Chapter 10
Integration of AI for Clinical Decision Support

Shyam Visweswaran, Andrew J. King, and Gregory F. Cooper

After reading this chapter, you should know the answers to these questions:

- What are the key challenges faced by clinicians that motivate integrating AI into clinical decision support?
- What are the main types of AI that have been developed for clinical decision support? How does data-derived AI clinical decision support differ from knowledge-based AI clinical decision support?
- What are typical degrees of automation and integration of AI in clinical decision support?
- Describe the types of clinical tasks that can be supported by AI clinical decision support?
- What are the pitfalls of data-derived clinical decision support?

Clinical decision support (CDS) aims to improve health and healthcare by providing clinicians, healthcare workers, and patients with situation-specific knowledge that aids critical clinical activities such as risk assessment, diagnosis, prognosis, and selection of therapy [1]. CDS systems assist clinicians in making decisions about patient care in various ways, such as by providing interpretations of patient data and clinical images, event monitoring and alerts, and recommendations. Some CDS systems guide patients and caregivers who integrate the clinical guidance from the CDS with their personal preferences to make informed decisions.

S. Visweswaran (✉) · G. F. Cooper
Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: shv3@pitt.edu; gfc@pitt.edu

A. J. King
Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: andrew.king@pitt.edu

T. A. Cohen et al. (eds.), *Intelligent Systems in Medicine and Health*, Cognitive Informatics in Biomedicine and Healthcare,
https://doi.org/10.1007/978-3-031-09108-7_10

Artificial intelligence (AI) enables computer systems to perform tasks that normally require human intelligence (see Chap. 1 for detailed definitions of AI). Because clinical decision-making predominates in medical practice, most applications of AI in clinical care are intended to enhance the quality of clinical decisions. Since the beginnings of AI in the 1950s, AI in medicine has been used increasingly for CDS, although the type of AI that drives CDS systems has evolved over the decades (see Chap. 2 for a historical account of AI in medicine). In the current era, modern AI that leverages large amounts of healthcare data to construct computational models is increasingly being used in such systems.

This chapter provides an overview of the rapidly developing field of **artificial intelligence-based clinical decision support** (AI-CDS) and the associated promising research efforts; it focuses on CDS that is targeted to clinicians, provides the motivation for integrating AI into CDS, describes the types of AI that are being developed for CDS systems, and explores a range of clinical tasks that AI-CDS can support. While the potential benefits of AI-CDS are enormous, significant challenges remain that must be overcome to ensure high-quality healthcare. This chapter describes some of the challenges (especially those stemming from **big data**), summarizes related regulatory developments, and closes with several predictions regarding future directions for AI-CDS.

Challenges Faced by Clinicians

Excellent clinical decision-making requires (1) up-to-date, pertinent medical knowledge, (2) access to accurate and complete patient data, and (3) good decision-making skills. CDS systems are increasingly important in aiding clinical decision-making due to the following key challenges faced by clinicians:

Exponential Growth of Medical Knowledge Provision of optimal care is dependent on the clinician's ability to obtain relevant, up-to-date knowledge. The body of medical knowledge in the era of Galenic medicine appears to have been quite static during the lifetime of a clinician (the Galenic era lasted for more than 1300 years from 300 CE to the seventeenth century, when Galen, a Greek physician, heavily influenced medicine). Today, however, medical knowledge is increasing in volume and complexity. In 1950, the doubling time of medical knowledge was estimated to be 50 years; it decreased to 7 years in 1980 and to a mere 73 days in 2020 [2]. Furthermore, the traditional histopathological classification of disease, which has been the way medical knowledge has been organized and taught for over a century, is giving way to a more fine-grained molecular and functional subtyping of disease. The volume and rapidly evolving genomic, proteomic, metabolomic, and other - omic characteristics of disease and health make it impossible for a clinician to remember and apply them in clinical care without some form of assistance.

Rapid Accumulation of Patient Data The amount of clinical data per individual is rising, driven by the widespread adoption of **electronic health record** (EHR)

systems and the rapid growth of new laboratory tests, investigations, and imaging that are increasingly used in clinical care (see Chap. 3). For example, in critical care, it is estimated that a patient generates an average of 1460 new data points daily [3], and a clinician is exposed to an average of 4380 data points during a shift of 12 h [4]. The large amount of patient data has led clinicians to spend more time reviewing and collating data in the EHR that are pertinent to the current clinical context.

Increase in Inference Complexity Human clinicians have limited cognitive capacity and can simultaneously consider only a few variables at a time in decision-making (see Chap. 5). With the exponential growth of medical knowledge and the rapid accumulation of patient data, good medical decision-making requires the consideration of many facts. With individual genomic and proteomic data becoming available for making decisions, inevitably, the number of salient facts to consider for a clinical decision will rise steeply [5]. As the number of facts to consider for clinical decision-making outstrips human cognitive capacity, CDS systems are needed to aid the clinician [6] (see Chap. 5).

Clinical Data Capture Clinicians, particularly in the United States, face an increasing amount of clinical documentation that reduces the time available for direct patient care. For example, primary care physicians spent 42% of their time (5.9 h of an 11.4-h workday) in the EHR, of which half the time is spent on documentation, order entry, billing, and coding [7].

Artificial Intelligence-Based CDS

AI has a long history that traces its modern roots to the 1956 Dartmouth meeting, where computer scientists discussed the notion of AI with the ultimate aim of building machine systems that can perform human-like intellectual and cognitive tasks (see Chap. 2). **Machine learning** (ML), which has come to constitute the largest subset of AI in recent years, refers to AI systems that can achieve some aspects of human-like intelligence without being explicitly programmed by human authors. In particular, **deep learning**, an important subfield of ML, relies on learning large neural networks, often from massive datasets (See Chaps. 1 and 6). Since the inception of AI, medicine has been identified as one of the most promising application areas. Many AI-CDS systems have been described and implemented for a panoply of tasks in medicine, from risk assessment and diagnosis to prognosis and therapeutics to patient monitoring and interpretation of human genomes.

The typical structure of an AI-CDS system has two main components: a knowledge component that represents medical knowledge in a computable form and an inference component that applies the knowledge to a patient's data to provide decision support (see Fig. 10.1). Different ways have been developed for representing

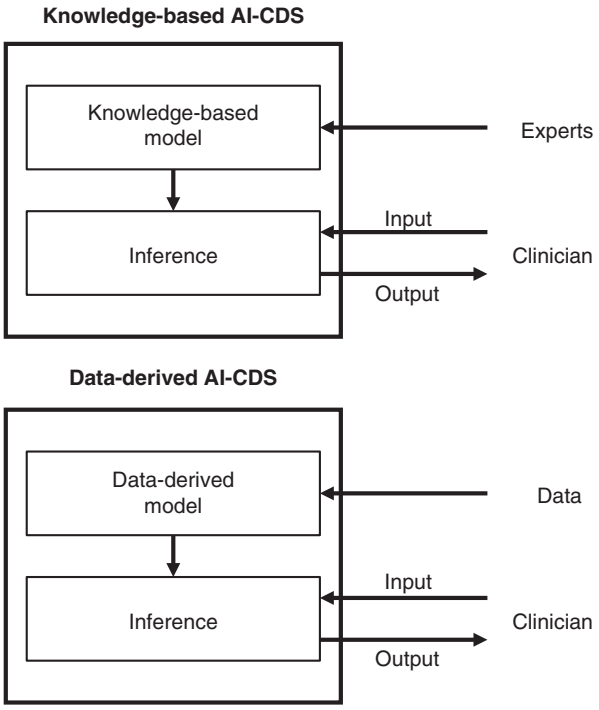

Fig. 10.1 The key components of a knowledge-based AI-CDS system include a knowledge base such as expert-derived rules and an inference mechanism for clinical application such as chained inference for rules. The key components of a data-derived AI-CDS include a model, such as a data-derived neural network, and an inference mechanism for clinical application, such as forward propagation in a neural network model

knowledge, such as rules and Bayesian networks, and a variety of inference mechanisms have been created, including chained inference for rules and probability calculations for Bayesian networks. Historically, the knowledge base is explicitly derived from human experts. With the advent of ML and deep learning, the knowledge component is replaced with computational models, such as classification trees and neural networks that capture relations among domain concepts. Models are then applied to a patient's data to provide outputs such as predicting a clinical outcome. Typically, models are derived from data and often big data. We can view a knowledge base as constituting a model as well. By doing so, models become a unified representation that can be constructed from data, knowledge, or both. We will refer to systems in which models are derived primarily from data as *data-derived systems*. Similarly, *knowledge-based systems* will refer to systems derived primarily from human knowledge (see Chap. 4).

