S. Kagerbauer[1] · M. Blobner[1] · B. Ulm[1] · B. Jungwirth[1,2]

Tomorrow is already here [1]

How machine learning is influencing anaesthesiology and intensive care medicine

▶ **Citation:** Kagerbauer S, Blobner M, Ulm B, Jungwirth B: Tomorrow is already here [1] – How machine learning is influencing anaesthesiology and intensive care medicine. Anästh Intensivmed 2020;61:85–94. DOI: 10.19224/ai2020.85

1 Klinik für Anästhesiologie und Intensivmedizin, Technische Universität München, (Direktor: Prof. Dr. G. Schneider)
2 Klinik für Anästhesiologie, Universitätsklinikum Ulm (Ärztliche Direktorin: Prof. Dr. B. Jungwirth)

Conflict of interest
The authors have no financial or other competing interest to disclose.

Keywords
Artificial Intelligence – Machine Learning – Big Data – Risk Prediction

Summary

Artificial intelligence has become an everyday part of modern medicine. Increasing storage capacity and new ways of processing data are leading to ever increasing quantities of data being collected and analysed, especially in the areas of anaesthesia and intensive care medicine, both of which commonly use electronic patient data management systems. Being able to use machine learning on this data requires that it is not simply stored but readily found, accessible, interoperable and reusable in accordance with the FAIR-principles. Analysis utilises a variety of supervised and unsupervised learning methods leading, amongst other things, to hypotheses for prospective randomised trials, analysis of rare complications, risk stratification and development of decision support tools. The aim of this article is to provide an overview of methods and application of machine learning in anaesthesiology and intensive care medicine. In addition, potential pitfalls associated with the technology and possible solutions are discussed. In anaesthesiology and intensive care medicine – as in other areas of medicine - machine learning can help provide individualised care with the aim of avoiding complications and increasing the quality of care provided.

Introduction

“I myself have an app on my mobile which, fed with the answers to 20 or 30 questions, nails the diagnosis more accurately than many doctors do, because it can access so many more studies and so much more information than any one doctor could by themselves.” This comment made by Federal Minister for health Jens Spahn in relation to a health platform based on artificial intelligence (AI) shows that AI has now made its way into healthcare policy and society. Many of the buzzwords used – such as “artificial intelligence”, “machine learning” or “big data” – are often not clearly defined and stir up anxiety not only in relation to data protection and privacy but also centred around the question of whether humans might become dispensable. Could AI systems replace anaesthesiologists? After all, use of electronic patient data management systems in anaesthesiology generates enormous amounts of data which are predestined for use with AI technology. Using these new methods, data can be recorded and analysed in a structured fashion and put to systematic use. AI technology promises to avoid complications, streamline processes and optimise use of resources. As such, responsible use of AI in anaesthesiology and perioperative medicine has the potential to increase the quality of patient care. The aim of this article is to introduce the clinician to this new technology and to subsequently provide greater detail on special use cases in anaesthesiology and intensive care medicine.

Introduction to Artificial Intelligence in Anaesthesiology and Intensive Care Medicine

Artificial intelligence (AI) is used to mimic decision making structures in the human thought process. To this end, software with self-induced behaviour is developed based on machine learning processes.

By using AI, the intention is to provide personalised treatment to the individual based on large quantities of data. The data sets have to fulfil a number of prerequisites before they can be analysed using machine learning methods. It is for this reason that the following sections will detail the data structure before elucidating on machine learning processes.

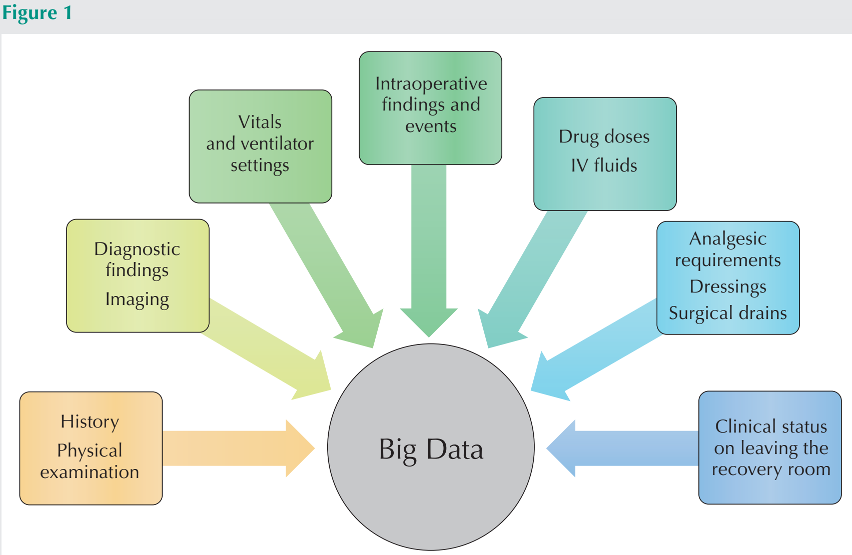
Data structure

Big Data

The term "big data", which is not strictly defined, describes large and sometimes convoluted quantities of data which cannot be managed within traditional database formats and equally cannot be analysed using conventional statistical methods [2]; in addition, the term encompasses the methods used for processing and analysing this data. "Big data" is characterised by the so-called 5-V-concept: "volume" describes the quantity of data, "velocity" the speed with which it is acquired, "variety" the diversity of data sources, "veracity" the precision and accuracy, and "value" the value of the data in an economical sense [3,4].

