Medicine and Health Care as a Data Problem: Will Computers Become Better Medical Doctors?

Michael Duerr-Specht[1(✉)], Randy Goebel[2], and Andreas Holzinger[3]

[1] Duke University Hospital, 40 Duke Medicine Circle, Durham, NC 27710, USA
mikespecht@me.com

[2] Department of Computing Science, University of Alberta, 2-21 Athabasca Hall, Edmonton, AB, Canada
goebel@cs.ualberta.ca

[3] Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Auenbruggerplatz 2/V, 8036 Graz, Austria
a.holzinger@hci-kdd.org

Abstract. Modern medicine and health care in all parts of our world are facing formidable challenges: exploding costs, finite resources, aging population as well as deluge of big complex, high-dimensional data sets produced by modern biomedical science, which exceeds the absorptive capacity of human minds. Consequently, the question arises about whether and to what extent the advances of machine intelligence and computational power may be utilized to mitigate the consequences. After prevailing over humans in chess and popular game shows, it is postulated that the biomedical field will be the next domain in which smart computing systems will outperform their human counterparts. In this overview we examine this hypothesis by comparing data formats, data access and heuristic methods used by both humans and computer systems in the medical decision making process. We conclude that the medical reasoning process can be significantly enhanced using emerging smart computing technologies and so-called computational intelligence. However, as humans have access to a larger spectrum of data of higher complexity and continue to perform essential components of the reasoning process more efficiently, it would be unwise to sacrifice the whole human practice of medicine to the digital world; hence a major goal is to mutually exploit the best of the two worlds: We need computational intelligence to deal with big complex data, but we nevertheless – and more than ever before – need human intelligence to interpret abstracted data and information and creatively make decisions.

Keywords: Medical decision support · Medical reasoning · Big data · Data centric medicine · Medical informatics · Smart health

1 Introduction: The Case of Watson Winning Jeopardy!

“If our brains were simple enough for us to understand them, we’d be so simple that we couldn’t.”
Ian Stewart, The Collapse of Chaos: Discovering Simplicity in a Complex World

A. Holzinger et al. (Eds.): Smart Health, LNCS 8700, pp. 21–39, 2015.
DOI: 10.1007/978-3-319-16226-3_2

On January 14th, 2011, the well known game show *Jeopardy!* offered something completely new to their large audience. For the first time since its inception in 1964, one of the participants answering the challenges offered by moderator Alex Trebek was not human – but a computer.

Watson, an IBM computer system, took the central seat between two human champions, Ken Jennings and Brad Rutter. 14 years after the famous victory of “Deep Blue” over Gary Kasparov in chess, a software system – some would describe it rather as artificial intelligence, or smart computing device – beat its human counterparts by a significant margin, though not with a perfect score.

Advances of computer systems have been remarkable since their early beginnings in the 1950’s. We have advanced far from arrays of soldered transistors on large circuit boards in air-conditioned rooms shielded with copper plates against interfering electromagnetic radiation. Capacities in computing power and data storage have doubled every 18–24 months over the past decades (Moore’s law [1, 2]) and the computing power today has reached a level that allows systems to compete in the human field in complex environments. For proponents of computer technology and artificial intelligence this marks the beginning of a new era in human–computer interaction; and the next battlefield to be conquered by computer technology is human medicine and health – smart health [3].

In fact, the challenges in the medical field are enormous [4] and seem to be increasing daily. Medical knowledge is growing at an exponential rate, far beyond what an individual medical doctor can be expected to absorb. It is estimated that a family physician would have to read new medical literature for more than 600 h per month to stay current [5]. The global availability of information on the press of a button is tempting, but the cognitive capacity reaches its natural limit. More information does not lead automatically to better decisions: the US American Institute of medicine (IOM) estimates that medical errors in the United States alone cost 98,000 lives a year at a financial cost of $29 billion a year [6]. The financial burden of health care on just the US economy is staggering and has been steadily increasing over decades raising some immediate questions:

(1) Are future computer systems going to provide a solution to these issues?
(2) Are we looking at a future where diagnoses and treatment decisions for the majority of health problems will be rapidly, accurately and efficiently made by smart Watson-like systems and their offspring?
(3) Or will computer systems, just like technological medical advances in the past, simply accelerate expenses in the health care arena without significantly impacting life expectancy or well-being for the average person?