Types of AI-CDS Broadly speaking, AI-CDS can be categorized into knowledge-based and data-derived systems (see Fig. 10.1). Early AI-CDS systems that were developed in the 1970s and 1980s used knowledge-based approaches in which medical knowledge represented as rules, expert constructed Bayesian networks, and semantic networks are stored in a **knowledge base**. In rule-based systems (Chaps. 2 and 4), knowledge is expressed in IF … THEN ... expressions; for example, in a diagnostic system, the IF part would typically encode symptoms, and the THEN part would encode diseases that manifest those symptoms. **Knowledge-based**

AI-CDS enjoyed early success and AI-CDS systems were used, for example, to choose appropriate treatments, interpret electrocardiograms, and generate diagnostic hypotheses in complex clinical scenarios. Their key advantages include knowledge that is represented in a form that is easy for clinicians to comprehend and the ability to explain inference in clinically meaningful ways (see Chaps. 1, 4, and 8). However, the construction of knowledge bases is typically manual, which can be time-consuming and tedious, and updates to the knowledge are also manual and slow. Additionally, the construction of stores of numerical knowledge, such as probabilities, as for example, in Bayesian networks, is tedious and difficult.

The data-derived approach to developing AI-CDS systems began in the 1990s. In these systems, knowledge of the earlier AI-CDS was replaced by models that were automatically derived from data. Typically, these models are computational objects that have structural and numerical components. For example, in a neural network model, the network architecture consisting of connections among layers of nodes constitutes the structure, and the weights assigned to connections constitute the numerical component. ML and deep learning methods (see Chap. 6) have been used to derive a wide range of models. ML methods have been developed to derive from data rules and probabilistic networks, resembling manually constructed knowledge-based AI-CDS models. The key advantages of **data-derived AI-CDS** include the ability to rapidly construct models that can have excellent performance. A key disadvantage is that the models are often opaque to human experts, and the explanation of inference using these models remains impenetrable to human users (see Chap. 8).

The impetus for the widespread application of ML to medical data came from advances in data availability, the development of a broad range of ML methods, and powerful and ubiquitous computing capability. First, data on health and disease are increasingly available and include a broad range of data types. In addition to experimental data that are typically collected in research studies under controlled conditions, observational data are becoming available from sources such as EHRs, social media, and monitoring through mobile smartphones. Second, a broad range of ML methods has been developed and is readily available as computer programs for application. Third, access to faster and ever more powerful computers is becoming inexpensive and ubiquitous.

Until recently, data-derived AI-CDS systems were static, implying that the computable knowledge learned from data is not updated. **Static AI-CDS** provides the same result each time the same input is provided, and they do not evolve over time and do not use new data to alter their results. This approach has the limitation that a static model may become obsolete when the conditions in which it was applicable change, for example, changes in the characteristics of a hospital's patient population. This limitation has led to the development of **adaptive AI-CDS** in which the CDS is dynamic in that it can learn and change performance over time, incorporating new data and new methods for learning from data [8]. An adaptive CDS that predicts the risk of cardiovascular disease would refine the predictive model in several ways: for example, the model might be slightly different at each institution where it is deployed, reflecting geographic or population variations, or an

institution's model may be continuously updated based on more recent data from that institution.

Machine Learning and Data-Derived AI-CDS Systems As described in the previous section, data-derived models typically consist of structural and numerical components. While most models are derived automatically from data using ML approaches, they can be hand-crafted by human experts or constructed by a hybrid process where the model structure is hand-crafted, and the numerical component is derived automatically from data. ML models capture patterns in data, and these patterns are often used to make predictions and also may lead to the discovery of new knowledge. ML methods can be categorized broadly into supervised, unsupervised, semi-supervised, deep, and causal learning (see also Chap. 6).

Supervised ML leverages data that contain cases that consist of input variables (such as symptoms, signs, and laboratory test values) and corresponding output labels (such as the presence or absence of myocardial infarction). By analyzing the patterns in the data, supervised ML constructs a model that seeks to produce the correct output when provided with the input on new cases. When the output is discrete and has a limited number of labels (e.g., presence or absence of myocardial infarction), the supervised ML is called classification. When the output is numerical and has a large number of values (e.g., height), the supervised ML is called regression.

In contrast to supervised ML, **unsupervised ML** uses data that contain cases with only input variables but no output labels. Unsupervised ML infers patterns in the data such as clusters, outliers, and low-dimensional representations. Clusters are groups of cases that are similar in some way. Outliers are cases that are very different from the other cases in the data. Low-dimensional representations represent cases with a smaller number of features (variables) than are present in the raw data.

Semi-supervised ML is concerned with learning from a combination of data that contain outputs (e.g., diagnostic labels) and data that do not. This type of ML extends the applicability of both supervised and unsupervised ML, which traditionally can use only labeled and unlabeled data, respectively.

The current advances in ML are largely driven by **deep learning,** which involves training artificial neural networks with many layers on large amounts of data. Compared to the other types of ML described so far, deep learning has the advantage of automatically selecting relevant features in the data, creating complex features from simpler ones, and deriving a large number of relations, both simple and complex, from big datasets.

Another advance in ML that is particularly applicable to medicine is **personalized ML**. The typical ML approach is to derive a single model from training data, such that the model is optimized to perform well on average on all future individuals. This **population ML** approach has been quite successful; however, it may ignore important differences among patients, such as differences in the mechanisms causing disease or in treatment response. An approach for better capturing individual differences is personalized ML, where the model is tailored to the characteristics of the current individual and is optimized to perform especially well for that individual, but not necessarily for all future patients [9–11]. For example, the breast cancer of a current

patient may have a mutation *W* that is highly predictive of the cancer course, although it is rare. A mutation *X* is much more common in the breast cancer population; however, it is only modestly predictive. A population model is likely to include *X* as a predictor, but not *W*, because mutation *X* is so common. That model would predict the cancer course of the current patient only fairly well. In contrast, a personalized model would likely include *W* as a predictor and predict the cancer course quite well.

Causal ML is concerned with modeling and discovering causal relationships [12, 13]. Such relationships predict the values of one or more variables after we *set* the values of other variables independently. Such predictions are important when making decisions to optimize expected outcomes, as is common in healthcare. For example, deciding on the best therapy for a patient involves making causal predictions. In contrast, most of the research and methods in ML have focused on learning models that predict one or more variables after we *observe* the values of other variables. Patient diagnosis in light of existing patient information is an example of an observational prediction.