With regard to medicine this means that most data which was previously only recorded on paper is now available in a digital form. Recording vitals and labs, clinical observations and diagnostic findings leads to large quantities of data ("volume") being produced – sometimes, e.g. when vitals are recorded continuously, at high speed ("velocity"). The data exists in a variety of formats including text, audio or video files ("variety") and at different resolutions, levels of precision and accuracy ("veracity"). The benefit ("value") which can be reaped from the data, not only from an economical point of view but also with respect to the quality of care provided to the patient, is largely dependent on the quality of the data, its storage and processing, and the available expertise (Fig. 1).

Figure 1



From preoperative evaluation for anaesthesia through the intraoperative phase up to discharge from the recovery room, large quantities of data are produced in anaesthesiology. Their quantity and diversity make them "big data".

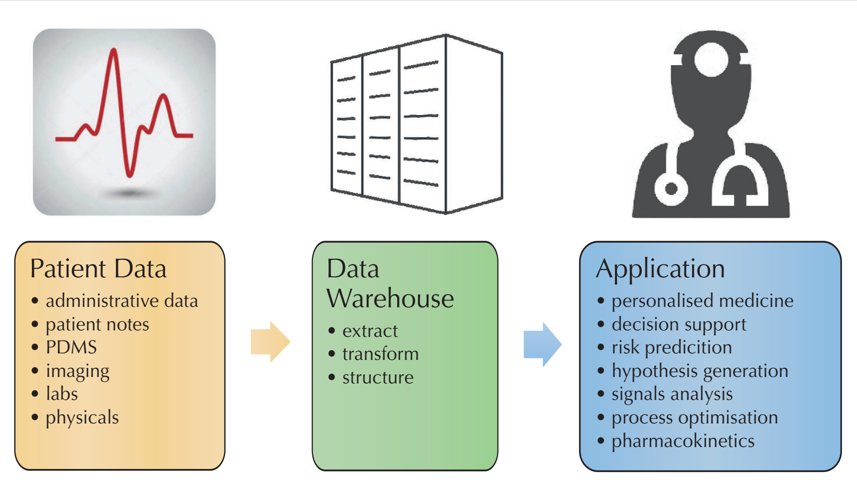
Smart Data

The term "smart data" focuses on the fifth V – "value" and comprises the integration of expert knowledge into the database, making it possible to attain usable information from the raw data [5]. The basis for this process is the fourth V – "veracity", which is ensured through high precision amongst other things. In a clinical context, precision is the equivalent of a positive predictive value, i.e. the probability that a patient with a positive test result will actually be suffering from the condition tested for. The second component, accuracy, shows how close the predicted and the actual value are to one another, so is a measure of correct prediction made by an algorithm. High precision and accuracy increase the veracity of the data. Only correct data is useful if clinical knowledge is to be implemented. Depending on what type of use is intended, smart data can be divided into four types: descriptive, predictive, prescriptive (determining) and cognitive (recognising). Consider the following example: the number of patients in a cohort suffering PONV can be described; the predictive value lies in the ability to preoperatively foresee the risk of PONV in a patient. Strategies for avoiding PONV can then be prescribed. The new insight can be used cognitively to explain the causes, effects and prevention of PONV to physicians and patients alike [3].

Patient Data Management Systems and Medical Databases

Automatic recording of anaesthesiologic data was implemented by individual centres towards the end of the 1970s and has become routine in many operating theatres today. In intensive care medicine PDMS (patient data management systems) have not yet been established comprehensively. A current survey showed that in total only 40% of all intensive care units used a digital documentation system, although the proportion was higher at university hospitals, reaching 58%. It is not only essential that these systems should digitally record, archive and analyse vitals, observations, medication and

Figure 2



Structured and unstructured data have to be consolidated and reorganised so as to be able to make them available to the respective applications. This is done using special, optimised databases – so-called "data warehouses".

interventions but in addition that the PDMS should also communicate with the other components of the hospital information system (HIS) and in turn be able to access results from other information systems, such as labs or radiology reports [6]. These data acquired from various sources are then collated, stored, filtered, reorganised, structured and compressed in a data warehouse, a special, optimised database, before being made available to the respective applications [7] (Fig. 2).

Special databases are required for academic use such as developing algorithms or improving on or correcting existing systems. The MIMIC-database, an example of a successful project which has been in service for some time, is an intensive care medicine database which is based in Boston and provides free access for academic use; innumerable publications have emanated from it. Amongst other things, MIMIC contains vitals, labs, diagnoses and procedures from intensive care patients over a period of more than 10 years [8]. More recently, a number of consortia aiming to bring together data acquired from patient care and academic work have been founded in Germany in association with the medical informatics initiative (www.medizininformatik-initiative.de). DIFUTURE (Data Integration for Future Medicine), HiGHmed (Heidelberg – Göttingen – Hannover Medical Informatics), MIRACUM (Medical Informatics for Research and Care in University Medicine) and SMITH (Smart Medical Information Technology for Healthcare) are examples, and more are currently planned. Use of universal nomenclature and a universal data format is a decisive factor in effective shared use of data originating from multiple centres. SNOMED CT (systematized nomenclature of human and veterinary medicine clinical terms) is one such universal nomenclature for symptoms, findings, diagnoses and procedures [9]. The OMOP (observational medical outcomes partnership) common data model is often used to convert data from multiple databases to one common format [10]. It is this harmonisation which makes data exchange across sites and systematic processing of data from multiple centres possible. Consolidating data from multiple centres provides detailed information even in regard to less common diseases and complications, which in turn can significantly increase the quality of patient care.

Artificial Intelligence in Data Analytics

Artificial Intelligence

In its original definition, artificial intelligence comprises mimicking human behaviour using mathematical methods and informatics. AI systems imitate human decision-making structures [11]. They must be able to learn and contend with a certain degree of uncertainty. Machine learning is a subsection of AI, despite the two terms "artificial intelligence" and "machine learning" often being used erroneously as synonyms for one another.