Certainly we first need to understand the limitations of the human decision making process to understand how to design and build such systems. We also have to face the challenges of the enormous *complexity* of the medical domain [7].

2 Glossary

Abductive Reasoning: a logical inference that leads from an observation to a hypothesis explaining the observation, seeking to find the simplest and most likely explanation (allows inferring a as an explanation of b; deduction in reverse) [8].

Big Data: a buzz word to describe growing amounts of large data sets, having strong relevance from a socio-technical perspective and economy [9] and a recent important topic in biomedical informatics [10].

Content Analytics: umbrella term for the application of machine intelligence to any form of digital content.

Deductive Reasoning: logical inference that links premises with conclusions (allows deriving b from a, where b is a formal logical consequence of a;) [11].

DeepQA: core of the Watson project at IBM on a grand challenge in Computer Science: to build a computer system, that is able to answer natural language questions over an open and broad range of knowledge [12].

Evidence Based Medicine (EBM): is about the integration of individual clinical expertise with the best external evidence [13], originally coming from medical education for improving decision making strategies [14].

Inductive Reasoning: allows inferring b from a, where b does not necessarily follow from a; this reasoning is inherently uncertain (e.g. Billo is a boxer, Billo is a dog > all dogs are boxer)

Machine Learning: field of study and design of algorithms that can learn from data, operate by building a model based on inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions. The more data the better, hence big data is good for machine learning [15].

Reasoning: associated with thinking, cognition, and intelligence it is core essence in decision making as it is one of the ways by which thinking comes from one idea to a related idea (cause-effect, truth-false, good-bad, etc.) [16].

Watson: synonym for a cognitive technology system of IBM, using DeepQA that tries to process information more like a human than a computer with the goal of understanding natural language based on hypothesis generation and dynamic learning [17].

Watson Content Analytics: a business solution, based on Watson technology for knowledge discovery from unstructured information aiming to help enterprises to validate what is known and reveal what is unknown, having much potential for medicine and health care [18].

3 What Is This Watson?

In 2007, the IBM Research labs began with the grand challenge of building a computer system that could eventually compete with human champions at the game quiz show *Jeopardy!* In 2011, the open-domain question-answering (QA) system, called Watson (in honor of Thomas J. Watson (1874–1956), CEO of IBM from 1914 to 1956), beat the two highest ranked players in a nationally televised two-game *Jeopardy!* competition.

Technologically, Watson is the result of a systems integration of many diverse algorithmic techniques, performed at champion levels [12, 19], and was created to demonstrate the capability of so-called DeepQA technology [20]: this architecture was designed to be *massively parallel,* with an expectation that low latency response times could be achieved by doing parallel computation on many distributed computing systems. A large set of natural-language processing programs were integrated into a single application, scaled out across hundreds of central processing unit cores, and optimized to run fast enough to compete in a real-world application [21].

Because Watson cannot hear or see, when the categories and clues were displayed on the game board, they were inputted manually (as text) to Watson. The program also monitored signals generated when the buzzer system was activated and when a contestant successfully rang in. If Watson was confident of its answer, it triggered a solenoid to depress its buzzer button and used a text-to-speech system to speak out loud its response – to make the output more appealing. Since it did not hear the host's judgment, it relied on changes to the scores and the game flow to infer whether the answer was correct or not. The Watson interface program had to use what were sometimes conflicting events to determine the state of the game, without any human intervention [22].