Sometimes correct causal and observational predictions yield the same answers, but other times they do not. Figure 10.2a shows a situation in which *X* causally influences *Y*, and there are no other sources of statistical dependence between *X* and *Y*. In this example, the causal prediction of *Y* given that we independently set *X* equals the prediction of *Y* given that we observe *X*. In such a scenario, we can estimate model parameters using observational data and apply the resulting model to make causal predictions. Figure 10.2b is an example in which the causal and observational predictions differ, due to the presence of a hidden (latent) variable *H*. Here the observational prediction of *Y* given that we observe *X* is determined by the association due to *X* directly causing *Y* and association due to the path from *X* to *H* to *Y*, which is not due to *X* causing *Y*. In contrast, the causal prediction of *Y* given that we independently set $X = x$ [13] involves the situation shown in Fig. 10.2c. By independently setting *X*, we break the non-causal source of association between *X* and *Y* that goes through *H,* and we predict *Y* based only on the causal influence of *X* on *Y* [14].

For most of the past century, the predominant, formal method for causal discovery in healthcare and beyond has been the randomized controlled trial (RCT) [15]. By randomizing the setting of the value of *X* (e.g., a treatment selection), its value

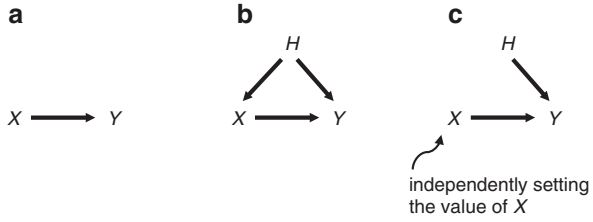

Fig. 10.2 Examples of causal Bayesian networks. *X* and *Y* are measured variables. *H* is a hidden (latent) variable. (**a**) *X* causally influences *Y*, and there is no confounding. (**b**) *X* causally influences *Y*, and there is hidden confounding. (**c**) Independently setting *X* removes the hidden confounding of *X* and *Y*

is set independently of the values of any of the other measured variables; thus, for example, the situation in Fig. 10.1c results, where the only dependency between *X* and *Y* is due to direct causation. On the other hand, RCTs are often expensive, sometimes infeasible, and frequently they study only a small, selected subset of patients, relative to the broader population of interest. Conversely, observational data, such as EHR data, are relatively plentiful, contain a rich variety of types of information, and more faithfully represent "real-world populations." However, care must be taken in deriving causal knowledge solely from observational data. A commonly used causal model is the causal Bayesian network, which is a Bayesian network in which a directed edge from *X* to *Y* represents that *X* is a direct cause of *Y*, relative to a set of modeled variables (as, for example, in Fig. 10.2).

ML methods have been developed that derive causal models from data, including observational-only data, or from a combination of knowledge and data. For instance, methods exist for learning Bayesian networks from expert knowledge and data [16, 17]. Expert knowledge could define an initial model for a system that provides diagnostic, prognostic, or therapeutic advice, for example. As data accumulate, the model adapts to represent the causal relationships consistent with the data. Causal modeling could also support the development of adaptive CDS systems. In the context of a given clinical task, such a system could compare its causal model of a domain with its model of a clinician's causal knowledge of the domain to provide advice to the clinician that optimally augments what he or she is likely to already know [18, 19].

Degree of Automation in AI-CDS

The early AI-CDS systems were standalone; the clinician interacted with the system by manually providing relevant patient data as input and then incorporating the system's output with their judgment to make clinical decisions. The widespread adoption of EHR systems has enabled increased integration of AI-CDS with such systems. AI-CDS may be integrated with EHR systems to a varying extent that enables the AI-CDS to obtain inputs automatically from the EHR, make recommendations, and provide those recommendations to the clinician and output them to the EHR (see Fig. 10.3).

Based on the degree of automation and integration with EHR systems, AI-CDS can be broadly categorized into three types [20]. In conventional AI-CDS, the CDS system collects patient data from the EHR and provides recommendations that the clinician receives, clarifies and considers in making the final decision. In integrative AI-CDS, the CDS system actively obtains patient data from the EHR, provides recommendations to the clinician, and also automatically records them in the EHR. The clinician still makes the final decision. In fully automated AI-CDS, the CDS system gathers information about and from a patient, makes decisions autonomously, and records results in the EHR. The clinician may monitor the recommendations and clarify the CDS system's recommendations. For some clinical tasks, fully automated decision support may be suitable, for example, some steps in robotic

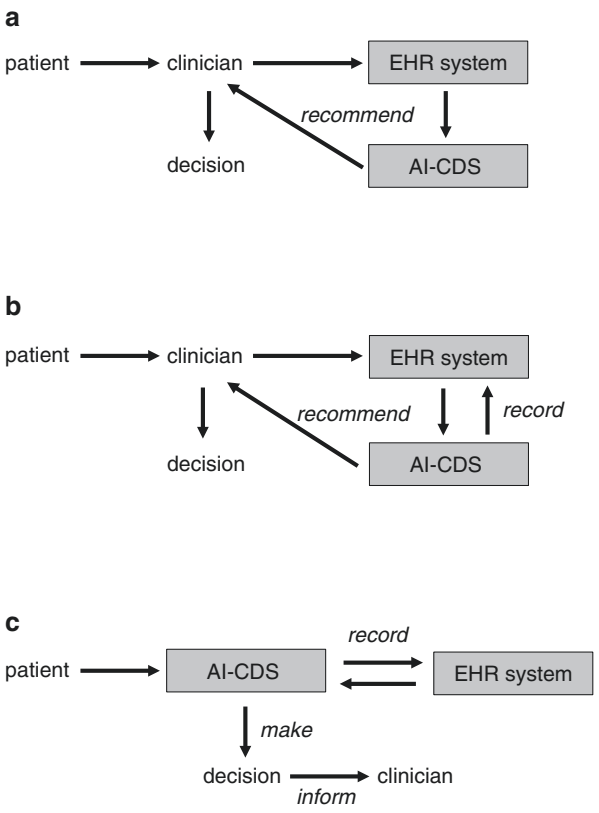

Fig. 10.3 Categories of AI-CDS based on the degree of automation and integration with EHR systems. (**a**) Conventional AI-CDS obtains patient data from the EHR system and provides recommendations to the clinician who makes decisions. (**b**) Integrative AI-CDS obtains patient data from the EHR system, provides recommendations to the clinician, and records them in the EHR system. (**c**) Fully automated AI-CDS collects information from the patient, makes decisions, and records them in the EHR system

surgery or in insulin dose adjustments by an insulin pump; in many more tasks, however, integrative decision support will be more practical in the foreseeable future with the final decisions made by the clinician.

AI-CDS systems are often based on an input-process-output workflow. Inputs can come from various sources such as data from EHR and medical imaging systems and devices such as mobile smartphones, Fitbit, Apple, and other health trackers (see Box 10.1). Outputs can be delivered in many ways. Examples include diagnoses, recommendations, alerts and reminders, order sets, relevant medical knowledge, and context-aware summaries (see Box 10.1).

Box 10.1 Example Input and Output Modalities for AI-CDS
Examples of input modalities for AI-CDS

- **EHR data**. Radiology reports of patients with acute traumatic injury are input to an AI-CDS that identifies incidental findings for follow-up care.
- **Medical images.** Screening mammograms are input to an AI-CDS that automatically identifies potential cancer.
- **Health sensors.** A wrist band with gyroscopic sensors worn by an individual at risk for fall provides input to an AI-CDS that automatically detects falls.

Examples of output modalities for AI-CDS

- **Highlighting in EHR.** An AI-CDS identifies important new patient data in the EHR and highlights them to the clinician.
- **Alerts.** On detecting strokes in CT images of the brain, an AI-CDS sends alerts to stroke clinicians.
- **Discharge summaries.** An AI-CDS automatically generates discharge summaries to support communication during the transition of care from hospital to community care.

