Journal of Biomedical Informatics 116 (2021) 103723

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

A new approach to medical diagnostic decision support

Gerald E. Loeb

University of Southern California – MC 1111, 1042 Downey Way, Los Angeles, CA 90089, United States

ARTICLE INFO

Keywords:
Diagnostic tests
Bayesian
Electronic health records
Cost-effectiveness
Artificial intelligence
Learning health system

ABSTRACT

Data mining is a powerful tool to reduce costs and mitigate errors in the diagnostic analysis and repair of complex engineered system, but it has yet to be applied systematically to the most complex and socially expensive system – the human body. The currently available approaches of knowledge-based and pattern-based artificial intelligence are unsuited to the iterative and often subjective nature of clinician-patient interactions. Furthermore, current electronic health records generally have poor design and low quality for such data mining. Bayesian methods have been developed to suggest multiple possible diagnoses given a set of clinical findings, but the larger problem is advising the physician on useful next steps. A new approach based on inverting Bayesian inference allows identification of the diagnostic actions that are most likely to disambiguate a differential diagnosis at each point in a patient's work-up. This can be combined with personalized cost information to suggest a cost-effective path to the clinician. Because the software is tracking the clinician's decision-making process, it can provide salient suggestions for both diagnoses and diagnostic tests in standard, coded formats that need only to be selected. This would reduce the need to type in free text, which is prone to ambiguities, omissions and errors. As the database of high-quality records grows, the scope, utility and acceptance of the system should also grow automatically, without requiring expert updating or correction.

1. Introduction

Misdiagnosis of a patient's pathology often results in substantial harm to the health of the patient and to the finances of the healthcare system. The extent of the problem has been much studied and many factors that lead to such errors have been identified [1,2]. Many of these factors are related to the increasing complexity of clinical medicine, a trend that is inevitable, accelerating and generally desirable. Advances in physiology and molecular biology provide ever more definitive understanding of pathology and opportunities to diagnose and treat it, but at the risk of overwhelming the practitioner's ability to recall, integrate or even locate relevant information when needed. Similar challenges arise in other fields such as the maintenance of complex engineered systems, where they are increasingly addressed by technology for informatics and various forms of automated data mining and artificial intelligence (AI), e.g. [3,4]. Many ambitious attempts to apply such technology to healthcare have yet to produce comparable results [5–13].

Engineered systems have many advantages over humans for prevention and diagnosis of their faults. They automatically generate objective, quantitative data rather than responding with subjective answers to subjective questions. They can be assumed to be functionally identical when operating correctly rather than subject to many idiosyncracies that may or may not be considered normal. Their internal mechanisms are fully known and modellable rather than the subject of ongoing scientific discovery. Nevertheless, physicians are trained to follow logical and orderly procedures when arriving at diagnoses, suggesting that at least some of the process should be supportable, if not replaceable, by artificial intelligence.

Several medical diagnostic support systems have been developed over the past 30 years (e.g. DXplain, Isabel, VisualDx, Iliad, QMR, Ada DX) [14–18], but none has attained widespread implementation. All consist of stand-alone software into which an operator can enter various clinical observations and then receive a list of clinical diagnoses that might be consistent with those observations. Most use Bayesian decision-making based on hand-curated probabilities [8,19,20], but none has been integrated into the electronic health records (EHRs) of individual patients. In theory, such integration would allow the Bayesian probabilities to be calculated and updated based on accumulating clinical experience, as is done for machine maintenance, but current EHRs are not structured for this. Some medical support systems employ decision trees developed by experts to chart a path through additional test data that should be obtained, but the paths are fixed, depriving the physician of the many subjective judgments and patient specific cost and benefit considerations that are required when dealing

E-mail address: gloeb@usc.edu.

https://doi.org/10.1016/j.jbi.2021.103723
Received 27 November 2020; Received in revised form 27 January 2021; Accepted 19 February 2021
Available online 9 March 2021

with humans rather than machines.

An ambitious concept of self-learning Knowledge Nets derived from clinical records and journal articles was proposed by Lawrence Weed [21–23], developer of the widely accepted problem-oriented medical record. A simplified version (Isabel Healthcare Inc.) based on selected textbooks and journals was able to include the correct diagnosis in a list of 30 possibilities for 96% of cases in which key findings were selected and entered by hand but in only 74% when given the entire EHR [24]. A general knowledge-based implementation by IBM Watson Health underperformed expectations because it underestimated the complexity of such knowledge [11].

The two general applications of artificial intelligence (AI) to this problem can be summarized as pattern-based and knowledge-based. Both aim to take advantage of large electronic databases of clinical experience such as EHRs. Each has been applied with some success to circumscribed diagnostic applications in which previously gathered information about a patient is used to identify candidate diagnoses from a circumscribed set of possibilities, as illustrated in Fig. 1. This article describes a new algorithm to address the larger requirement of working iteratively with the physician to suggest cost-effective next steps in the diagnostic process and to integrate new clinical data automatically as it becomes available. It builds on an iterative decision-making tool that achieved success exploring and identifying unknown haptic objects [25,26], an application that is qualitatively similar to differential diagnosis, albeit much simpler in scope. Methods are described to add the additional capabilities required for clinical diagnostic support.

2. Requirements

Defining the requirements of all stakeholders is now a required first step in the design of all Class II and III medical devices and software [27]. Stakeholders for medical diagnostic software include patients, clinicians, clerks, institutional purchasers, programmers, manufacturers, regulators and payors. Diagnostic decision-support software constitutes a very large and complex medical product that can succeed only if it meets their requirements.

2.1. The clinician's interface must be convenient and easy to use.

Many diagnostic decisions are made by primary care physicians during visits that are under economic pressure to be as brief as possible. Those caregivers are already resentful of time spent with EHR software instead of the patient because such software is currently designed and required for billing payors rather than to add value for the patient. Nevertheless, the EHR is an essential component of any automated learning health system, so decision-support should be provided and closely integrated within such software rather than diverting the clinician to a separate interface or external resources [9]. If such decision-support came to be perceived as valuable to the clinician's diagnostic process by saving time or preventing oversights, clinicians would be more enthusiastic about its procurement and use.