Deep Blue, the computer that played chess and beat world champion Garry Kasparov in 1996 first, and in 1997 the six-game match [23], was also impressive, but far not as impressive as Watson [24]. Deep Blue operated in a finite and well specified problem space. Though the chess problem space is large (estimated to be greater than 10^{120} positions) making it impossible for computers to calculate every potential outcome [25] it could certainly calculate the merits of every immediate move and the possible alternatives two or three moves ahead [26]. Combined with some strategic knowledge, it was able to beat any opponent at chess.

The problem space that Watson took on was much less well defined and required the interpretation of natural language to form and select an appropriate answer. Exactly this is the big problem: Whereas chess programs tend towards performing "super-human", i.e. perform better than all humans, natural language processing, i.e. word sense disambiguation is traditionally considered an AI-hard problem [27, 28].

For those of us who study both human and artificial intelligence, the question arises as to what extent to which Watson mimics human intelligence [29, 30]. In the past, human intelligence researchers and many artificial intelligence researchers have dismissed the possibility of any strong similarity between artificial and human intelligence [31]. This was almost certainly correct for any past accomplishment in artificial intelligence, especially which focused on games and search.

Could Watson be different? It is very likely that Watson would do quite well on many test items that compose intelligence tests including general information, vocabulary, similarities, and nearly anything dependent on verbal knowledge. Nevertheless, it is very likely that Watson would do quite poorly on many other kinds of tests that require reasoning or insight. In its current state, it would be difficult for Watson to understand directions for the various and different subtests that usually make up an intelligence test, something that children as young as three or four do easily.

Tests of computer intelligence are as old as computers themselves. The most famous is the Turing test proposed by Alan Turing (1912–1954) more than 60 years ago. Turing suggested that a machine would be intelligent when an observer would have a conversation with a computer and a real person and not be able to distinguish which was which [32].

Numerous other approaches have been proposed, including the construction of a unique battery of tests that would provide an actual IQ score for artificial intelligence systems, similar to the way human IQ scores are determined [33]. This challenge is supported by the editorial board of the journal "Intelligence" and members of the International Society for Intelligence Research.

A recent paper by Rachlin [34] speculates on the abilities Watson would need, in addition to those it has, to emulate true human behaviour: essential human attributes such as consciousness, the ability to love, to feel, to sense, to perceive, and to imagine. Most crucially, such a computer may exhibit self-control and may act altruistically.

At this point, the external perception of Watson's performance in *Jeopardy!* exposes only "question in – single answer out" with no detailed explanation of *how* the answer was found. However, internally, Watson uses a form of hypothetical reasoning called *"probabilistic abduction*", e.g., see [35], which creates and ranks alternative answers based on the alternatives that can be inferred from a variety of text resources within the time limit for a response.

Currently the IBM team is working on a vision for an evidence-based clinical decision support system, based on the DeepQA technology, that affords exploration of a broad range of hypotheses and their associated evidence, as well as uncovers missing information that can be used in mixed-initiative dialog [17]. Whereas Watson used simple but broad encyclopaedic knowledge for the *Jeopardy!* task, the extended medical Watson uses medical information gleaned from sources also available to the practicing physician: medical journals and textbooks.

Considering the fact that medicine is turning more and more into a data intensive science, it is obvious that integrated machine learning approaches for knowledge discovery and data mining are indispensable [36].

The grand goal of IBM is having the Watson technology ready as a medical doctor's assistant (in German: Arzthelfer) available on a mobile computing device by the year 2020 [37]. This is a grand challenge exactly at the intersection of human-computer interaction (HCI) and knowledge discovery and data mining (KDD) [38].

4 Computers and Medicine

Undeniably, computers offer ever-increasing capabilities in areas where we as humans have trouble competing. Information entered can be reproduced accurately and without degradation innumerable times to as many users as desired. Data elements can be grouped, parsed, abstracted, combined, copied, and displayed in any conceivable way, offering seemingly infinite options to view even very large sets of data. Numeric values can be instantly analyzed and participate in complex calculations, the result of which is immediately available. Network environments allow multiple users to access and share identical information in real time.