Application of AI-CDS in Clinical Care

As described in the section on "Challenges Faced by Clinicians", clinicians face challenges in the daily practice of medicine that arise from the exponential growth of medical knowledge, rapid accumulation of a diversity of patient data, and the increased complexity of clinical decision-making. Clinicians perform a range of tasks, such as assessing the risk of developing a disease in the future (risk assessment and stratification), determining the presence or absence of disease at the current time (diagnosis), forecasting the likely course of disease (prognosis), predicting treatment response (therapeutics), and monitoring in acute care, such as in critical care and during surgery, as well as outside the hospital for chronic diseases (see Chaps. 11, 12, and 14). The remainder of this section provides examples of areas of rapid progress in the development of AI-CDS.

Providing Relevant Medical Knowledge Studies have shown that clinicians have knowledge needs in many aspects of clinical decision-making, including diagnosis, prognosis, and therapy during patient encounters [21]. CDS systems have been developed that provide relevant medical knowledge at the right time and at the right place, such as the **Infobutton** that collates medical knowledge from the literature, textbooks, and other sources of information and presents knowledge relevant to a particular clinical context [22]. More recently, AI-CDS approaches have been described for generating medical evidence for treatments in a specific clinical context when such knowledge is lacking in the medical literature or in published treatment guidelines. One approach to this situation that has been described is to generate evidence from the EHR and other health utilization data of similar patients [23]. For a clinical question, the approach specifies the relevant population, intervention, comparator, outcome, and timeframe to select data from a large database such as a hospital's EHR data warehouse, which is used for treatment-effect estimation and survival analysis. However, such estimates may be subject to bias due to idiosyncrasies in the hospital's EHR data and due to hidden confounding and selection bias.

Prioritization of Patient Data In a specific clinical context, relevant patient data should be readily available for optimal decision-making. However, in current EHR

systems, retrieval of patient data relevant to a clinical task is cumbersome and time-consuming, exacerbated by confusing layouts, workflows, poor prioritization, and weak search capabilities. Clinicians spend substantial amounts of time searching large volumes of data to identify clinically meaningful patterns and important patient details, predisposing them to information overload. AI-CDS systems are needed that intelligently identify and display clinically relevant patient data that enhance the clinician's ability to rapidly assess the clinical context and make optimal decisions. The learning electronic medical record (LEMR) system uses ML models to highlight data and are trained from output labels that clinicians have identified in past patient cases. In a research study, the LEMR system was able to identify and highlight salient patient information to summarize the clinical status of the patient for morning rounds in the critical care setting (Fig. 10.4) [24].

Risk Assessment Data-derived AI-CDS is increasingly developed to predict the risk of developing a disease or monitor adverse clinical events. For example, a deep learning strategy that combines results from cognitive testing and magnetic resonance imaging (MRI) of the brain predicts the risk of developing Alzheimer's disease [25]. As another example, an ML-based system that predicts the risk of hypoxemia in the near future and explains the risk factors during general anesthesia was developed to aid anesthesiologists in early intervention [26].

Diagnosis The application of ML and deep learning approaches for diagnosis in medical imaging has rapidly grown in recent years in the areas of radiology, ophthalmology, dermatology, pathology, cardiology, and gastroenterology. In radiology, clinicians rely primarily on imaging for diagnosis, and deep learning methods have rapidly improved the performance of diagnostic tasks in images. For example, the automated diagnosis of common lung diseases with chest radiography [27], the detection of lung nodules with computed tomography (CT) [28], and the identification of breast tumors using mammography [29] have achieved expert-level diagnostic accuracies. In dermatology, clinicians rely on visual inspection of skin lesions to diagnose and differentiate between benign and malignant lesions. For example, neural networks can identify malignant melanomas from a single photograph of the lesion at a dermatologist's level of accuracy [30]. In ophthalmology, fundus photographs are visually examined by ophthalmologists to detect and monitor various diseases, such as glaucoma and diabetic retinopathy. In a recent application of deep learning, neural network models were able to identify diabetic retinopathy at an accuracy comparable to that of ophthalmologists [31]. In pathology, histopathological assessment under the microscope of biopsied specimens by pathologists is used to diagnose many types of cancer. Deep learning models have been shown to be useful in detecting prostate cancer from biopsy specimens [32, 33] and identifying breast cancer metastasis in lymph nodes [34]. Cardiologists use electrocardiograms and echocardiograms, and deep learning methods have recently been shown to perform at expert-level accuracy for diagnosing heart attacks, as well as cardiac abnormalities like hypertrophic cardiomyopathy, from electrocardiograms [35]. Identification of small polyps during colonoscopy is an arduous task for gastroen-

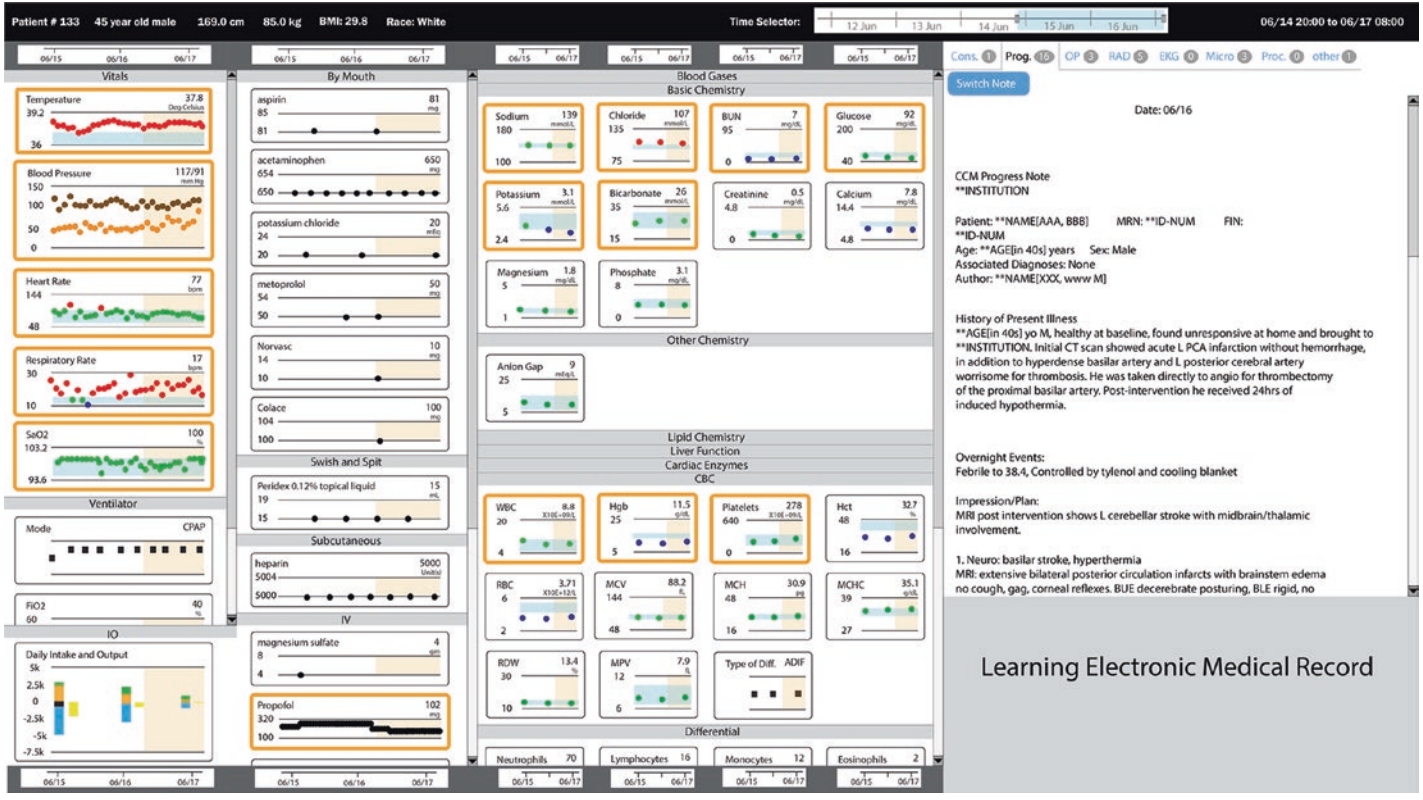

Fig. 10.4 The learning electronic medical record (LEMR) system showing relevant patient data that are highlighted with orange borders. Patient data are arranged in four vertical columns: the left-most column displays vital sign measurements, ventilator settings, and fluid intake and output; the second column shows medication administrations; the third column displays laboratory test results; and the right-most column displays free-text notes and reports. Names have been removed, and the demographic data and dates have been changed to preserve patient confidentiality