Machine Learning

Regardless of whether for advertising, online image searches, filtering spam e-mails or unlocking smartphones using facial recognition, machine learning has become well-established in everyday life. It makes use of complex statistical processes which are utilised to recognise patterns, compile prognoses and generate hypotheses from data sets, using automated algorithms. This way, in contrast to conventional statistical methods, the artificial system creates models which have not been pre-programmed and so is able to provide better and better solutions, optimising itself continuously [4] – a process which is termed "learning".

Supervised Learning

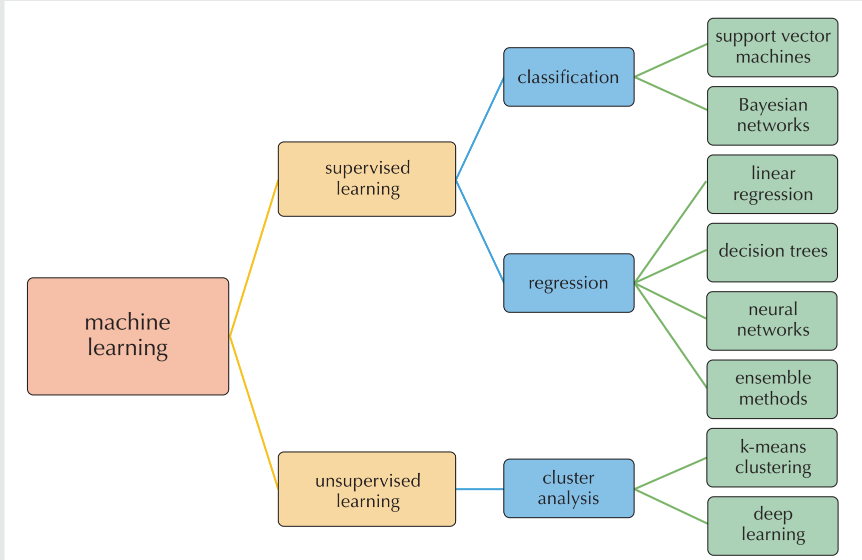
As a rule, supervised machine learning processes are used, especially in the fields of risk calculation and decision support. This entails training the algorithms used with a training data set with known results – i.e. using known input und known output. The training data set used for this purpose – the so-called classifier – should be as large as possible; the intention is for rules governing the relationship between individual data elements to be recognised. A further data set is then used for validation, whilst a third set – the test data set – confirms the predictive value of the algorithm which was generated [11,12]. Typically, training and validation are performed with data taken from one set, whilst the test data set is an independent, external set. This "oversight" is only undertaken during the training phase, whilst the system is not actively supervised during live operation. Because of this, it is important to avoid so-called overfitting. An increased number of errors encountered using test data sets with a decrease seen when using the training data sets can be a sign of overfitting – the model produces accurate prognoses in the context of training data, but incorrectly assigns unknown data. This is caused by an algorithm too tightly tied to the training data set, meaning that divergence between data sets leads to errors [12]. If on the other hand too few training data are provided, underfitting can occur. It is important to avoid both by using sufficient tests. Methods used in supervised learning include decision trees, neural networks, Bayesian networks and support vector machines.

Unsupervised Learning

In contrast with supervised learning, unsupervised learning does not use marked data, and the output is not known [12]. As such, it is not possible to train the algorithm. Instead, unsupervised processes attempt to recognise a structure in the data. As a rule, this is done using clustering methods, assigning individual observations to certain subsets (clusters) which feature internal similarities differing from others. Human intervention is not required until any structures identified require interpretation. An example taken from a clinical setting is the classification of patients with heart failure with preserved ejection fraction (HF-pEF), a very heterogenic patient cohort. The individual phenotypes were divided into three groups using clustering; those three differed from one another in clinical characteristics, haemodynamics and outcome. This group model can be used for risk stratification [13].

An overview of machine learning and commonly used methods is provided by Table 1 and Figure 3.

Figure 3



Machine learning can be divided into supervised and unsupervised learning. Classification and regression are domains of supervised learning, whilst unsupervised learning primarily includes cluster analysis.

Table 1
Commonly used machine learning methods.

Method	Description
Decision tree	Used to automatically classify objects. Made up of root nodes, inner nodes and leaves. Every node represents a logical rule, every leaf an answer.
Random forest	Classification method made up of numerous different decision trees.
Reinforcement learning	Determines how a computer program capable of self-induced behaviour must act to maximise rewards.
Bayesian networks	Develop graphic models of random variables and their conditional dependencies.
Support vector machines	Are used for classification and regression. Most commonly, Support Vector Machines are used to solve classification problems, segregating classes in the data by so-called hyperplanes (lines that separate the dataset).
Artificial neural networks	Are made up of interconnected artificial neurons capable of processing information.
Deep learning	Neural networks with numerous hidden (deep) layers which provide the network with a certain "depth".
Clustering	Aggregation of unsupervised learning methods. The system is unaware what it is aiming to recognise and splits observations into various categories (clusters).
Ensemble methods	Meta-algorithms which combine various machine learning methods.

Applications for Artificial Intelligence in Anaesthesiology and Intensive Care Medicine

AI opens up a range of applications within the realms of personalised medicine, decision support and risk prediction. It can also be put to use in signals analysis and for developing pharmacokinetic models. In epidemiologic research, AI shows great potential in investigating rare diseases and complications, and in hypothesis generation for prospective randomised clinical trials.