2.2. The software should provide support for clinician's memory and judgment.

Informatics engineers aspire to create complete solutions that supersede human judgment and prevent human errors. Clinical interactions with patients, however, require the types of subjective and affective judgment that defy recording in the EHR, much less quantification and analysis by AI. Any decision-support system should provide only salient, objective information that the clinician may have overlooked and do so in a way that is as unobtrusive as possible. The physician should be able to ignore or override irrelevant or inaccurate advice, which is inevitable in such automated systems.

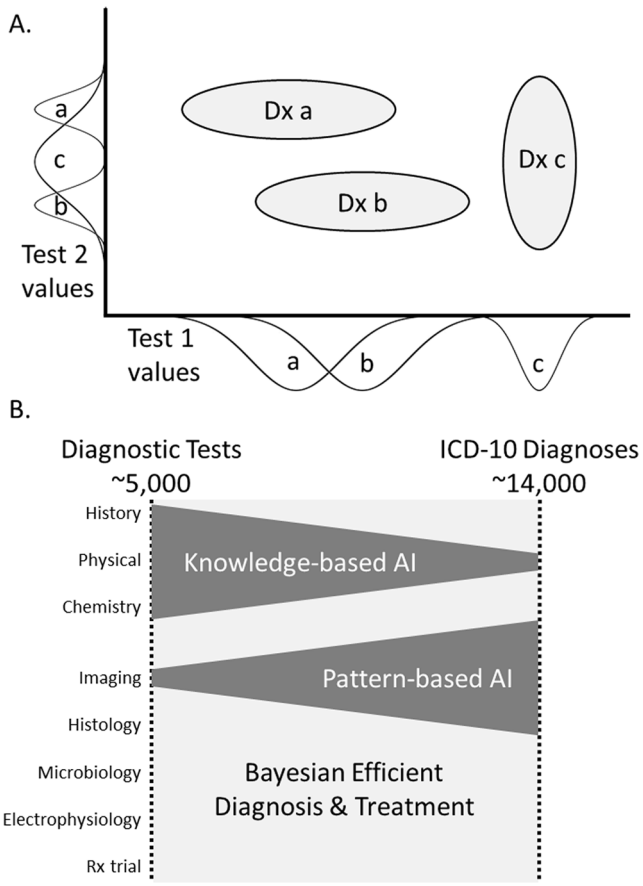

Fig. 1. A. Differential diagnosis is based on identifying information from clinical data (herein referred to as tests) that might distinguish among possible diagnoses a, b, c…. A simplistic starting point is to assume that patients in different conditions will have different distributions of certain test results (values) from each other and from nominal normal values. B. Knowledge-based procedures such as checklists, flowcharts and decision-trees have been constructed by panels of experts to standardize screening for a limited subset of common diagnoses using commonly available diagnostic test data. Pattern-based artificial intelligence (AI) has been used on specialized test data such as radiological and histological images to identify a broader subset of diagnoses for which such tests are already believed to be discriminative. Current AI methods require massive amounts of training data that represents adequately all possible combinations of test data and diagnoses. This is impractical for the complete and growing ~5,000 dimensional problem of all clinical tests available (but obtained inconsistently) to discriminate among ~14,000 diagnoses. In the language of informatics, this results in a high dimensional hyperspace that will inevitably be sparsely populated, resulting in "the curse of dimensionality" [40] when attempting to apply pattern-based AI (https://deepai.org/machine-learning-glossary-and-terms/curse-of-dimensionality#). Various cures for this curse have been proposed but they require special conditions that are not relevant to the general problem of clinical diagnosis [41,42].

2.3. The software should avoid provoking liability claims.

Knowledge-based protocols, checklists, standard operating procedures and other proscriptive "recommendations" for the diagnostic work-up of patients invite liability. When diagnostic errors inevitably occur, the experts who generated the standards and the supplier that incorporated them may be liable for a defective product. If its recommendations result in excessive costs, insurers may refuse to reimburse them. If its recommendations are ignored by clinicians because they are deemed to be erroneous or excessive, this may constitute evidence in a malpractice lawsuit. Efforts to mitigate such problems require frequent chart reviews and case studies to identify them, which is time-consuming and may be threatening to institutions and clinicians. The

information provided to the clinician should be helpful rather than proscriptive and should derive from sources and processes that pose no liability.

A determination of negligence is usually based on prevailing standards of care, which are difficult to establish for a rapidly evolving technology such as diagnostic support software. Responsibility for a diagnostic misadventure might result from blind adherence to its inappropriate recommendations or from failure to employ technology that would have made an appropriate recommendation [28].

2.4. *The software should encourage efficiency.*

The cost of healthcare has been rising unsustainably for two reasons that are inextricably linked: increasingly powerful and expensive diagnostic and therapeutic tools (a good thing) that inevitably tax the clinician's ability to remember them and to understand their benefit/cost (a bad thing). The true cost of any diagnostic test must include its nominal charge, delays associated with obtaining results and any risks of adverse events from the test itself (e.g. radiation, embolization, allergic reaction, etc.). The costs of delay and adverse events are patient-specific, as is the benefit of a specific test toward establishing a definitive diagnosis at any given time in any given patient.

2.5. *Diagnostic advice should update automatically as standards of practice evolve.*

The decision-support software should recognize when new items appear in the EHR database, monitor their utility based on collective clinical experience and provide the salient items to the clinician during the decision-support process. The growth of diagnoses, diagnostic tests and therapeutics is accelerating and best methods of practice are constantly evolving. Research to extract and obtain consensus on such methods is expensive, slow and generally limited to a few high-profile clinical challenges. Extracting such information from journal articles via natural language processing (e.g. IBM-Watson) and inductive learning [29] remains challenging [11], especially for the complex sentence structures often found in scientific writing [30].

2.6. *The software should support iterative decision-making.*

Clinical diagnosis is inherently iterative, unlike extracting a pattern from a complete image. Given only a chief complaint and essential demographics, the busy clinician is already prioritizing the order of questions and observations during the history and physical according to a probability-weighted differential diagnosis [31]. This is generally not quantified, it is only partially recorded and it may be almost subconscious, but it is an essential process for efficient diagnosis and treatment. It is anathema to artificial intelligence based on pattern recognition by deep-learning neural networks, which require reasonably complete and well-curated EHRs.