terologists. Recently, an ML-based approach that identifies polyps in images from a colonoscopic camera was shown to enhance the clinician's speed and accuracy of detecting polyps during colonoscopy [36, 37]. The future translation into clinical applications of the successful application of AI, especially deep learning for image-based diagnosis, will significantly change current medical practice. Curt Langlotz, a pioneer in AI in radiology, posed the question, "Will AI ever replace radiologists?" then answered, "I say no – but radiologists who use AI will replace radiologists who don't" [38].

Early diagnosis of rapidly developing clinical conditions is another area of abundant application of ML approaches. For example, in critical care, early diagnosis of sepsis using ML models has been shown to be more accurate than traditional tools such as the quick Sepsis Related Organ Failure Assessment (SOFA) score [39].

Prediction of Clinical Outcomes Prediction of clinical outcomes with ML has grown rapidly with the increased availability of large volumes of EHR and health insurance claims data. ML models can learn the patterns of health trajectories from EHR and other data of a large number of individuals, and such models can anticipate future events at an expert clinical level. For example, accurately forecasting the likely clinical course in a patient with community-acquired pneumonia enables decision-making about whether to treat the patient as an inpatient or as an outpatient [11]. Similar ML-based forecasting can identify recently discharged patients who are likely to develop complications requiring readmission or patients who are at risk for prolonged hospitalization [40]. Such information can be used proactively to provide additional resources or initiate more intensive management. Furthermore, Bayesian networks have been developed to predict mortality, readmission, and length of hospital stay using EHRs from the emergency department [41], and deep learning applications have been developed to predict in-hospital mortality, 30-day readmissions, and prolonged length of hospital stay [40]. ML has also been applied to identify patient characteristics in the medical notes to classify cancer patients with different responses to chemotherapy [42].

Therapy Therapeutic CDS systems that aid in choosing the best therapy have been developed since the inception of CDS systems. One of the earliest such systems was MYCIN (see Chap. 2), a rule-based system that uses backward chaining inference to identify causative bacteria in infections and recommend appropriate antibiotics and dosages. Examples of AI-CDS are found in the field of radiomics that use AI-based analyses of clinical images to characterize tumor phenotypes and predict treatment response. For example, a deep learning approach using radiomic features in CT scans of non-small cell lung cancer was able to predict treatment response to various therapeutic agents [43].

Alerting Alerting CDS systems have been developed for a long time to draw the clinician's attention to the important data at the right time. One of the earliest such systems was the **HELP** system that was developed at LDS Hospital in Salt Lake City in the 1960s and generated automated alerts about abnormalities in patient data

[44]. Several alerting AI-CDS systems have been described in recent years. A deep learning approach has been developed and deployed that sends alerts to stroke clinicians on detecting strokes in CT images of the brain [45]. More recently, an ML approach for detecting anomalous patient-management decisions in the critical care unit was developed and evaluated at the University of Pittsburgh [46].

Patient Monitoring In-hospital patient monitoring is an essential clinical activity in operating rooms, intensive care units, and emergency rooms. Real-time detection of critical events from data generated by monitoring devices is an area where ML is increasingly applied. For example, ML methods can identify seizures from continuous electroencephalographic monitoring [47]. As another example, such methods can predict hypoxemia events during surgery from continuous physiological monitoring [26]. These results suggest that the application of ML methods to continuous patient-monitoring data can achieve accurate and timely predictions, thus relieving information overload on clinicians.

Clinical Data Capture A significant contributor to clinician frustration and burnout is the undue length of time spent in documenting encounters, often at the cost of decreased time spent interacting with patients. Clinical scribes, who work alongside clinicians to translate and record information in clinical encounters, were introduced to reduce the burden of documentation on clinicians. More recently, digital scribes that leverage speech recognition and natural language processing are being developed to capture and document the spoken portions of the clinical encounter automatically [48]. Advances in **human-computer interaction** technologies, such as speech and gesture recognition and ambient listening and seeing, will likely lead to the development of autonomous digital scribe systems that allow clinicians to migrate from interacting with a standalone computer to speaking in an intelligent room where the environment itself becomes the automated scribe.

Pitfalls of AI-CDS

In a recent application of ML to detect pneumonia in chest X-rays, the ML model performed successfully, detecting pneumonia with an accuracy of 93% when the model was evaluated on a different batch of X-rays at the institution where the model was developed. However, when the model was evaluated on a batch of X-rays from a different institution, its performance in detecting pneumonia fell to 73% [49]. It was subsequently found that the X-rays of pneumonia had been mostly taken from very sick patients lying down with portable chest X-ray machines, and X-rays of patients lying down look very different from X-rays of patients who are standing up, and the model had learned to discriminate between X-rays of patients lying down from standing up, rather than identifying features of pneumonia. This is an example of a pitfall of data-derived AI-CDS due to an inadvertently introduced bias

in the data that was used to derive the ML model. Translation of ML research into clinically robust AI-CDS requires mitigating a range of such pitfalls.

AI-CDS has significant pitfalls that include dataset shift, algorithmic bias, automation complacency and inscrutable explanations. We discuss each of these problems in the remainder of this section.

Dataset shift is a common pitfall in ML that occurs when data characteristics differ between the training phase and the application or deployment phase. It is common and occurs for reasons that range from the bias in the training data to the application of the ML system to an inappropriate clinical context. The availability of high-quality training data may be limited if, for example, portions of the data need manual review by experts, if the outcome is poorly defined, or the available data are a convenience sample that is not representative of the entire population. Sometimes, dataset shift is introduced by the process of training; for example, the training data may have been adjusted to contain an equal number of cases and controls to maximize the performance of the ML system; however, at application time, it is rarely the situation in medicine that the condition of interest occurs 50% of the time.

Dataset shifts are common across locations and across time. Thus, ML models developed at one location may perform poorly at a different location because disease patterns are different across the two locations. Further, even within the same healthcare system, models that are developed from data on patients who attend a specialty clinic may perform poorly on the general population. For example, an ML model that is trained on photographs of skin lesions in a dermatology clinic may have lower accuracy when applied to patients seen in a primary care clinic where the appearance of lesions, and the risk profile of patients, are different.

Even at the same location, disease patterns can change over time, leading to a decrease in performance in the future. Models developed only from historical data will reinforce existing practice and may not reflect new medical developments and changes in policies and guidelines. For example, an AI-CDS system might erroneously recommend a drug after it has been withdrawn due to safety concerns or will not recommend a medication appropriately whose use has been expanded to the treatment of new conditions.

It is important to monitor and update ML models because unanticipated dataset shifts will almost certainly occur, and the performance of deployed models is likely to deteriorate. Thus, AI methods are needed to detect when shifts have occurred, identify the nature of the shifts, and continually update the models using more recent data.