Not least, advances in natural language processing allow an increasingly accurate analysis of data contained within unstructured written documents and reports. In fact, IBM's *Jeopardy!* contestant Watson is the first system to show the power of this technology in quasi real time [39, 40].

In medicine, electric and electronic devices as well as computerized systems have an impressive track record. Assistance in diagnosis dates back to 1895 when Conrad Wilhelm Roentgen (1845–1923) demonstrated that the use of electromagnetic radiation, invisible to the naked eye was able to penetrate tissues and visualize the bones [41]. Advances in this basic technology combined with the computing power of modern semiconductors led to the development of sophisticated imaging technology such as computerized tomography offering 2-D and 3-D images of internal tissues and organs as well as real time fluoroscopy, providing essential information for tricky medical interventions. Magnetic resonance imaging provides digital images constructed from atomic nuclear resonance of body tissues allowing for accurate visual diagnoses of many pathological conditions. Computer generated images based on the analysis of ultrasound waves reflected from body tissues have become an indispensable tool in the evaluation of internal organs and prenatal care.

In the realm of therapy, computers led to significant medical advances. Cardiac pacemakers generate a cardiac rhythm in cases of failure of the innate sinus node. Implanted automatic defibrillators continuously analyze the electrical system of the heart and sophisticated algorithms determine the selection of multiple response modes which allow the device to save a failing heart and prolong the patient's life.

The use of electronic medical records has been steadily growing over the past 10 years. Today about 42 % of US hospitals utilize some type of electronic documentation [42]. In 2011 the US government introduced the concept of "Meaningful Use" initially offering financial incentives to increase the adoption of computerized record systems. The advance of electronic documentation has created a fertile basis for diagnostic and therapeutic decision support systems, which have diversified significantly over time. Some projects have taken into account newest findings in neurocognitive research ("cognostics") for their human–computer interface development, and are adapting principles and methods to ideally support the human cognitive process with its interactive analytical and constellatory operations [43]. Most recently, the market penetration with small computing devices such as smart phones and table computers has shifted medical referencing from printed media to electronic devices, though the analysis of accumulated patient data is yet to be abstracted and codified in a manner that would easily amplify the abilities of the clinical decision maker – but that is anticipated to come.

5 The Digital Challenge

In view of these impressive advances in technology and computing it is not surprising that a debate has been sparked as to whether the continuation of this development will lead to a future situation in which computers will eventually outperform human doctors and consequently assume larger roles in medical diagnostics and therapeutic decisions.

The venture capitalist Vinod Khosla states in an interview published by Techcrunch on January 10th, 2012 that

> *"we cannot expect our doctor to be able to remember everything from medical school twenty years ago or memorize the whole Physicians Desk Reference (PDR) and to know everything from the latest research, and so on and so forth. This is why, every time I visit the doctor, I like to get a second opinion. I do my Internet research and feel much better."* [44].

In order to better understand the issues involved in the potential of computing technology and artificial intelligence as it is integrated into medical decision making, it might be worthwhile to differentiate between three different aspects of the process of delivering health care to a patient: data formats and relationships amongst those, the accumulation of large volumes of medical data contained within databases, and the reasoning process used to interpret and apply those data to benefit a specific patient.

6 Data Formats and Relationships

Here we follow the definitions of Boisot & Canals [45], who describe data as originating in discernible differences in physical states of the world. Significant regularities in this data constitute *information*. This implies that the information gained from data, depends on the expectations, called: *hypotheses*. A set of hypotheses is called *knowledge* and is constantly modified by new information. This definition fits well to the human information processing model by Wickens [46]: The physical stimuli (cues) are selected by the attentional resources and the perceived information builds working hypotheses H1, H2, H3 … etc., which are constantly compared and judged against available hypotheses, already present in the long-term memory. On this basis the best possible alternative will be chosen and actions A1, A2, A3, … etc., performed according to likelihoods and consequences of the outcomes – which can be perceived again via the feedback loop. Wickens described the input "filter" as the "nebula of uncertainty" and this emphasizes perfectly a general problem in decision making: we deal always with probable information. Each information chunk, item or whatever you call it, has always a certain probability aspect (refer to lecture 7 in [10]).