2.7. *The software should propose treatment options and consider responses.*

The decision-support system should have knowledge of treatments and provide advice based on responses to treatment regardless of when or where they occurred. Failure to arrive at a definitive diagnosis before initiating a treatment is not a diagnostic error. Often the most cost-effective way to rule out a common cause of a problem is to see if the patient responds to a course of its common treatment. Diagnostic challenges such as unusual presentations and rare diseases are usually entertained only *if* such treatment fails, but unfortunately not necessarily *when* [32].

3. Design

The description below identifies the structure and major components of a decision-support system called Bayesian Efficient Diagnosis & Treatment (BEDT). It combines classical Bayesian inference with an inversion of Bayes' theorem called Bayesian Exploration (Fig. 2). Its success will depend on buy-in and substantial investment from many stakeholders both in the clinic and in the larger healthcare system. Those stakeholders will need to refine the current design before implementation. The conceptual design described here can be used to identify specific challenges that are amenable to existing technology or that require new technology. When stakeholders are satisfied with the design and engineers have determined that the required technology is available or feasible, it then becomes possible to create the partnerships, budgets and schedules required to implement a realistic plan.

In the descriptions below, the term "test" is used to include the collection of any piece of clinically useful information: basic demographics, items on an intake questionnaire, clinical history questions, physical exam observations, laboratory tests, therapeutic trials, etc. A test returns a value, which may be a continuous variable, a yes/no answer or a probability of an observation (e.g. "indicative of" or "suggestive of" or "consistent with" a diagnosis or other conclusion). The terms "benefit" and "cost" are also formally defined below.

3.1. *Bayesian exploration*

Bayesian exploration is a new form of AI that appears to overcome the limitations of pattern-based and knowledge-based AI (Fig. 1). Bayesian inference has long been used to determine the probability of a particular event such as a diagnosis given an evolving set of observations (tests) [19,29,33]. Bayesian exploration inverts Bayesian inference to decide which next observation (test) will best disambiguate the currently probable events [25,26].

The anticipatable benefit of any diagnostic action depends on which diagnoses are currently being entertained in a specific patient. For Bayesian exploration, the current EHR database is regularly mined (which can be done offline) to produce confusion matrices (Fig. 3 top) whose values reflect the overlap in probability density functions of the results reported for each test in patients with each diagnosis (Bhattacharyya coefficient, most easily computed for tests that tend to produce normal distributions of values but can be extended to skewed, Boolean and other data types). The benefit of using a given test at any given point in a differential diagnosis can be estimated by weighting the confusion matrix according to the current (Bayesian prior) probabilities of each diagnosis in the current patient (Fig. 3 bottom). The sum of the weighted cells in the confusion matrix reflects the likely confusion that will remain after performing the test with this patient, so the test with the lowest sum is the statistically optimal choice with the largest anticipated benefit for the diagnostic process.

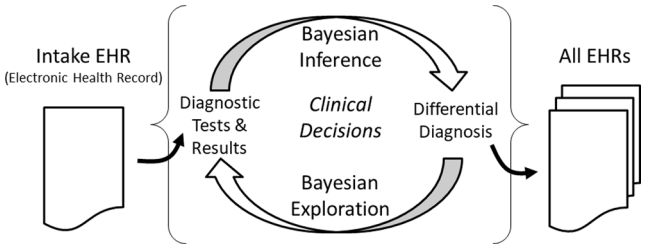

Fig. 2. Bayesian Efficient Diagnosis & Treatment consists of two steps that form an iterative cycle: forward Bayesian inference uses currently available diagnostic data in a patient's EHR to identify the probabilities of items in their differential diagnosis; inverse Bayesian exploration uses the collective clinical experience in all EHRs to prioritize next items of diagnostic data according to their ability to disambiguate the differential diagnosis.