Algorithmic bias refers to errors in an AI-CDS that systematically underperform for one group of individuals relative to others. Algorithmic bias exacerbates existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability, and sexual orientation, and it may amplify inequities in healthcare systems. Bias arises due to many factors; however, the common problem is that the data used in training ML models often do not represent the whole population, leading to poor performance in underrepresented groups. Most data used for ML are

observational data that are often limited by low diversity in race, gender, geographical location, economic conditions, and other important attributes. Training with such biased data can lead to biased ML models that are not valid for parts of the population, and the application of such models has the potential to exacerbate existing healthcare disparities.

As examples, ML models trained with gender-imbalanced data perform poorly at reading chest X-rays for the underrepresented gender [50]; and ML models trained primarily on light-skinned individuals perform poorly in detecting skin cancer affecting individuals with darker skin [51]. A recent study reviewed over 70 publications and noted that most of the data used to train ML models came from just three states in the United States [52], suggesting the potential for geographic bias. As another example, a commercial risk model for predicting future risk of needing complex healthcare exhibited racial bias. For the same level of predicted risk, black patients were found to be sicker than white patients because the model was trained on healthcare costs as a proxy for healthcare needs. Since less money had been spent on black patients who have the same level of need, the model inaccurately predicted that black patients are healthier than white patients [53].

Beyond problems with the data, algorithmic bias can arise at any point in the development of an AI-CDS system from data collection and cleaning, model choice, the protocol used in training and evaluation, and implementation and dissemination. Preventing algorithmic bias requires that the teams that develop AI-CDS include experts who have knowledge about how to prevent bias and not simply data scientists who are technical experts in ML. Particularly, clinicians and even patients should be included in the teams, as they can provide deep insights into the clinical context [54].

Automation Complacency With the deployment of autopilots in aircrafts and, more recently, in automobiles, it has been observed that pilots often failed to monitor important flight indicators, and drivers in autonomous automobiles frequently failed to watch the road. Similar behavior has been noted to occur with clinicians using AI-CDS systems. If an AI-CDS system were completely accurate and reliable, then clearly following its recommendations would lead to positive outcomes; however, practical AI-CDS systems are not perfect and can increase errors if incorrect advice is followed. Over-dependence on CDS in conjunction with reduced vigilance in information seeking and processing is termed automation complacency, which can lead to errors that would not normally occur in the absence of CDS [55]. Automation complacency can result in omission errors in which the clinician fails to consider relevant medical knowledge or patient information because the CDS did not recommend it, and commission errors where the clinician complies with incorrect CDS recommendations.

For example, an AI-CDS system that aids in detecting cancers in screening mammograms can increase the rate of cancer detection by uncovering those that the radiologist would otherwise miss. However, omission errors by the AI-CDS will result in cancers going undetected, and commission errors may result in individuals without cancers receiving unnecessary interventions [56]. Similar errors due to

automation complacency occur in the computerized interpretation of electrocardiograms [57], decision support in e-prescribing [58], and answering questions about clinical scenarios [59].

The factors causing automation complacency are multifactorial; they include complex tasks that impose a greater cognitive load, low clinician experience with a task, and high trust in the AI-CDS system, especially as familiarity with the system grows over time. Mitigating automation complacency is a challenging, open problem, and interventions, such as providing clinicians with information on the AI-CDS system's reliability, have had little impact. One potential solution to this problem is having an AI-CDS that balances sometimes offering advice upfront with sometimes only offering critiques post facto.

Inscrutable Recommendations With the increasing complexity of AI models that underlie CDS, explanations that describe the basis of recommendations or predictions are important to detect error or bias, as well as to engender trust in the system (see Chaps. 1, 2, and especially Chap. 8). The insight that an explanation provides about why a patient is at high risk of developing a disease can help a clinician understand the reasoning, which helps gain trust in the AI-CDS system. In knowledge-based AI-CDS, such as rule-based systems, and some ML models, such as classification trees, the reasons for the resulting predictions can be clearly explained. Other ML models, such as random forests and artificial neural networks, often perform better than earlier models, but their black-box nature makes their recommendations more inscrutable (see Chap. 8).

A wide range of methods, which can be broadly categorized into ante-hoc and posthoc approaches, are being developed to provide explanations for AI-CDS systems. In the ante-hoc approach, the AI-CDS system is designed to be interpretable, and such systems have a long tradition in medicine and are created from expert knowledge and employ human-AI interaction. For example, MYCIN was designed as a consultation system with explanatory capabilities to advise clinicians on diagnosing and treating bacterial infections. The MYCIN system conducts a question-and-answer dialog to elicit relevant patient data, and the execution of the rules forms a coherent explanation of MYCIN's reasoning [60].

Posthoc approaches aim to provide explanations for a specific recommendation and are more applicable to modern ML models that are not designed for interpretability (see Chap. 8). For example, in deep learning-based AI-CDS systems for medical imaging, a post-hoc approach uses saliency maps. In a saliency map, the explanation highlights the salient regions in the image that are important to the system's recommendation, such as the regions on the chest X-ray or the picture of a skin lesion that most contributed to the recommendation. Beyond image analysis, model-agnostic explanatory methods that focus on explaining individual recommendations of a black-box ML model have been developed. Examples of such methods include Local Interpretable Model-Agnostic Explanations (LIME) [61] and SHapley Additive exPlanations (SHAP) [62]; these methods estimate the impact of input features for a specific prediction from analysis of the behavior of the model when the inputs are varied.

Regulation of AI-CDS

As AI-CDS systems become more complex, automated, and adaptive, they will surpass the ability of clinicians to independently verify their veracity, which makes regulatory oversight vital to shield patients from the pitfalls of such systems (see also Chap. 18). Depending on the complexity, the regulatory requirements for AI-CDS can range from none at all to substantial compliance burden. For example, in the outpatient clinic, a clinician receives a CDS recommendation to offer colonoscopy for a patient who is 45 years of age. The clinician can easily verify the accuracy of the recommendation, given the U.S. Preventive Services Task Force guidelines on which the recommendation is based. Such a CDS system would not require regulation.

In contrast, consider an AI-CDS system that uses an ML model to recognize cardiopulmonary instability from continuous physiological monitoring of the cardiac and respiratory systems. Such a system may be deployed in the critical care unit to monitor and predict the elevated risk of cardiopulmonary instability, and a prediction of elevated risk may lead to decisions such as initiation of medication to increase the blood pressure or mechanical ventilation. In this situation, the clinician cannot readily assess the accuracy of the assessment provided by the AI-CDS, and such a system would need to be regulated to ensure patient safety.

AI-CDS systems consist of software, and software may be deemed a medical device if it is used to guide clinical decision-making. The U.S. Food and Drug Administration (FDA) has created guidelines for regulating Software as a Medical Device (SaMD) that encompasses AI-CDS. The FDA guidelines are based on recommendations from the International Medical Device Regulators Forum (IMDRF), an international group of medical device regulators that develops guidelines for the uniform regulation of medical products worldwide.

There are many important factors in the regulatory framework of AI-CDS, including risk assessment, unbiased training, reproducibility, and whether the AI methods in the CDS are static vs. adaptive. The FDA provides a framework for the clinical evaluation of SaMD that is adopted from the IMDRF. The goal of the clinical evaluation is to assess a SaMD's clinical safety, effectiveness, and performance as intended by the developer of the SaMD. The clinical evaluation consists of three components that include scientific validity, analytical validation, and clinical validation (see Table 10.1). A SaMD must pass all three components successfully to be considered validated. Further, following the IMDR, the FDA stratifies SaMD into four risk levels based on the intended medical purpose of the SaMD (treat or

Table 10.1 Components of clinical evaluation of Software as a Medical Device (SaMD)

Clinical evaluation		
Valid clinical association (scientific validity)	Analytical validation	Clinical validation
Is there a valid clinical association between the SaMD's output and the SaMD's targeted clinical condition?	Does the SaMD correctly process input data to generate accurate, reliable, and precise output data?	Does the use of SaMD's accurate, reliable, and precise output data achieve the intended purpose in the target population in the context of clinical care?