Based on these definitions, the commonly used term "unstructured data" might just capture random state descriptors – uncertainty – noise [47]. In Informatics, particularly, it can be considered as unwanted non-relevant data without meaning within a given data model – or, even worse, with an interpretation assigned in error, hence modelling of artefacts is a constant danger in medical informatics [48].

The question "what is information?" continues to be an open question in basic research. Any definition depends on the view taken. For example, the definition given by Carl-Friedrich von Weizsäcker (1912–2007): "Information is what is understood," implies that information has both a sender and a receiver who have a common understanding of the representation within a shared modelling system and the means to communicate information using some properties of the physical systems. His addendum: "Information has no absolute meaning; it exists relatively between two semantic levels" implies the necessity of context [49].

There is a distinct and insurmountable difference between human and computer data formats. Computers – at least as current electronic methods of computation on the basis of Von Neumann machines [50] are concerned – operate exclusively with digital data formats. Content is stored as strings of binary data elements. Meaning and relationships between content items are added by method of (human) assignment or (machine) calculation, i.e., they are subsequently provided as additional data layer to the original content data.

All data elements are provided to computer systems by means of human interaction (e.g., keyboard, touch pad, etc.), technical devices (sensors, microphones, cameras, etc.) or secondary data generated through calculation of previously available data. As outlined above, regardless of the original data format, making data available to computer technology has an absolute requirement of translating data into a binary format, regardless of the original complexity. All relationships between data elements are initially stripped from the occurrence observed, though some may be added back and preserved for future use (e.g. machine learning) by adding additional documentation. As such, every pixel of a picture file has no "knowledge" of its neighbor pixel and the composition of the picture is provided by a separate instruction, providing placement of the pixel within a defined grid format. This feature provides the power and flexibility of digital image processing, as individual elements can be altered without affecting the remainder of the composition.

Humans, on the other hand, have the luxury of a primary experience of their environment. We traditionally speak of our five senses, sight, hearing, taste, smell and touch, though other senses also provide data, such as temperature, balance, pain, time and another less well-understood class of senses commonly referred to as "intuition". As we experience life, it appears that the input from the individual senses interact and is stored in a complex fashion with a high degree of connectedness between individual data items. The human subjective experience overrules the (measurable) objective state and data content is difficult to differentiate from data interpretation. This has been demonstrated quite impressively in pictures such as the checker-shadow optical illusion [51, 52]. In spite of the lack of objectivity, however, the skewed data perception by the human observer has advantages, as it puts acquired data content into a *contextual perspective* thereby allowing the all-important differentiation between relevant and irrelevant items.

This ability differentiates human perception from machine analysis. It is used effectively to block automated software responses, e.g. with the so-called CAPTCHA ("Completely Automated Public Turing test to tell Computers and Humans Apart") [53–55]. It is also the basis of the human ability to analyze visual motion in general scenes which – as of now – exceeds the capabilities of even the most sophisticated computer vision algorithms [56].

7 On Data and Context

Regardless of whether examining a computer system or the human brain, calculations and conclusions can only be based on data available to the system. For technical systems this means that the basis of all operations is the availability of binary input data. Though large computer systems such as IBM's Watson have access to databases of many terabytes and can process data at the rate of more than 500 gigabytes per second, access is still restricted to data published or digitally available in other formats. This makes these systems vulnerable to the GIGO ("garbage in, garbage out") phenomenon, which plagues large data environments [57].

As in other fields with computerized data entry, medical documentation in health records is biased by documentation guidelines, template requirements and constraints on entry formats, as well as reimbursement requirements, etc., and does not accurately reflect the complete array of signs and symptoms of a patient's presentation.