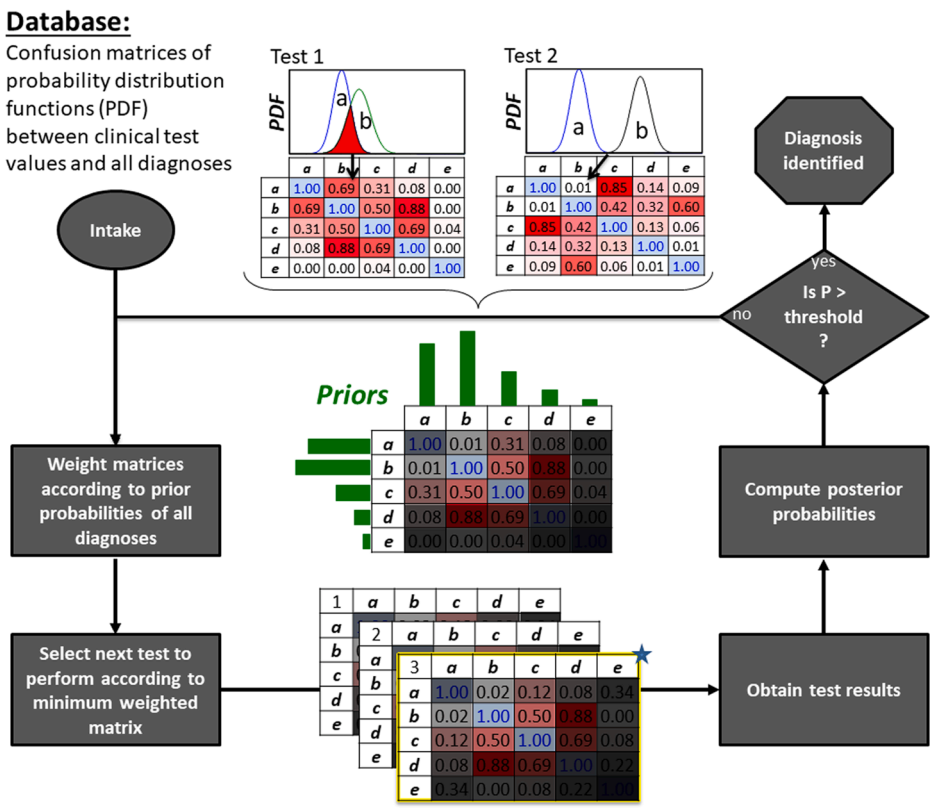

Fig. 3. Bayesian Exploration is a fully automatable and iterative process that starts with the initially available clinical data to identify the currently most probable diagnoses (Priors) according to commonly used Bayesian inference. A clinical Database drawn from the collective EHRs of a large, diverse population of patients quantifies the degree to which any given clinical test is likely to distinguish among diagnoses a, b,…, by trying to minimize overlap (red) between the distributions of test values. Each quantified overlap constitutes an entry in the confusion matrix for each test. During use with a patient, the values in the confusion matrix for each test are multiplied by the Prior probabilities of each diagnosis (green bars, resulting in weight depicted inversely as shading for each cell) and summed (omitting the blue diagonal cells that are necessarily 1.0). Tests with the lowest sum are most likely to be useful and may be selected by the clinician. When new results are available, Bayesian inference computes the new (posterior) probabilities of each diagnosis. If one diagnosis has a high enough probability, the process can be stopped; otherwise, these probabilities become the priors for the next iteration of the differential diagnostic process. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Like both knowledge-based and pattern-based AI, Bayesian exploration automatically abstracts information from large databases but without the expert curation required of current Learning Health Systems [12]. Like pattern-based AI, it extracts information autonomously from whatever events have been recorded in its database, but it copes much better with inherently high-dimensional and sparse data such as EHRs. This is because, like humans, it solves problems in steps rather than looking for a gestalt in one pass [25]. Like knowledge-based AI, Bayesian exploration informs iterative processes such as differential diagnosis, but it does not need to be taught "good" vs. "bad" procedure, about which it is agnostic. Unlike the extensive off-line training required by both pattern-based and knowledge-based AI, Bayesian exploration mines the EHR database at time of use, informing the clinician about the diagnoses that were eventually confirmed in similar patients and the diagnostic tests that other clinicians have employed. If new diagnostic tests (i.e. dimensions) or diagnoses appear in its continuously growing database, they will be considered and presented automatically.

Bayesian exploration is not a form of machine learning. No matter what data it is fed, its underlying algorithm is deterministic, transparent and unchanging, unlike deep-learning neural networks [34]. It claims no expertise and provides no information that the clinician couldn't have extracted by (very) extensive review of the EHR database. The advice provided to the clinician reflects the standard of practice represented in the cumulative actions of all clinicians in the database rather than contentious or changing opinions of experts.

3.2. Cost estimation

The diagnostic action that promises the largest benefit isn't necessarily the only or even best course of action. The clinician needs suggestions that are practical because they consider costs as well as benefits. This requires objective information regarding availability, charges, and delays associated with each test procedure, as well as subjective considerations that are patient-specific such as comorbidities, reliability, mobility and ability to pay. The objective information can be used to compute a benefit/cost ratio so that the diagnostic options most likely to be cost-effective are at the top of a rank-ordered list (Fig. 4). That list should be presented alongside a rank-ordered list of all diagnoses over a threshold level of probability so that the clinician can understand the utility of the options for diagnoses that he/she might have overlooked (see below). In the end, however, the clinician must evaluate subjective factors to make appropriate selections, which the software can then automatically record and implement efficiently as coded orders. The machinations of the BEDT algorithms in computing those lists can and should be hidden.

The different aspects of cost (financial, risk and delay) need to be combined into a single cost estimate in order to compute the benefit/cost ratio. Financial cost tends to be locally variable and may already be available in the billing systems associated with EHRs. Risk cost consists of the historical expenses associated with each adverse event associated with a test multiplied by the probability of its occurrence, which is often available from the literature. Delays tend to be highly local; the cost of any anticipated delay depends on the probability that the patient has a condition for which such a delay in diagnosis and treatment incurs costs associated with increased mortality or morbidity. Fig. 5 illustrates how the iterative nature of Bayesian exploration can generate such an estimate by extrapolating the current situation according to a series of "if-then" scenarios (akin to the stochastic mathematical analysis called Markov chains).

3.3. Clinician's interface

The current user interfaces for EHR databases are generally unpopular with clinicians. They include hierarchical pull-down menus with pick-lists of coded diagnoses and test procedures. Many clinicians ignore these in favor of entering free text that is often colloquial shorthand, perhaps misspelled, and almost useless for automatic billing or data mining. Human clerks then comb through the clinician's notes and