Adapted from [63]

Table 10.2 Regulatory requirements for Software as a Medical Device (SaMD) by the intended medical purpose and the nature of the patient's condition

	Intended medical purpose		
Nature of the patient's condition	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

I: least regulatory requirements, IV: greatest regulatory requirements. Adapted from [63]

Table 10.3 Examples of AI-CDS that have received FDA clearance as SaMDs

Name of device or algorithm	Name of parent company	Short description	FDA approval number	Date	Medical specialty
Arterys Cardio DL	Arterys Inc	Analysis of cardiovascular magnetic resonance images	K163253	2016/11	Cardiology
ContaCT	Viz.ai	Automated stroke detection on CT images	DEN170073	2018/02	Radiology
EyeArt	Eyenuk, Inc	Automated detection of diabetic retinopathy on retinal fundal images	K200667	2020/06	Ophthalmology

diagnose, drive clinical management, inform clinical management) and the nature of the patient's condition (critical, serious, non-serious). A higher level of risk requires increased oversight, more regulatory requirements, and more evidence for the efficacy and safety of the SaMD (see Table 10.2).

The FDA certified the first AI-CDS system in 2016 when Arterys became the initial company to receive clearance to use deep learning in a clinical setting for the analyses of cardiovascular images. As of January 2021, a total of 71 AI-CDS systems have been cleared by the FDA as SaMDs. The largest number of AI-CDS systems certified by the FDA are in the fields of radiology and cardiology [64]. Table 10.3 provides examples of AI-CDS systems that have received FDA clearance.

Conclusions

CDS is at a critical juncture for the safe and effective integration of AI into clinical care. The technical capacity to develop, implement, and maintain AI-CDS in the clinical enterprise is increasing by leaps and bounds, and the promise of AI in clinical decision-making offers considerable opportunities to improve patient outcomes, reduce costs, and improve population health.

AI-CDS is poised to advance the learning health system in which clinical experience and patient data are systematically integrated to provide higher quality, safer, more efficient care. Clinical trials and similar research underlie one of the key ways of generating new knowledge and evidence for improving clinical care. The clinical enterprise of treating patients and the research enterprise of evaluating new therapies, for the most part, are segregated into two disparate enterprises. However, to realize the learning health system, there is a need to treat patients and evaluate therapies at the same time [65, 66]. In the future, AI-CDS systems will support patient care and support research tasks that include screening, enrollment, adaptive treatment assignment, data collection, and dynamic data analysis.

With new approaches for measuring and analyzing a wide range of biomedical data, including molecular, genomic, cellular, physiological, clinical, behavioral, and environmental data, ML models that power AI-CDS will integrate heterogeneous multimodal data to provide broader, more accurate and nuanced recommendations and predictions. As data is generated at increasing volumes and rates, adaptive AI-CDS systems will grow and continuously learn and adapt to optimize overall healthcare. Such systems will intelligently adapt to the patient (e.g., taking into account patient preferences and life circumstances), to the clinician (e.g., physician vs. nurse vs. pharmacists, etc.), to the clinical task (e.g., diagnosis, prognosis, medication reconciliation, etc.), and to the clinical context to help optimize the overall delivery of healthcare to individuals and society. Current AI-CDS systems collaborate very little, if at all, with clinician users, and as they begin to interact with thousands of users every day, human-AI cooperative systems will be increasingly developed [19].

Questions for Discussion

- What are the pros and cons of knowledge-based and data-derived AI-CDS? Discuss how to improve data-derived AI-CDS by incorporating biomedical knowledge.
- The current popular paradigm is to use big data (e.g., EHRs and billing data) to develop AI models for CDS. Describe the pitfalls of this paradigm and suggest methods to mitigate these pitfalls.
- The development of a new therapeutic (e.g., a drug or vaccine) involves rigorous assessment and validation of safety and efficacy. Do you agree that a new AI-CDS system should undergo a similar rigorous assessment and validation of safety and performance? Why or why not? How does validating an AI-CDS system differ from validating a new therapeutic? How does the nature of software complicate the application of traditional evaluation and regulation approaches?
- Hospitals typically have antimicrobial stewardship programs to monitor antibiotic prescribing and resistance patterns and to guide appropriate antimicrobial use. If you were the Chief Medical Information Officer of a large hospital that has deployed a large number of AI-CDS tools, propose the design for an AI-CDS stewardship program. What factors will you monitor and how will you accomplish doing so?

Further Reading

Greenes RA, editor. Clinical Decision Support: The Road Ahead. Elsevier; 2011 Apr 28. (Revised edition to be published in early 2023).

- This book provides a comprehensive description of the computational challenges in development of CDS systems and detailed discussions of their deployment.

Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New England Journal of Medicine. 2019 Apr 4;380 (14):1347–58.

- This review provides an overview of the uses and key challenges of machine learning for clinical applications.

Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine. 2019 Jan;25 (1):44.

- This article surveys the clinical applications of AI and deep-learning and describes their impact on clinicians, patients, and health systems.

Montani S, Striani M. Artificial intelligence in clinical decision support: A focused literature survey. Yearbook of Medical Informatics. 2019 Aug;28 (1):120.

- This survey of the literature found data-driven AI to be prevalent in CDS either used independently or in conjunction with knowledge-based AI.

Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Quality & Safety. 2019 Mar 1;28 (3):231–7.

- This article provides an overview of short-term, medium-term, and long-term safety and quality issues related to clinical deployment of AI in medicine.

References

1. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc. 2007;14(2):141–5.
2. Densen P. Challenges and opportunities facing medical education. Trans Am Clin Climatol Assoc. 2011;122:48–58.
3. Manor-Shulman O, Beyene J, Frndova H, Parshuram CS. Quantifying the volume of documented clinical information in critical illness. J Crit Care. 2008;23(2):245–50.
4. Gal DB, Han B, Longhurst C, Scheinker D, Shin AY. Quantifying electronic health record data: a potential risk for cognitive overload. Hosp Pediatr. 2021;11(2):175–8.
5. Institute of Medicine. Evidence-based medicine and the changing nature of healthcare: 2007 IOM annual meeting summary. Washington, DC: National Academies Press (US); 2008.
6. Stead WW, Searle JR, Fessler HE, Smith JW, Shortliffe EH. Biomedical informatics: changing what physicians need to know and how they learn. Acad Med. 2011;86(4):429–34.
7. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan WJ, Sinsky CA, Gilchrist VJ. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med. 2017;15(5):419–26.