The human brain, on the other hand, does not rely on single-bit data streams as its input method. We have a complex and analog (not binary digital) experience of our surroundings, which delivers simultaneous perception data from all senses. Acquired information from one sensory organ is therefore never one-dimensional but experienced in the multidimensional context of the other senses, thereby adding meaning to the data. In contrast to technical systems however, the human brain reduces the original environmental data quantity, according to principles of interest or "meaningfulness," led by attentional resources [58] though the exact mechanism of information condensation and subsequent storage still remains poorly understood.

Some catchy numbers shall highlight the comparison between human and computer (although the following numbers are theoretical and the exact function of human information processing is not known yet):

The human eye has been estimated to be able to perceive 6 Mbits/s of primary visual information, however less than estimated 1 % of this information reaches the visual cortex of the brain for further cognitive processing [59, 60]. By the time information reaches our consciousness, the rate of information flow has been estimated to shrink to about 100 bits/s for visual sources, 200 bits/s for all sensory sources combined [61]. At this rate it would take about 300,000 years for a human to obtain the data utilized by Watson in the 2011 *Jeopardy!* contest.

8 Reasoning Process

Reasoning is according to the common definition "the process of thinking about something in a logical way in order to form a conclusion or judgment" [62, 63]. In medicine, we apply reasoning to come to a conclusion about which diagnosis would be appropriate for a patient presenting with a certain constellation of signs and symptoms. The reasoning process typically applied by medical doctors has been described to include abductive, deductive and inductive reasoning elements [35]. Upon presentation of the patient, the physician will first generate domain-specific hypotheses based on an initial set of observations (abduction). These initial hypotheses are confirmed and refined by additional observations (induction). Textbook knowledge of the disease entity is then employed to select the appropriate treatment to improve the patient's health (deduction).

Adopting this concept of the medical reasoning process, it can be argued that man and machine have complementary strengths and weaknesses (see Table 1). The abductive component provides the basis for hypothesizing known causes or diseases that imply the observed symptoms. This initiates the diagnostic process and is uniquely supported by the human high dimensional intrinsically correlated mechanism of perception which includes concrete observations that can be measured and documented in the record and is highly supported by soft factors (sense of illness severity and distress, emotional state, social environment, etc.), typically not documented or even consciously recognized. The human brain is capable of rapidly associating this overall picture with known disease patterns and can thereby not only very efficiently postulate the hypotheses within the abductive process but also intuit measures to manage cases in which the observed pattern does not sufficiently match known entities. The medical literature is full of examples of unique presentations in which the treating physician invoked a creative process of expanding hypotheses beyond what had previously been known or documented (as an example and fun reading please refer to a recently published report by Rice et al. [64]).

Table 1. Medical Reasoning: Human vs. Computer

Reasoning Process	Human	Computer
Abductive Hypothesis generation	Uniquely capable of complex pattern recognition and creative thought. "the whole is greater than the sum of its parts"	Matches multiple individual correlations from extensive data banks based on preconceived algorithms. Secondary construction of relationships. "the whole equals the sum of its parts"
Inductive Symptom → Disease	Limited database. Subject to biases - Anchoring bias - Confirmation bias - Premature closure	Extensive database. Probability based on Bayesian statistics, no significant bias. Limitation based on available data.
Deductive Disease → Symptoms, Treatment	Limited database. Personal intuition and experience affect decision making.	Extensive database. Application of rules of evidence based medicine with potential biases.

Computerized systems, on the other hand, only can use data supplied as binary code which is processed in pre-conceived algorithms and have no original perception of "Gestalt." Relationships are based on correlations extracted secondarily from extensive databases, however the whole remains equal to the sum of the parts. What seems elusive is an information model or structure that emulates the emergence of genuinely novel concepts and ideas the human mind is capable of, without succumbing to a reductionist view of acknowledged but unformulated physician insights. At the very least, the information system support would have to consider a wider breadth of feasible hypotheses guided by the deeper experiences of physicians, which would require a much more sophisticated and extensive conceptual underpinning of the language in which hypotheses are expressed.