Interpretation of Clinical Trials, Hypothesis Generation, and Detection of Rare Complications

Classical retrospective trials involving manual sighting of a limited number of patient notes are labour-intensive and subject to several limitations, such as incomplete detection of confounders, contradictions within patients' notes, and differing assessments of clinical events [14]. Nowadays, retrospective analyses are based upon large quantities of data, typically from thousands of patients. Often, automated analytic processes belonging to the field of machine learning are used. This has led to the almost complete displacement of retrospective cohort and case-control studies [15]. In addition, analysis of large quantities of data can be used to generate hypotheses which can then be further investigated within prospective randomised clinical trials. This way, precise inclusion criteria can be defined [14]. Bailly et al., for example, investigated the efficacy of treatment with an antimycotic on sepsis patients using an intensive care database [16]. The hypothesis was then further reviewed within a selective randomised double-blind trial [17].

Machine learning methods make the analysis of results of prospective trials with large case numbers easier and are superior to conventional statistical methods, particularly when a multitude of subgroups or the influence of multiple factors is to be analysed. An investigation comprising 22,000 patients which showed that the use of neuromuscular blocking agents increases the rate of postoperative pulmonary complications [18] is an example.

Rare diseases or events can be analysed with the help of large data sets from as many hospitals as possible, something which would require an almost impossibly large number of cases using randomised clinical trials. In anaesthesiology especially, this concerns cohorts such as children or pregnant women, for whom there are sparse data from randomised controlled trials. An example might be the risk of neuraxial haematoma in obstetric anaesthesia in the presence of thrombocytopenia [19] or the identification of predictors of failure of laryngeal masks in children – something which differs significantly when compared with an adult population. 11,910 cases of paediatric anaesthesia were required to identify 102 children in whom a laryngeal mask was used and who required secondary intubation [20].

It's not just scientific questions which can be worded, suitable study designs developed, or large quantities of data analysed with the aid of machine learning. These tools can help make literature searches more efficient too – something which is necessitated in medical research by the ever-increasing flood of information. Data mining – using automated processes to extract useful information from quantities of data so large as to be convoluted – is also called for in the development of clinical guidelines. The current ESA recommendations concerning the evaluation of patients undergoing non-cardiac surgery were developed using data mining tools such as “PubReMiner” or “TerMine” [21].

Risk Prediction and Decision Support Systems

Knowing the precise individual risk with respect to treatment success is essential if tailor-made and hence personalised anaesthesia is to be provided to the patient. The aim is to reduce perioperative morbidity and mortality, something commonly accepted to be the anaesthesiologist's core competence [22].

This has led to the increasing use of risk calculators in everyday clinical life. These are complex algorithms on the basis of stochastic or machine learning methods [23]; their development is based on large databases. Today, they are used across numerous fields of medicine and have been incorporated into guidelines. In their current guidelines, the ESA, for example, recommend the use of the NSQIP (The National Surgery Quality Improvement Project) risk calculator for preoperative evaluation [21]. This tool was developed as a decision aid for patients and surgeons using clinical data from 393 hospitals encompassing more than 1.4 million patients. Based on known preoperative risk factors, regression models for postoperative outcome were created. These included cardiac, pulmonary or infectious complications, but also thrombosis or the risk of requiring revision surgery. The risk that an individual patient will suffer one or more of these complications can be calculated with the NSQIP web-based tool and can aid in decision making pro or contra surgery [24].

Examples in Anaesthesiology

Whilst the NSQIP still uses classical regression models, current literature in the field of risk stratification increasingly uses machine learning algorithms. These can be used, for example, to calculate the perioperative risk of delirium, the PONV risk or the risk of deterioration in the recovery room [25–27]. A team surrounding Olsen, for example, used a random forest classifier to recognise clinical deterioration in the recovery room at an early stage [27]. Davoudi's work to determine the risk of delirium compared various methods, amongst them the random forest model, support vector machines and neural networks [25]. The transition from risk calculation models to decision support is fluid. Initially, the data recorded and analysed by PDMS is used to calculate risks. These calculations become clinically relevant if, in an additional step, algorithms within the system provide the physician with treatment recommendations [28]. As such, when implementing their on-screen tool for determining risk of PONV, Kappen et al. found that simply displaying the risk did not lead to a decreased incidence of PONV. It wasn't until a directive approach using concrete treatment recommendations was implemented that the outcome was improved [26]. To be able to derive direct clinical consequences from risk calculations requires the use of real-time processes, e.g. for directing fluid therapy during major abdominal surgery [29] or for predicting intraoperative hypotension based on the arterial blood pressure waveform [30].

Examples in Intensive Care Medicine

Computer-based decision support systems, e.g. to aid in determining ideal antibiotic treatment for intensive care patients, have entered into everyday clinical practice in the past few years [31]. Algorithms for risk prediction are also increasingly being developed and deployed.

For use specifically in intensive care medicine and based solely on machine learning, Piracchio and colleagues developed an "super ICU learner" algorithm which predicts in-patient mortality. The prognoses made by this algorithm were more reliable than conventional scores [32].

A study protocol was introduced in paediatric intensive care, aiming to record the cardiorespiratory signals provided by 250 preterm infants and analyse these using machine learning algorithms. These data are to be used to predict the ideal time for extubation of ventilated premature infants [33]. A promising reinforcement learning model for guiding sepsis therapy was developed. Analysis of a retrospective data set had shown that the probability of survival was increased when treatment provided by clinicians corresponded with the suggestions made by the system [34].

These examples are exemplary for a whole range of decision support systems which have been tested in the context of trials. Routine integration of these algorithms into PDMS holds future potential.

Table 2

Applications for machine learning in anaesthesiology and intensive care medicine with examples from a hospital environment.