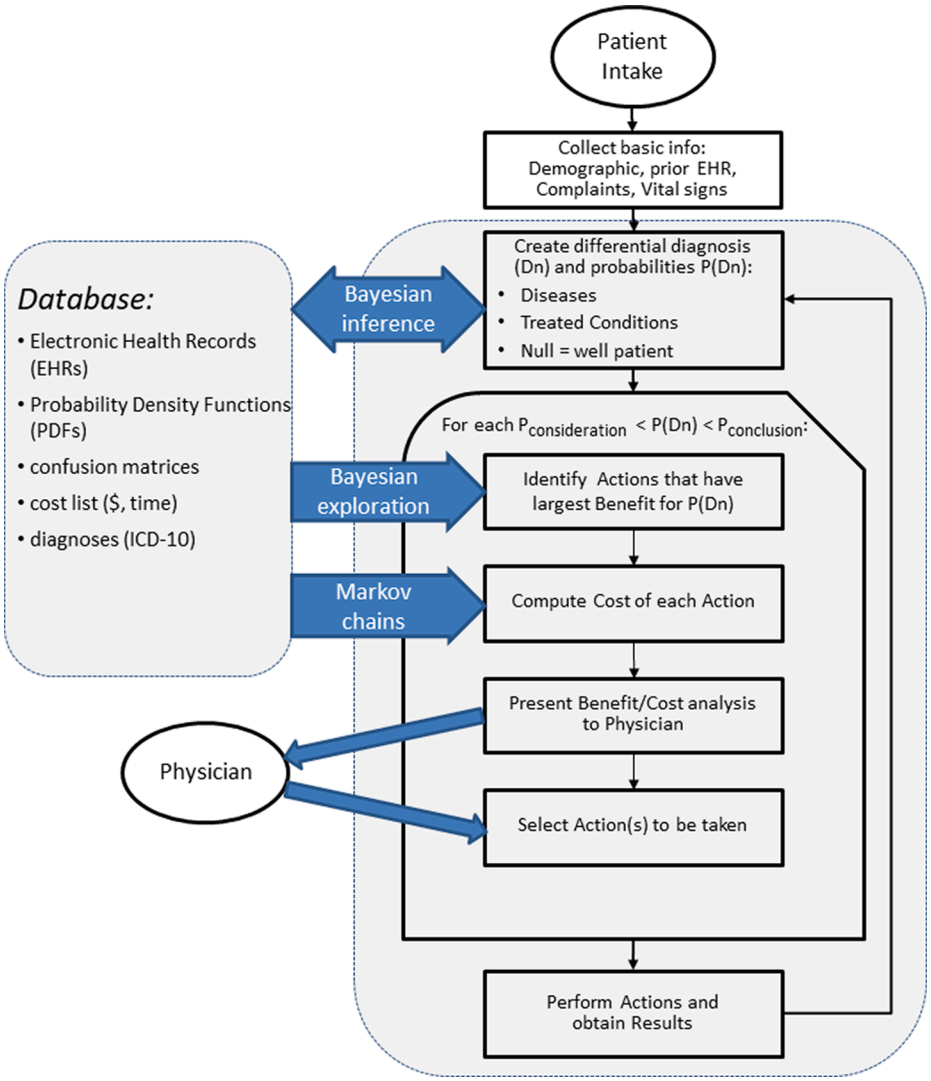

Fig. 4. A flowchart for the interaction of the BEDT software with physician and patient. The intake information from the patient plus any preexisting EHR provides input to Bayesian inference to create an initial differential diagnosis (Dn) from the most probable diagnoses (P(Dn)) in the database of all EHRs, delimited by a threshold probability (Pconsideration) and rank-ordered according to descending probability. The Bayesian exploration algorithm identifies the Actions that are likely to provide the greatest diagnostic benefit and extrapolates their Costs via Markov chains (see Fig. 4) to determine a Benefit/Cost ratio that is used to rank-order the Actions for the physician to select as desired. When the Results of the Actions are available, Bayesian inference is repeated to recompute Dn and P(Dn) for presentation to the physician.

orders to extract billing codes [35]. Informatics specialists painstakingly curate databases so they can be used for chart-based research. BEDT follows and supports the thinking process of the clinician, so the most relevant observations, diagnoses and tests can be offered to the clinician (complete with standard nomenclature and codes) for one-click selection in the user interface, eliminating deeply nested pick-lists or manual entry of free text and post-processing.

The highly simplified example illustrated in Fig. 6 assumes that there are only three diagnoses and four diagnostic procedures in the database. Based on the basic intake data (presenting complaint, age, sex etc.), the BEDT software presents the diagnoses rank-ordered according to descending probability and the diagnostic tests according to decreasing benefit/cost. Based on subjective impressions regarding the patient's condition, the clinician might tell the patient to try aspirin while waiting for an MRI test, which was automatically scheduled when selected. When the patient returns in two weeks, the MRI results provide a definitive diagnosis and a clear indication of the most useful next action. Again, the clinician needs only to select the now-obvious choice rather than type more text.

4. Path to implementation

4.1. Quality of EHRs

This is a chicken-and-egg problem for BEDT. Its probabilistic nature makes it relatively resistant to errors and omissions, but the current state of most uncurated EHR databases will be a serious obstacle for BEDT or any learning health system [13]. One possibility is to start with a nucleus of well-curated, albeit limited, EHRs in a database sufficient to provide useful information for a limited range of diagnoses and tests. Clinicians would then need to type in many items that appear to be missing, which is essentially what many are now doing all the time. Because BEDT updates its knowledge every time it queries the database, the new items will be available immediately for selection by other clinicians; their diagnostic utility will become apparent from the collective experiences of their patients. Possible sources for such a database nucleus include pediatric hospital admissions, which often arise from circumscribed and eventually well-diagnosed causes, or rare disease consortiums [32], which deal with problems that are often misdiagnosed at great expense and for which unfamiliar new diagnoses and specific test methods are now appearing rapidly.

Like any automated decision support system, BEDT assumes that all or at least most of the relevant clinical information about a patient's history is in that patient's EHR. One challenge is to provide a clinician's interface that efficiently captures important observations that do not lend themselves to continuous variables, such as types of pain or the texture of skin. This would be facilitated by the ability of BEDT to prioritize the display of the most likely and important observations in precoded form. Another challenge is the scattering of health records in the balkanized US healthcare system, where a given patient may have