Human reasoning in the practice of medicine is, however, hampered in the inductive process of confirming and refining abductive hypotheses due to biases and poor understanding of probability calculations. The tendency of physicians to retain their initial hypotheses even in the light of contradicting data is a well-described phenomenon [65,66].

Multiple concepts of bias are described in this context:

- Anchoring bias: focusing on a single concept before sufficient data is available to support it,
- Confirmation bias: gathering only information to support an hypothesis, and
- Premature closure: terminating the reasoning process and eliminating evaluation of alternative explanations prematurely;

Computer systems are not prone to these biases. The computer has no urge to favor one hypothesis over another but rather uses information from extensive medical databases as entry data for probability calculations, often along the lines of Bayesian statistics. Conceptually this approach is supported by the probabilistic nature of information, and the role of reasoning to calculate and identify the most probable hypotheses. In contrast to computer algorithms, a recent study reports that most physicians misunderstand the underlying probabilistic logic of significance tests and consequently often misinterpret their results. The study concludes that a solid understanding of the fundamental concepts of probability theory is becoming essential to the rational interpretation of medical information per se [48].

One of the earliest software tools in medical reasoning, MYCIN, developed by the Stanford Medical Center in the 1960's was based on the concepts of inductive reasoning [67, 68]. In the same manner, well-designed clinical reasoning software could be of significant value in alerting physicians about possible bias in their decision process, assisting in the probability calculations and helping to minimize or avoid clinical error.

Sophisticated access to the knowledge of large medical databases could also assist in the deductive phase of medical reasoning. In selecting the most likely diagnosis among a selection of differential diagnostic considerations, specific tests and exams are necessary. Physicians generally have a very poor track record in selecting the course of clinical tests that provides for the most efficient information gain. Often studies are ordered according to individual habits with limited understanding or consideration of how the test results affect the likelihood of a disease being present or not. Software with access to extensive data regarding prevalence of disease entities in specific populations as well as the sensitivity and specificity of diagnostic studies would offer guidance to an efficient selection of tests to confirm or refute a diagnosis as it relates to a particular patient presentation.

Once a diagnosis has been established, the decision on therapeutic interventions can also be assisted by medical software. Unlike their human counterparts, computers have access to all published information and recommendations and can suggest the intervention that is most current. In addition, the broader influence of historical data as well as subtle trends can be considered, which is difficult and time challenging for humans. Since 2011, IBM Watson's capabilities in assisting in treatment decisions are being studied by multiple medical facilities, including Columbia University, the Memorial Sloan-Kettering Cancer Center as well as the Cleveland Clinic [69]. In the German speaking world progress has been made particularly at Graz University Hospital [18], and Vienna University Hospital [70] – two of the largest hospitals in Europe.

Computerized assistance in medical treatments are based on the principles of "evidence based medicine," an approach that is led by the idea that the best treatment is one based on the results of published trials and applies findings to the individual patient. While this treatment philosophy represents an understandable ideal, it is subject to significant limitations, among others: selection of study population, publication bias, bias based on financial incentives and errors in study results due to incomplete understanding of the biological system (e.g., Simpsons paradox [71]). In addition, computer generated treatment recommendations exclude the personal experience and intuition of the treating physician. Recent research further elaborates on the dual processing theory of human cognition [72] and a recent study reports that reasoning and decision-making can be described as a function of both an intuitive, experiential and affective system (system 1) and/or an analytical, deliberative processing system (system 2) [73].

9 Future Challenges

Faced with unsustainable costs and enormous volumes of under-utilized data, health care needs more efficient practices, research, and tools. It is our opinion that there is tremendous potential in harnessing the power of modern information technology and applying it to the medical sciences.