Application	Examples
Hypothesis generation	Efficacy of antimycotic treatment in patients suffering sepsis [16]
Detecting rare complications	Neuraxial haematoma in obstetric anaesthesia [19] Paediatric airway management [20]
Literature research	ESA guidelines on preoperative evaluation of patients prior to non-cardiac surgery [21]
Decision support	Predicting PONV [26] Guiding sepsis therapy [34]
Risk prediction	NSQIP risk calculator [24] Mortality in intensive care units (Super ICU Learner) [32]
Signals analysis	Depth of anaesthesia monitoring [36] Interpretation of somatosensory potentials [37]
OMICS	Biomarker for sepsis [41] Duration of recovery following general anaesthesia [42]
Process optimisation	Optimising theatre and recovery room capacity [43] Transfusion lists [44]
Pharmacokinetic models	Drug side effects [46]

Further Applications

The biggest steps forward have been made in the field of signals analysis, particularly with respect to ECG interpretation. Today, cardiac arrhythmias can be detected at home by patients using smart watches [35]. Depth of anaesthesia monitoring and intraoperative analysis of somatosensory evoked potentials can also be provided with the aid of machine learning algorithms [36, 37].

Machine learning algorithms can also be used in the development of anaesthesiologic closed-loop systems which automatically control hypnosis, analgesia and neuromuscular block. The combined use of the automated anaesthesia system "McSleepy" with the da Vinci robotic surgical system caused something of a public stir [38]. A system designed to sedate patients for endoscopic procedures (Sedasys) was quickly taken back off the market – ostensibly due to a lack of acceptance by patients, physicians and nursing staff [39]. Despite that, the development of closed-loop systems has not come to a standstill. Trials involving small patient cohorts have shown what is technologically feasible and delivered promising results [40]. Unfortunately, trials involving large patient cohorts, and especially patients with complex medical histories, are lacking – presumably due to the technical complexities and costs involved.

Omics data is used in intensive care medicine to find new biomarkers for sepsis which can be used to differentiate between infectious and non-infectious inflammation and predict treatment results and response to specific therapies [41].

A paper looking into anaesthesia was able to show that postoperative recovery times are influenced by the patient's genotype such that it is possible to identify patients with a prolonged recovery time following general anaesthesia prior to surgery [42].

Machine learning algorithms are also used to optimise internal and logistical processes in hospitals, e.g. matching theatre and recovery room capacities, optimum use of which reduces costs significantly and helps plan the staff rota [43]. Further applications include planning blood transfusion requirements and compilation of transfusion lists for specific surgical procedures [44].

AI and machine learning have also gained entry into pharmacological research. By using a deep learning model, the bispectral index (BIS) can be predicted when a target-controlled infusion (TCI) of propofol and remifentanil is used [45]. Models for use in predicting adverse effects of drugs have also been developed – these can be used to enhance patient safety [46].

A summary of applications can be found in Table 2.

Personalised Medicine

Medicine in the 21st century is undergoing a transition from being purely "reactive" to a predictive, preventative and personalised approach; the aim is to identify at-risk patients before the complication is encountered [47]. This becomes particularly relevant with the knowledge that every year across the globe 4.2 million people die within 30 days of surgery. That makes postoperative mortality the 3rd most common cause of death [48]. Algorithms to identify at-risk patients at an early stage have already been developed for the field of intensive care medicine. Masino et al., for example, developed a model which can predict sepsis in newborns even before they develop clinical signs [49]. For adults a model exists which can predict mortality in cardiogenic and septic shock [50]. Using tools such as these means that individual life-saving decisions can be taken at an early stage [50]. Establishing targeted preventative measures and developing individual, tailor-made treatments promises to notably advance the care particularly of patients with complex diseases [2]. Large multicentre trials showing better patient outcomes using machine learning methods when compared with conventional therapy are still lacking. Large quantities of data from PDMS and electronic patient notes, analysed with the help of machine learning algorithms, could, however, help reduce mortality, increase the quality of care and enhance patient satisfaction [51,52].

Problems and How to Approach Them

Risks

In their statement published 2017, Cabitza et al. summarise the major concerns voiced by critics of machine learning. Put bluntly, the risks of machine learning are characterised as follows: the use of computer-aided algorithms causes physicians to lose their ability to diagnose and treat unaided. For various reasons, the algorithms are error prone. These errors cannot be detected, because machine learning algorithms are “black boxes” and as such are inscrutable and not amenable to fault analysis [53].

That so-called “de-skilling” – which leads to physicians tending to accept erroneous decisions by the system despite the fact that their clinical expertise puts them in a position to make the right decision – is a problem that cannot be dismissed out of hand. Use of computer-assisted systems led to less experienced physicians making more correct decisions. However, when in doubt, experienced diagnosticians were more likely to latch onto an incorrect decision made by the system, whereas without it they would have had a higher diagnostic accuracy [53,54].

That machine learning algorithms are error prone is due to the fact that in contrast to humans they cannot recognise context in certain situations or apply it incorrectly. This is particularly likely to happen if the training data set deviates from real-life conditions. For example: an AI system was to recognise horses in pictures. Approximately one fifth of pictures in the training data set contained a reference pointing to the source of the picture. When the tag containing the reference was removed, the horse was no longer recognised as such [55]. “Intrinsic uncertainty” in medicine is another problem [53]. If data which has already been analysed and interpreted by a human is fed to an algorithm, this algorithm will be influenced by subjective factors, which may also be a source of error.

The lack of explicability means that results or recommendations produced by machine learning algorithms are not transparent, making them “black boxes” [56,57]. This way, fault analysis becomes almost impossible. However, anaesthesiologists in particular will be aware that development of modern anaesthesia would have been significantly hampered if all “black boxes” had been banished from operating theatres from the word go [56]. After all, anaesthetics were used long before there was any understanding of their mode of action. Obviously, ethical and qualitative standards of the 19th century cannot be applied today. It is for this reason that responsible use of machine learning and artificial intelligence is subject to prerequisites, in particular with regard to training of medical staff and the type and quality of data used.

