Chest 154 (5), 1239–1248. https://doi.org/10.1016/j.chest.2018.04.037.
Scala. The Scala Programming language. https://scala-lang.org/. (Accessed 12 June 2023).
Schweingruber, N., Mader, M.M.D., Wiehe, A., Röder, F., Göttsche, J., Kluge, S., Westpha, l M., Czorlich, P., Gerloff, C., 2022. A recurrent machine learning model predicts intracranial hypertension in neurointensive care patients. Brain 145 (8), 2910–2919. https://doi.org/10.1093/brain/awab453.
Shillan, D., Sterne, J.A.C., Champneys, A., Gibbison, B., 2019. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. Crit. Care 23, 284. https://doi.org/10.1186/s13054-019-2564-9.
Takeuchi, H., Kumabe, A., Zhang, P., Kogure, Y., Akutagawa, M., Kinouchi, Y., Zhang, Q., 2009. The development of remote monitoring system using international standard. In: Dössel, O., Schlegel, W.C. (Eds.), World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany. IFMBE Proceedings, 25/5. Springer, Berlin, Heidelberg, pp. 330–333. https://doi.org/10.1007/978-3-642-03904-1_92.
Teplan, M., Bajla, I., Rosipal, R., Rusnak, M., 2017. Feature clustering of intracranial pressure time series for alarm function estimation in traumatic brain injury. Physiol. Meas. 38 (11), 2015–2043. https://doi.org/10.1088/1361-6579/aa8a51.
TRACK-TBI. https://tracktbi.ucsf.edu/. (Accessed 12 June 2023).
van de Sande, D., van Genderen, M.E., Huiskens, J., et al., 2021. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med. 47, 750–760. https://doi.org/10.1007/s00134-021-06446-7.
Vokinger, K.N., Feuerriegel, S., Kesselheim, A.S., 2021. Mitigating bias in machine learning for medicine. Commun. Med. 1, 25. https://doi.org/10.1038/s43856-021-00028-w.
Vollmer, S., Mateen, B.A., Bohner, G., KirÃ¡ly, F.J., Ghani, R., Jonsson, P., 2020. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 368, l6927. https://doi.org/10.1136/bmj.l6927.
Ye, G., Balasubramanian, V., Li, J.K., Kaya, M., 2022. Machine learning-based continuous intracranial pressure prediction for traumatic injury patients. IEEE J. Transl. Eng. Health Med. 2 (10), 4901008 https://doi.org/10.1109/JTEHM.2022.3179874.
Zhou, Q., Chen, Z.-H., Cao, Y.-H., Peng, S., 2021. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence

prediction tools: a systematic review. NPJ. Digit. Med. 4, 1–12. https://doi.org/10.1038/s41746-021-00524-2.