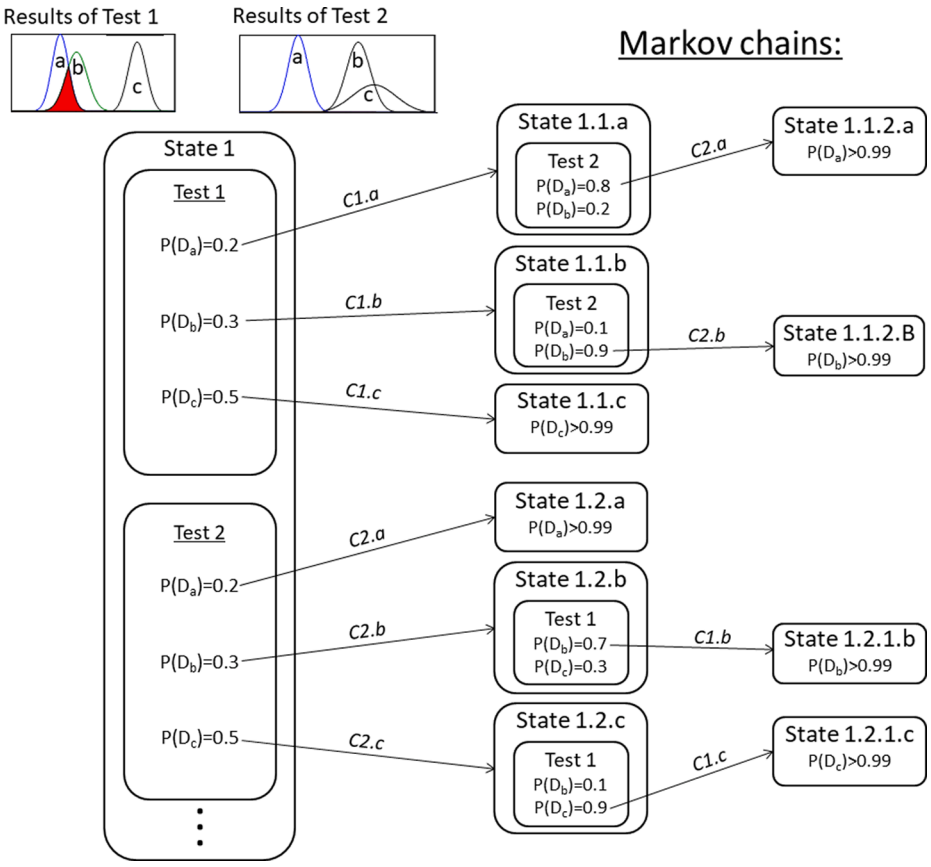

Fig. 5. State 1 illustrates a point in the diagnostic work-up where three diagnoses are being considered (Da, Db, Dc) with initial Bayesian probabilities P(Da), P(Db) and P(Dc). Two further diagnostic tests are under consideration, with PDFs for each diagnosis illustrated at the top. If the patient actually has Da (whose probability is only 0.2) or Db (whose probability is 0.3), the most likely outcome of Test 1 (the mean value of its PDF for Da) will still be ambiguous regarding Db, so this path will entail both the probability-weighted cost C1.a and the further cost C2.a of conducting Test 2 to arrive at a final diagnosis. In this case, it would have been better to start with Test 2, incurring only the cost C2.a. If the patient actually has diagnosis Dc (the highest probability at the starting point), then it is less costly to start with Test 1. The probability-weighted sums of the costs of starting with Test 1 or Test 2 are divided into the diagnostic Benefit of each Test as computed by Bayesian exploration (Fig. 3) to generate the Benefit/Cost value for each Test that is presented to the clinician.

INTAKE VISIT

Pt. #12345 Name: John Doe Sex: M DOB: 01/01/1970 Zipcode: 11111

Presenting complaint: headache for 3 days

Diagnoses (code) % probability to consider:	Diagnostic tests (code) benefit/cost to consider:
❑ *Tension headache (D001) 90%*	❑ *Aspirin trial (T001) 6.0*
❑ *Viral encephalitis (D002) 7%*	❑ *Spinal tap (T002) 5.4*
❑ *Meningioma (D003) 3%*	❑ *MRI (T003) 3.1*
	❑ *Craniotomy (T004) 0.1*

FOLLOW-UP VISIT

Headaches persist despite aspirin trial for two weeks

Test results

MRI (T003): 4 cm diameter, well-circumscribed anterior fossa tumor

Diagnoses (code) % probability to consider:	Diagnostic tests (code) benefit/cost to consider:
✓ *Meningioma (D003) 99%*	❑ *Craniotomy (T004) 9.9*
❑ *Tension headache (D001) <1%*	❑ *Aspirin trial (T001) 0.3*
❑ *Viral encephalitis (D002) <1%*	❑ *Spinal tap (T002) 0.0*
	❑ *MRI (T003) 0.0*

Fig. 6. A highly simplified example to illustrate the general nature of the clinician's interface with BEDT. At each stage of the work-up, the physician sees a list of the current items in the differential diagnosis with their diagnostic codes, rank-ordered according to current probability. BEDT presents another coded list of the pertinent diagnostic tests with codes, rank-ordered according to their benefit/cost value at this stage of the work-up. After selecting the desired tests and obtaining their results at a follow-up visit, the differential diagnosis has been reordered, as is the list of actions and their new benefit/cost. Note that the option "Craniotomy" might be diagnostic or therapeutic. If the list of diagnoses includes "headaches resolved", this is now the procedure most likely and cost-effective to produce that outcome, but the low initial probability of meningioma placed it at the bottom of the initial list of diagnostic tests.

key parts of their medical history distributed over multiple doctors' offices, out-patient clinics, emergent care facilities and hospitals, many of which use mutually incompatible EHRs and databases or undigitized records. This suggests that BEDT should be developed and tested initially in a walled-garden such as an HMO or in a country with a national healthcare system and comprehensive, centralized EHRs. Success there might motivate the myriad policy and information technology changes required for data sharing in less centralized systems.

A key piece of information for BEDT is a correct final diagnosis for each patient's condition(s). It doesn't matter how long or tortuous the path, but such a diagnosis needs to be in the EHR to be able to quantify the utility (or lack) of any steps along the way. This requirement is particularly problematic for unusual presentations and rare diseases that are often not diagnosed until the patient is referred out of their local healthcare system or comes to autopsy. The declining rate of autopsies [36] and the isolation of such records from the clinical EHRs will need to be addressed for BEDT or any other AI approach to diagnosis to perform well.

4.2. Temporal property of clinical data

Diseases evolve and their clinical manifestations change in their nature and significance over time. The EHR generally includes timestamps on patient visits and clinicians usually note the duration of symptoms and treatments, so temporal information is available, but it cannot be used as a continuous variable without causing the dimensionality of the problem to explode. One possibility is to quantize it into bins that reflect the different pathophysiological processes that are responsible for such evolution. For example, a sore throat or chest pain that has been present for a month is likely to arise from causes that are different from those likely to be responsible after a week, a day or an hour. A small number of such crude (essentially logarithmic) bins can be used to subdivide general observations into discrete but useful temporal categories.