Training

Electronic patient notes and telemedical consultations are part of everyday clinical practice. And whilst most clinicians will use information from large databases as part of their daily routine, only a few will be acquainted with the analytical methods used to interpret that data. A failure to grasp the context and uncritical use of available data can lead to misinterpretation and consequently to misdiagnosis and incorrect treatment. It is therefore important to train physicians and particularly clinical researchers to handle large data sets and electronic documentation and decision support tools. Equally, tools used in practice and research must be thoroughly validated. Knowledge and understanding of clinical context will remain important in the hospital of the future.

Data Diversity

Medical data are heterogenic, exist in both structured and unstructured forms and are derived from a multitude of sources [7]. Structured data from electronic patient notes, e.g. vitals recorded by a PDMS, can be archived in databases, structured and analysed relatively simply. Traditionally, however, patient notes contain a good deal of unstructured data, mainly handwritten medical and nursing notes and free text diagnostic findings. Illegible handwritten notes in particular impair the veracity of data and lead to misinterpretations. It has long been known that the easiest way to avoid prescription errors is to use digital rather than handwritten orders.

Unfortunately, structured input masks are often rejected by clinical staff used to free text inputs. However, to obtain structured data it is necessary to reduce free text input to a minimum and use predefined input fields. As such, to gain acceptance amongst medical and nursing staff, it is essential that user-friendly software and input masks are developed [2].

Until structured input masks have been universally introduced, information will still have to be extracted from unstructured free text. Conventional methods of text mining or statistical or probabilistic methods of extracting information from large quantities of free text fail especially when spelling mistakes are made or ambiguous terms are used; grammatical features, especially in non-English texts, can also be problematic. Consider negations which can, for example, be presented in numerous forms ("can be excluded", "free of symptoms", "no indication of") and recognition of which requires adaption of current tools or development of new ones [58]. Natural language processing as a domain of deep learning is increasingly making it possible to extract meaning and context from complex texts.

Data Acquisition

When special monitoring procedures are used, interfaces to PDMS will often be unavailable. Consider, for example, the use of separate neuro-monitoring or extended haemodynamic monitoring. When multiple hospitals are involved, it is important to identify the necessary interfaces prior to implementing a trial; the compatibility of the systems must be ensured, and universal nomenclature, a universal definition of clinically relevant events and a universal archiving format agreed upon. The technical and human resources required are sometimes underestimated; however, harmonisation of the acquired data is a decisive factor in the usability of that data [59]. Professional societies are increasingly making recommendations regarding data acquisition and data management in the context of multicentre trials [60]. As a rule, the FAIR principle should govern large data sets: data should be findable, accessible, interoperable and reusable [61].

Data Quality

When data sets are incomplete or contain invalid data, this increases the probability of type II errors when conventional statistical methods are employed. Big data, in contrast, is not as susceptible to incomplete or erroneous data [7]. Data can also be stored at different resolutions. Invasive blood pressure readings, for example, could be recorded every minute, every 5 seconds or even at a resolution of 100 Hz; it is necessary to find a compromise between loss of information and storage requirements.

Besides that, the quality of data is often influenced by artefacts. When paediatric anaesthesia was investigated, non-invasive blood pressure measurements showed artefacts in 5% of recordings, increasing to 7.3% for invasive measurements [62]. That makes it important to define aberrant values and to identify artefacts from those values. To achieve this, AI technologies such as recurrent neural networks (neural networks which possess recurrent connections within current or with past layers, creating feedback loops) and long short-term memory technologies (neural networks with a sort of memory of past events) can be used.

Data Protection and Data Safety

In adherence with current data protection laws and to avoid misuse of patient data, it is necessary to anonymise that data [7]. However, this is not always possible without a loss of data quality. Some data sets, e.g. genome data, cannot be anonymised in the traditional sense of the word. To ensure anonymity of data without making it unusable for diagnostic purposes, it must on the one hand be purged, i.e. have any unimportant data removed. On the other hand, noise, that is information not actually belonging to the data set, can be added to make identification more difficult (software-cluster.org, article dated 02.08.2017).

Ethical Questions

To discuss the whole range of ethical questions which arise with increasing use of AI in medicine would go beyond the scope of this article. The question of liability in case of injury to a patient, for example, has not been resolved. At the same time, however, the question arises of whether it is acceptable to withhold AI systems used for diagnostic and treatment purposes from a patient when these systems have proven to be more accurate than a physician. The German Association for the Digital Economy discusses these and similar questions in its discussion paper "Humans, Morals, Machines. Digital Ethics and Artificial Intelligence" ("Mensch, Moral, Maschine. Digitale Ethik und künstliche Intelligenz"; German; www.bvdw.org, accessed on 13.06.2019).

An additional point is the fact, already remarked upon above, that machine learning algorithms are often very complex and therefore inscrutable. However, methods are being developed which analyse input and output and can identify factors which have led to the respective conclusion or recommendation. The LIME model (local interpretable model-agnostic explanations) is an example. There is a trend towards developing comprehensible models [52,57].

Outlook

Medicine is changing. The availability of large data sets and technical options for storing and analysing that data opens up new opportunities in anaesthesiology and intensive care medicine. Epidemiological studies, analysis treatment success and risk stratification are simplified using big data methods. Subgroup analysis makes it possible to plan and carry out targeted prospective trials involving specific patient cohorts. Structured recording, archiving and analysis of data together with targeted consolidation of data sets can capture more data on patients suffering rare diseases or rare perioperative complications. The aim is to provide an individual risk assessment and optimised treatment to each individual patient, to increase efficiency of perioperative processes and reduce the rate of complications.