8. Petersen C, Smith J, Freimuth RR, Goodman KW, Jackson GP, Kannry J, Liu H, Madhavan S, Sittig DF, Wright A. Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper. J Am Med Inform Assoc. 2021;28(4):677–84.
9. Liu X, Wang Y, Ji H, Aihara K, Chen L. Personalized characterization of diseases using sample-specific networks. Nucleic Acids Res. 2016;44(22):e164.
10. Cai C, Cooper GF, Lu KN, Ma X, Xu S, Zhao Z, Chen X, Xue Y, Lee AV, Clark N, Chen V, Lu S, Chen L, Yu L, Hochheiser HS, Jiang X, Wang QJ, Lu X. Systematic discovery of the functional impact of somatic genome alterations in individual tumors through tumor-specific causal inference. PLoS Comput Biol. 2019;15(7):e1007088.
11. Visweswaran S, Angus DC, Hsieh M, Weissfeld L, Yealy D, Cooper GF. Learning patient-specific predictive models from clinical data. J Biomed Inform. 2010;43(5):669–85.
12. Beebee H, Hitchcock C, Menzies P. The Oxford handbook of causation. Oxford University Press; 2009.
13. Pearl J. Causality. Cambridge: Cambridge University Press; 2009.
14. Spirtes P, Glymour CN, Scheines R. Causation, prediction, and search. Cambridge, MA: MIT Press; 2000.
15. Fisher RA. The design of experiments. New York, NY: Hafner; 1951.
16. Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data. Mach Learn. 1995;20(3):197–243.
17. Andrews B, Spirtes P, Cooper GF. On the completeness of causal discovery in the presence of latent confounding with tiered background knowledge. In: International conference on artificial intelligence and statistics. PMLR; 2020. p. 4002–11.
18. Rosenfeld A, Kraus S. Predicting human decision-making: from prediction to action. San Rafael, CA: Morgan & Claypool; 2018.
19. Dafoe A, Bachrach Y, Hadfield G, Horvitz E, Larson K, Graepel T. Cooperative AI: machines must learn to find common ground. Nature. 2021;593(7857):33–6.
20. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719–31.
21. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, Kim MS. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J. 2013;30(3):178–90.
22. Del Fiol G, Huser V, Strasberg HR, Maviglia SM, Curtis C, Cimino JJ. Implementations of the HL7 context-aware knowledge retrieval ("Infobutton") standard: challenges, strengths, limitations, and uptake. J Biomed Inform. 2012;45(4):726–35.
23. Gallego B, Walter SR, Day RO, Dunn AG, Sivaraman V, Shah N, Longhurst CA, Coiera E. Bringing cohort studies to the bedside: framework for a 'green button' to support clinical decision-making. J Comp Eff Res. 2015;4(3):191–7.
24. King AJ, Cooper GF, Clermont G, Hochheiser H, Hauskrecht M, Sittig DF, Visweswaran S. Using machine learning to selectively highlight patient information. J Biomed Inform. 2019;100:103327.
25. Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, Chang GH, Joshi AS, Dwyer B, Zhu S, Kaku M, Zhou Y, Alderazi YJ, Swaminathan A, Kedar S, Saint-Hilaire MH, Auerbach SH, Yuan J, Sartor EA, Au R, Kolachalama VB. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain. 2020;143(6):1920–33.
26. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, Liston DE, Low DK, Newman SF, Kim J, Lee SI. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018;2(10):749–60.
27. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. 2017;284(2):574–82.
28. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz CP, Patel BN, Yeom KW, Shpanskaya K, Blankenberg FG, Seekins J, Amrhein TJ, Mong DA, Halabi SS, Zucker EJ, Ng AY, Lungren MP. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 2018;15(11):e1002686.

29. Arevalo J, Gonzalez FA, Ramos-Pollan R, Oliveira JL, Guevara Lopez MA. Convolutional neural networks for mammography mass lesion classification. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2015; 2015. p. 797–800.
30. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.
31. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.
32. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, van der Laak J, Hulsbergen-van de Kaa C, Litjens G. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. 2020;21(2):233–41.
33. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, Bostwick DG, Evans AJ, Grignon DJ, Humphrey PA, Iczkowski KA, Kench JG, Kristiansen G, van der Kwast TH, Leite KRM, McKenney JK, Oxley J, Pan CC, Samaratunga H, Srigley JR, Takahashi H, Tsuzuki T, Varma M, Zhou M, Lindberg J, Lindskog C, Ruusuvuori P, Wählby C, Grönberg H, Rantalainen M, Egevad L, Eklund M. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. Lancet Oncol. 2020;21(2):222–32.
34. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, Olson N, Peng LH, Hipp JD, Stumpe MC. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. Arch Pathol Lab Med. 2019;143(7):859–68.
35. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully automated echocardiogram interpretation in clinical practice. Circulation. 2018;138(16):1623–35.
36. Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, Ohtsuka K, Urushibara F, Kataoka S, Ogawa Y, Maeda Y, Takeda K, Nakamura H, Ichimasa K, Kudo T, Hayashi T, Wakamura K, Ishida F, Inoue H, Itoh H, Oda M, Mori K. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med. 2018;169(6):357–66.
37. Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, Hu X, Liu P, Song Y, Zhang D, Yang X, Li L, He J, Yi X, Liu J, Liu X. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng. 2018;2(10):741–8.
38. Center for Artificial Intelligence in Medicine & Imaging. RSNA 2017: Rads who use AI will replace rads who don't. 2021. https://aimi.stanford.edu/news/rsna-2017-rads-who-use-ai-will-replace-rads-who-don-t.
39. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res. 2017;4(1):e000234.
40. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, Liu PJ, Liu X, Marcus J, Sun M, Sundberg P, Yee H, Zhang K, Zhang Y, Flores G, Duggan GE, Irvine J, Le Q, Litsch K, Mossin A, Tansuwan J, Wang D, Wexler J, Wilson J, Ludwig D, Volchenboum SL, Chou K, Pearson M, Madabushi S, Shah NH, Butte AJ, Howell MD, Cui C, Corrado GS, Dean J. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18.
41. Cai X, Perez-Concha O, Coiera E, Martin-Sanchez F, Day R, Roffe D, Gallego B. Real-time prediction of mortality, readmission, and length of stay using electronic health record data. J Am Med Inform Assoc. 2016;23(3):553–61.
42. Ng T, Chew L, Yap CW. A clinical decision support tool to predict survival in cancer patients beyond 120 days after palliative chemotherapy. J Palliat Med. 2012;15(8):863–9.
43. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, Mak RH, Aerts HJ. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiother Oncol. 2016;119(3):480–6.

44. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. J Med Syst. 1983;7(2):87–102.
45. FDA approves stroke-detecting AI software. Nat Biotechnol. 2018;36(4):290.
46. Hauskrecht M, Batal I, Hong C, Nguyen Q, Cooper GF, Visweswaran S, Clermont G. Outlier-based detection of unusual patient-management actions: an ICU study. J Biomed Inform. 2016;64:211–21.
47. Siddiqui MK, Morales-Menendez R, Huang X, Hussain N. A review of epileptic seizure detection using machine learning classifiers. Brain Inform. 2020;7(1):5.
48. Coiera E, Kocaballi B, Halamka J, Laranjo L. The digital scribe. NPJ Digit Med. 2018;1:58.
49. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med. 2018;15(11):e1002683.
50. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc Natl Acad Sci U S A. 2020;117(23):12592–4.
51. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol. 2018;154(11):1247–8.
52. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. JAMA. 2020;324(12):1212–3.
53. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.
54. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health. 2019;9(2):010318.
55. Parasuraman R, Wickens CD. Humans: still vital after all these years of automation. Hum Factors. 2008;50(3):511–20.
56. 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(1):98–107.
57. Bond RR, Novotny T, Andrsova I, Koc L, Sisakova M, Finlay D, Guldenring D, McLaughlin J, Peace A, McGilligan V, Leslie SJ, Wang H, Malik M. Automation bias in medicine: the influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. J Electrocardiol. 2018;51(6s):S6–s11.
58. Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. BMC Med Inform Decis Mak. 2017;17(1):28.
59. Golchin K, Roudsari A. Study of the effects of clinical decision support system's incorrect advice and clinical case difficulty on users' decision making accuracy. Stud Health Technol Inform. 2011;164:13–6.
60. Buchanan BG, Shortliffe EH. Rule-based expert systems: the MYCIN experiments of the Stanford heuristic programming project. Addison-Wesley; 1985.
61. Ribeiro MT, Singh S, Guestrin C. "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; 2016. p. 1135–44.
62. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Proces Syst. 2017;30:4765–74.
63. FDA. Software as a Medical Device (SAMD): Clinical Evaluation 2017. https://www.fda.gov/media/100714/download.
64. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med. 2020;3:118.
65. Angus DC. Optimizing the trade-off between learning and doing in a pandemic. JAMA. 2020;323(19):1895–6.
66. Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? JAMA. 2015;314(8):767–8.