4.3. Diagnostic classification systems

The ICD system (International Classification of Disease) started as an epidemiological tool and evolved to facilitate billing and reimbursement rather than differential diagnosis. A great deal of work has gone into this but it may not be fit for the purpose of improving healthcare [37]. Clinicians and clerks tend to use different codes for a singular diagnosis and may combine codes inconsistently to reflect predisposing or complicating conditions, leading to calls for simplification [38,39]. This challenge needs to be addressed as part of any automated effort to provide broad support for diagnosis and treatment. It will be difficult to use existing EHRs based on ICD-10 codes for a demonstration of feasibility without requiring so much curation that it vitiates relevance to current clinical practice. If a BEDT system is built on a new coding system, it may not be acceptable in a healthcare system based on ICD-10 codes for reimbursement.

Many patients have more than one concurrent disorder. If the pathophysiological effects of each interacted linearly in the probability density functions of results from each test, the effects of one confirmed diagnosis could be used to normalize results in order to discriminate a secondary disorder. Unfortunately, these interactions are often complexly and nonlinearly synergistic in terms of presentation, diagnostic test results and response to treatment. One approach is to create new diagnostic codes for syndromes and complications, which is already being done in the ICD system to determine appropriate reimbursement levels. This remains a work in progress for which BEDT could provide a useful perspective, but meanwhile it is another argument for the importance of providing easily over-ridden diagnostic support rather than actual diagnoses.

4.4. Feasible extensions

The treatment part of BEDT is easily extended beyond therapeutic trials for diagnosis. If the diagnoses list itself contains items identifying outcomes such as "pneumonia resolved", "cancer in remission" and "well-patient", BEDT could be used to suggest treatments that might produce such outcomes and that were rank-ordered according to benefit/cost to do so. If the treatments list were further extended to include lifestyle advice such as "reduce caloric intake", "get more exercise" and "stop smoking", BEDT could remind physicians to discuss such preventive medicine with patients when most likely to be effective in achieving well-patient status. As genetic data become more common in EHRs, BEDT will automatically detect if single nucleotide polymorphisms are correlated with increased probability of certain diagnoses or different responses to treatment, a discovery that now usually requires large, expensive and narrowly focused clinical studies.

The norm-based strategy of BEDT makes it particularly suited to looking for systematic deviations from practice norms that might indicate various failings and abuses of public health. Higher than expected incidence of a diagnosis in a geographical locale could alert public health officials to look for an outbreak of an infectious disease or an environmental contaminant. Higher than expected use of expensive treatments with low benefit/cost might trigger accountable care organizations (ACOs) to do in-service training or insurers to look for fraud.

5. Where to Start?

Close integration of BEDT with the EHR and database software is essential for its computational function, for the usability of the clinician's interface and to achieve the virtuous cycle anticipated from eliminating much of the free text that now defeats the original goal of EHRs. This suggests that the project should be led by a software developer. If the project were to be funded as a profit-making enterprise, the improved usability, reduced clerical effort and reduced diagnostic oversights would all be marketable features to institutional purchasers. Providing decision-support to clinicians that is extracted from their own collective clinical records should minimize the regulatory burden associated with "Software as a Medical Device" (https://www.fda.gov/media/100714/download) and will probably not fall under the new "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" (https://www.fda.gov/media/145022/download). The completely anonymized probabilities and benefit/cost estimates should avoid patient privacy and data security issues.

The radically different purpose and design of BEDT from current EHR software may be unattractive to purveyors of existing products. EHR software has become a mature industry, despite its clinically dysfunctional nature. It has locked up customers with many legacy systems for accounting and reimbursement. BEDT will require close integration with the business plans of healthcare providers and complex reimbursement systems, which are particularly balkanized in the US. Rethinking diagnostic classification, quantifying diverse data types and sharing EHRs across platforms are massive challenges faced by any automated decision support system for general medical practice. This suggests that development of BEDT might need to be led by a large new player looking to take advantage of an extremely large and inefficient marketplace with many dissatisfied stakeholders. Fortunately, we live in an age when large-scale disruptive innovation of established industries has become feasible and investable once it looks technologically possible.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding acknowledgement

Development of the methods described herein was supported in part by SynTouch, Inc. and Biomed Concepts, Inc., which had no role in the preparation of this manuscript.