Literature

1. Jungk R: Die Zukunft hat schon begonnen. Entmenschlichung – Gefahr unserer Zivilisation. Bern/Stuttgart: Goldmann 1952. Amerikanische Ausgabe: Tomorrow is already here. New York: Simon and Schuster 1954
2. Raghupathi W, Raghupathi V: Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3
3. Hofer IS, Halperin E, Cannesson M: Opening the Black Box: Understanding the Science Behind Big Data and Predictive Analytics. Anesth Analg 2018;127:1139–1143
4. Bailly S, Meyfroidt G, Timsit JF: What's new in ICU in 2050: big data and machine learning. Intensive Care Med 2018;44:1524–1527
5. Funkner AA, Yakovlev AN, Kovalchuk SV: Data-driven modeling of clinical pathways using electronic health records. Pr 2017;121:835–842
6. Baumgartel M, Riessen R, John S: Digitization in Intensive Care

Medicine. Dtsch Med Wochenschr 2019;144:436–441
7. Austin C, Kusumoto F: The application of Big Data in medicine: current implications and future directions. J Interv Card Electrophysiol 2016;47:51–59
8. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al: MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035
9. Chu L, Kannan V, Basit MA, Schaeflein DJ, Ortuzar AR, Glorioso JF, et al: SNOMED CT Concept Hierarchies for Computable Clinical Phenotypes From Electronic Health Record Data: Comparison of Intensional Versus Extensional Value Sets. JMIR Med Inform 2019;7:e11487
10. Maier C, Lang L, Storf H, Vormstein P, Bieber R, Bernarding J, et al: Towards Implementation of OMOP in a German University Hospital Consortium. Appl Clin Inform 2018;9:54–61
11. Dilsizian SE, Siegel EL: Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 2014;16:441
12. Lee S, Mohr NM, Street WN, Nadkarni P: Machine Learning in Relation to Emergency Medicine Clinical and Operational Scenarios: An Overview. West J Emerg Med 2019;20:219–227
13. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al: Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation 2015;131:269–279
14. Sessler DI: Big Data-and its contributions to peri-operative medicine. Anaesthesia 2014;69:100–105
15. Bhatt DL, Mehta C: Adaptive Designs for Clinical Trials. N Engl J Med 2016;375: 65–74
16. Bailly S, Bouadma L, Azoulay E, Orgeas MG, Adrie C, Souweine B, et al: Failure of empirical systemic antifungal therapy in mechanically ventilated critically ill patients. Am J Respir Crit Care Med 2015;191:1139–1146
17. Timsit JF, Azoulay E, Schwebel C, Charles PE, Cornet M, Souweine B, et al: Empirical Micafungin Treatment and Survival Without Invasive Fungal Infection in Adults With ICU-Acquired Sepsis, Candida Colonization, and Multiple Organ Failure: The EMPIRICUS Randomized Clinical Trial. JAMA 2016;316:1555–1564
18. Kirmeier E, Eriksson LI, Lewald H, Jonsson Fagerlund M, Hoeft A, Hollmann M, et al: Post-anaesthesia pulmonary complications after use of muscle relaxants (POPULAR): a multicentre, prospective observational study. Lancet Respir Med 2019;7:129–140
19. Klumpner TT, Bauer ME, Kheterpal S: Current applications of big data in obstetric anesthesiology. Curr Opin Anaesthesiol 2017;30:300–305
20. Mathis MR, Haydar B, Taylor EL, Morris M, Malviya SV, Christensen RE, et al: Failure of the Laryngeal Mask Airway Unique and Classic in the pediatric surgical patient: a study of clinical predictors and outcomes. Anesthesiology 2013;119:1284–1295
21. De Hert S, Staender S, Fritsch G, Hinkelbein J, Afshari A, Bettelli G, et al: Pre-operative evaluation of adults undergoing elective noncardiac surgery: Updated guideline from the European Society of Anaesthesiology. Eur J Anaesthesiol 2018;35:407–465
22. Boehm O: Morbidität und Sterblichkeit in der Anästhesiologie. Patientensicherheit als Kernkompetenz des Anästhesisten. Anästh Intensivmed 2019;60:488–500
23. Mansmann U, Rieger A, Strahwald B, Crispin A: Risk calculators-methods, development, implementation, and validation. Int J Colorectal Dis 2016;31:1111–1116
24. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al: Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013;217:833–842
25. Davoudi A, Ebadi A, Rashidi P, Ozrazgat-Baslanti T, Bihorac A, Bursian AC: Delirium Prediction using Machine Learning Models on Preoperative Electronic Health Records Data. Proc IEEE Int Symp Bioinformatics Bioeng 2017;2017:568–573
26. Kappen TH, Vergouwe Y, van Wolfswinkel L, Kalkman CJ, Moons KG, van Klei WA: Impact of adding therapeutic recommendations to risk assessments from a prediction model for postoperative nausea and vomiting. Br J Anaesth 2015;114:252–260
27. Olsen RM, Aasvang EK, Meyhoff CS, Dissing Sorensen HB: Towards an automated multimodal clinical decision support system at the post anesthesia care unit. Comput Biol Med 2018; 101:15–21
28. Simpao AF, Tan JM, Lingappan AM, Galvez JA, Morgan SE, Krall MA: A systematic review of near real-time and point-of-care clinical decision support in anesthesia information management systems. J Clin Monit Comput 2017;31:885–894
29. Joosten A, Alexander B, Duranteau J: Clinical decision support system clears the way for perioperative goal directed therapy protocol adherence improvement. Minerva Anestesiol 2019;85:691–692
30. 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;129:663–674
31. Tafelski S, Sacher A, Nachtigall I, Powollik R, Tamarkin A, Angermair S et al: Unterstützung der Antibiotikatherapie auf Intensivstationen in Deutschland. Analyse der Anwenderfreundlichkeit und Zufriedenheitsanalyse mit einem computerassistierten Stewardship-Programm. Anästh Intensivmed 2016;57:174–181
32. 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;3:42–52
33. Shalish W, Kanbar LJ, Rao S, Robles-Rubio CA, Kovacs L, Chawla S, et al: Prediction of Extubation readiness in extremely preterm infants by the automated analysis of cardiorespiratory behavior: study protocol. BMC Pediatr 2017;17:167
34. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA: The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018;24:1716–1720
35. Dorr M, Nohturfft V, Brasier N, Bosshard E, Djurdjevic A, Gross S, et al: The WATCH AF Trial: SmartWATCHes for Detection of Atrial Fibrillation. JACC Clin Electrophysiol 2019;5:199–208
36. Sadrawi M, Fan SZ, Abbod MF, Jen KK, Shieh JS: Computational Depth of Anesthesia via Multiple Vital Signs Based on Artificial Neural Networks. Biomed Res Int 2015;2015:536863
37. Fan B, Li HX, Hu Y: An Intelligent Decision System for Intraoperative Somatosensory Evoked Potential Monitoring. IEEE Trans Neural Syst Rehabil Eng 2016;24:300–307
38. Wehbe M, Arbeid E, Cyr S, Mathieu PA, Taddei R, Morse J, et al: A technical description of a novel pharmacological anesthesia robot. J Clin Monit Comput 2014;28:27–34