References

[1] National Academies of Sciences and Medicine. Improving Diagnosis in Health Care, National Academies Press, 2015.
[2] H. Singh, G.D. Schiff, M.L. Graber, I. Onakpoya, M.J. Thompson, The global burden of diagnostic errors in primary care, BMJ Qual. Safety 26 (6) (2017) 484–494.
[3] S.N. Huang, K.K. Tan, T.H. Lee, Automated fault detection and diagnosis in mechanical systems, IEEE Trans. Syst., Man Cybernet., Part C (Applications and Reviews) 37(6), pp. 1360–1364, 2007, doi: 10.1109/tsmcc.2007.900623.
[4] J. Zaytoon, S. Lafortune, Overview of fault diagnosis methods for Discrete Event Systems, Ann. Rev. Control 37 (2) (2013) 308–320, https://doi.org/10.1016/j.arcontrol.2013.09.009.
[5] M.L. Graber, Progress understanding diagnosis and diagnostic errors: thoughts at year 10, Diagnosis 7 (3) (2020) 151–159, https://doi.org/10.1515/dx-2020-0055.
[6] D. Liebovitz, Next steps for electronic health records to improve the diagnostic process, Diagnosis 2 (2) (2015) 111–116, https://doi.org/10.1515/dx-2014-0070.
[7] M.L. Graber, C. Byrne, D. Johnston, The impact of electronic health records on diagnosis, Diagnosis 4 (4) (2017) 211–223, https://doi.org/10.1515/dx-2017-0012.
[8] N. Riches, et al., The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis, PLoS ONE 11 (3) (2016), e0148991, https://doi.org/10.1371/journal.pone.0148991.
[9] G.D. Schiff, D.W. Bates, P. Hartzband, J. Groopman, G. Schiff, Can electronic clinical documentation help prevent diagnostic errors? New Engl. J. Med. 362 (12) (2010) 1066.
[10] R. El-Kareh, O. Hasan, G.D. Schiff, Use of health information technology to reduce diagnostic errors, pp. 2, pp. ii40-ii51, BMJ Qual. Saf. 22 (Suppl) (2013), https://doi.org/10.1136/bmjqs-2013-001884.
[11] E. Strickland, IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care, IEEE Spectr. 56 (4) (2019) 24–31, https://doi.org/10.1109/mspec.2019.8678513.
[12] D. Corrigan, et al., Requirements and validation of a prototype learning health system for clinical diagnosis, Learn. Health. Syst. 1 (4) (2017), e10026, https://doi.org/10.1002/lrh2.10026.
[13] S. McLachlan, et al., A framework for analysing learning health systems: Are we removing the most impactful barriers? Learn. Health. Syst. 3 (4) (2019), e10189 https://doi.org/10.1002/lrh2.10189.
[14] E. Vardell, C. Bou-Crick, VisualDx: A visual diagnostic decision support tool, Med. Reference Services Quart. 31 (4) (2012) 414–424, https://doi.org/10.1080/02763869.2012.724287.
[15] S. London, DXplainTM: A web-based diagnostic decision support system for medical students, Medical Reference Services Quart. 17 (2) (1998) 17–28, https://doi.org/10.1300/j115v17n02_02.
[16] E. Vardell, M. Moore, Isabel, a clinical decision support system, Medical Reference Services Quart. 30 (2) (2011) 158–166, https://doi.org/10.1080/02763869.2011.562800.
[17] J.B. Lemaire, J.P. Schaefer, L.A. Martin, P. Faris, M.D. Ainslie, R.D. Hull, Effectiveness of the Quick Medical Reference as a diagnostic tool, CMAJ 161 (6) (1999) 725–728.
[18] H.R. Warner, et al., ILIAD as an expert consultant to teach differential diagnosis, in: Proceedings of the Annual Symposium on Computer Application in Medical Care, American Medical Informatics Association, 1988, p. 371.
[19] P. Elkin, D. Schlegel, M. Anderson, J. Komm, G. Ficheur, L. Bisson, Artificial intelligence: Bayesian versus heuristic method for diagnostic decision support, Appl. Clin. Inform. 09 (02) (2018) 432–439, https://doi.org/10.1055/s-0038-1656547.
[20] P.L. Elkin, G.O. Barnett, K.T. Famiglietti, R.J. Kim, Closing the Loop on Diagnostic Decision Support Systems, in: Proc Annu Symp Comput Appl Med Care vol. 11, 1990, pp. 589–593.
[21] L.L. Weed, Knowledge coupling, medical education and patient care, Crit. Rev. Med. Inform. 1 (1) (1986) 55–79.
[22] L.L. Weed, Medical records that guide and teach, Clin. Problem Lists Electronic Health Record, vol. 19, 2014.
[23] A. Shukor, An alternative paradigm for evidence-based medicine: revisiting Lawrence weed, MD's systems approach, Permanente J. (2017), https://doi.org/10.7812/tpp/16-147.
[24] M.L. Graber, A. Mathew, Performance of a web-based clinical diagnosis support system for internists, J. Gen. Intern. Med. 23 (1) (2008) 37–40.
[25] G.E. Loeb, J.A. Fishel, Bayesian action & perception: representing the world in the brain, Front. Neurosci. 8 (2014).
[26] J.A. Fishel, G.E. Loeb, Bayesian exploration for intelligent identification of textures, Front. Neurorobotics 6 (4) (2012), https://doi.org/10.3389/fnbot.2012.00004.
[27] G.E. Loeb, F.J.R. Richmond, Making Design Controls Useful for R&D, Med. Dev. Diag Ind. 25 (4) (2003) 63–68.
[28] K.W. Goodman, Ethical and Legal Issues in Decision Support, Springer International Publishing, 2016, pp. 131–146.
[29] I. Kononenko, Inductive and Bayesian Learning in Medical Diagnosis, Appl. Artificial Intell. 7 (4) (1993) 317–337, https://doi.org/10.1080/08839519308949993.
[30] S. Läubli, S. Castilho, G. Neubig, R. Sennrich, Q. Shen, A. Toral, A set of recommendations for assessing human-machine parity in language translation, J. Artificial Intell. Res. 67 (2020), https://doi.org/10.1613/jair.1.11371.
[31] J. Groopman, How doctors think, Houghton Mifflin Harcourt, 2008.
[32] S. Ronicke, M.C. Hirsch, E. Türk, K. Larionov, D. Tientcheu, A.D. Wagner, Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study, Orphanet J. Rare Dis. 14 (1) (2019) 69.
[33] M. Bayes, M. Price, An essay towards solving a problem in the doctrine of chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S., Philos. Trans. 53 (1763) 370-418, doi: 10.2307/105741.
[34] D. Castelvecchi, Can we open the black box of AI? Nat. News 538 (7623) (2016) 20.
[35] G.D. Schiff, L. Zucker, Medical scribes: salvation for primary care or workaround for poor EMR usability? J. Gen. Intern. Med. 31 (9) (2016) 979–981, https://doi.org/10.1007/s11606-016-3788-x.
[36] K.G. Shojania, E.C. Burton, K.M. McDonald, L. Goldman, Changes in rates of autopsy-detected diagnostic errors over time, JAMA 289 (21) (2003) 2849–2856, https://doi.org/10.1001/jama.289.21.2849.
[37] G. Surjan, Questions on validity of International Classification of Diseases-coded diagnoses, Int. J. Med. Inf. 54 (2) (1999) 77–95, https://doi.org/10.1016/s1386-5056(98)00171-3.
[38] J. Stausberg, N. Lehmann, D. Kaczmarek, M. Stein, Reliability of diagnoses coding with ICD-10, Int. J. Med. Inf. 77 (1) (2008) 50–57, https://doi.org/10.1016/j.ijmedinf.2006.11.005.
[39] R. Wockenfuss, T. Frese, K. Herrmann, M. Claussnitzer, H. Sandholzer, Three- and four-digit ICD-10 is not a reliable classification system in primary care, Scand. J. Prim. Health Care 27 (3) (2009) 131–136, https://doi.org/10.1080/02813430903072215.
[40] R.E. Bellman, Dynamic Programming, Princeton University Press, 1957.
[41] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, Q. Liao, Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review, Int. J. Autom. Comput. 14 (5) (2017) 503–519.
[42] F. Bach, Breaking the curse of dimensionality with convex neural networks, J. Mach. Learn. Res. 18 (1) (2017) 629–681.