39. Goudra B, Singh PM: Failure of Sedasys: Destiny or Poor Design? Anesth Analg 2017;124:686–688
40. Zaouter C, Hemmerling TM, Mion S, Leroux L, Remy A, Ouattara A: Feasibility of Automated Propofol Sedation for Transcatheter Aortic Valve Implantation: A Pilot Study. Anesth Analg 2017;125: 1505–1512
41. Prucha M, Zazula R, Russwurm S: Sepsis Diagnostics in the Era of „Omics" Technologies. Prague Med Rep 2018; 119:9–29
42. Xie S, Ma W, Shen M, Guo Q, Wang E, Huang C, et al: Clinical and pharmacogenetics associated with recovery time from general anesthesia. Pharmacogenomics 2018;19:1111–1123
43. Fairley M, Scheinker D, Brandeau ML: Improving the efficiency of the operating room environment with an optimization and machine learning model. Health Care Manag Sci 2019;22:756–767
44. Pendry K: The use of big data in transfusion medicine. Transfus Med 2015; 25:129–137
45. Lee HC, Ryu HG, Chung EJ, Jung CW: Prediction of Bispectral Index during Target-controlled Infusion of Propofol and Remifentanil: A Deep Learning Approach. Anesthesiology 2018;128: 492–501
46. Dey S, Luo H, Fokoue A, Hu J, Zhang P: Predicting adverse drug reactions through interpretable deep learning framework. BMC Bioinformatics 2018;19:476
47. 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
48. Nepogodiev D, Martin J, Biccard B, Makupe A, Bhangu A, National Institute for Health Research Global Health Research Unit on Global Surgery: Global burden of postoperative death. Lancet 2019;393:401
49. Masino AJ, Harris MC, Forsyth D, Ostapenko S, Srinivasan L, Bonafide CP, et al: Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. PLoS One 2019;14:e0212665
50. Aushev A, Ripoll VR, Vellido A, Aletti F, Pinto BB, Herpain A et al: Feature selection for the accurate prediction of septic and cardiogenic shock ICU mortality in the acute phase. PLoS One 2018;13:e0199089
51. Contreras I, Vehi J: Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018;20:e10775
52. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018;2:749–760
53. Cabitza F, Rasoini R, Gensini GF: Unintended Consequences of Machine Learning in Medicine. JAMA 2017; 318:517–518
54. Povyakalo AA, Alberdi E, Strigini L, Ayton P: How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography. Med Decis Making 2013;33:98–107
55. Lapuschkin S, Wäldchen S, Binder A, Montavon G, Samek W, Müller KR: Unmasking Clever Hans predictors and assessing what machines really learn. Nature Communications 2019;10:1096. DOI: 10.1038/s41467-019-08987-4
56. Cannesson M, Shafer SL: All Boxes Are Black. Anesth Analg 2016;122:309–317
57. Tapan S, Wang D: A Further Study on Mining DNA Motifs Using Fuzzy Self-Organizing Maps. IEEE Trans Neural Netw Learn Syst 2016;27:113–124
58. Koza W, Filippo D, Cotik V, Stricker V, Munoz M, Godoy N, et al: Automatic Detection of Negated Findings in Radiological Reports for Spanish Language: Methodology Based on Lexicon-Grammatical Information Processing. J Digit Imaging 2019;32:19–29
59. Rodriguez A, Smielewski P, Rosenthal E, Moberg D: Medical Device Connectivity Challenges Outline the Technical Requirements and Standards For Promoting Big Data Research and Personalized Medicine in Neurocritical Care. Mil Med 2018;183:99–104
60. Roger VL, Boerwinkle E, Crapo JD, Douglas PS, Epstein JA, Granger CB, et al: Strategic transformation of population studies: recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. Am J Epidemiol 2015;181:363–68
61. Hawkins BE, Huie JR, Almeida C, Chen J, Ferguson AR: Data Dissemination: Shortening the Long Tail of Traumatic Brain Injury Dark Data. J Neurotrauma 2019 (Online First). DOI: 10.1089/neu.2018.6192
62. Hoorweg AJ, Pasma W, van Wolfswinkel L, de Graaff JC: Incidence of Artifacts and Deviating Values in Research Data Obtained from an Anesthesia Information Management System in Children. Anesthesiology 2018;128:293–304.

Korrespondenzadresse

Prof. Dr. med.
Bettina Jungwirth

Klinik für Anästhesiologie
Universitätsklinikum Ulm
Albert-Einstein-Allee 23
89081 Ulm, Deutschland

E-Mail: bettina.jungwirth@uni-ulm.de

ORCID-ID: 0000-0001-9749-7460