We believe that the challenges and future work needed to support medicine, health and well-being with software products can be categorized in three distinct areas: organizational (including administrative and political), technological and educational:

Area 1: Organizational/administrative/political

- data access and data ownership issues;
- balancing legitimate privacy concerns with the benefits of access to large amounts of anonymized open clinical data for public and personal health assessment;

Area 2: Technological

- building new software products based on existing technology and using available digitally stored data elements, with a special focus on visual representation of complex clinical data, trending of individual health parameters and weak signal detection;
- developing intuitive medical record systems to allow for improved documentation of the process of care and medical reasoning and promoting continuity of care during the hand-off process between health care providers
- enhancing digital data capture through newly designed intelligent user interfaces and/or secondary processing by means of natural language processing and content tagging
- developing new hardware products to automatically capture relevant physiological data, e.g. along the lines of the quantified self movement
- promoting preventative care by analyzing large amount of high quality clinical data to detect weak signals that serve to risk stratify for future health events
- continuing research in artificial intelligence and machine learning and testing concepts of software systems acting as legitimate sparring partners in sophisticated medical decision making, which is still the core area of biomedical informatics [56].

Area 3: Educational

- promoting and supporting interdisciplinary events in which software engineers and medical professionals exchange ideas and concepts and develop a common language in describing domain specific needs.

We envision a future where medical doctors can ask questions to the available data and have an integrative overview of both the clinical patient data and -omics data (e.g. genomics, proteomics, metabolomics, etc.) [74]. Software support in personal and global health data would allow the expert to find and diagnose diseases in advance, before symptomatically apparent. In this form of *data-centric medicine,* prevention could really become efficient, and the dream of a personalized medicine approach can become true [75]. Although both science and engineering are making significant progress, a lot of work remains to be done within the coming years for this vision to become a reality.

The integration of technology into clinical medicine includes at least three broad classes of challenges. In our discussion regarding the role of Artificial Intelligence, it is clear that there are a large variety of technologies that can begin by augmenting and amplifying the value of clinical practioners:

(1) improvements in diagnostic sensing and imaging; capture and rapid deployment of new medical knowledge,
(2) logistics and management improvements in both small clinics and hospitals, and
(3) improvement in the capture, security, and use of medical data.

Not all of these challenges are technical. In fact 2 and 3 are largely organizational challenges, partly due to educational lag and the pace with which modern medical management adopts technologies that are already available. These include not just actionable medical knowledge and technology, but operational management and technology procurement. Challenge 3 is largely about the development and exploitation of patient data, where two major impediments exist. One is simply the evolutionary adoption of standards of data capture and use, partly at capture time, where capture, storage and open access must be addressed. Subsequent to that, medical ontology systems, which provide the foundation for aggregating data, and using analytics (machine learning) to find trends and help improve clinical practice.

A more serious challenge is the development and deployment of medical data governance models, into which public, government, and medical organizations can collaborate to develop the trust to actually use medical data. Many jurisdictions are recognizing that data security methods have never been better, so that the governance of medical data, and building public trust for its value is the key.

Thus, in the effort to achieve superior and cost-effective medical care by virtue of integration of physician expertise and computerized clinical decision support systems (CDSS), the following issues need to be addressed:

Issue 1. Negotiating the contradiction between structured digital data capture and the expressive narrative in clinical documentation: Whereas downstream use and reuse of clinical data in decision support systems requires data that is highly structured and standardized, practicing clinicians demand documentation systems that afford flexibility and efficiency and easily integrates into busy and hectic workflows [76]. In order to successfully implement computerized clinical support systems, EHR solutions will have to be developed in a way that satisfy the needs for clinical workflow support, documentation of unsuspected circumstances, machine readability and structured data capture.

Issue 2. Development and adoption of a standardized biomedical "language". Automated data capture processes and electronic health records are producing data sets too large to be manually analyzed or processed. Therefore it is important that clinical data can be tagged according to a common biomedical ontology to allow for widespread international data sharing and analysis [77].

Currently several competing ontologies are being used, serving various interests in the biomedical domain (e.g. UMLS, MeSH, GALEN, SNOMED, ICD), however, all these are difficult to use and rather impossible to map to each other due to inconsistent representation.

Issue 3. Regulatory and legal framework. Legal exposure to practicing physicians can result from errors due to flawed design or functionality of computerized clinical support systems, or their improper use. Currently there exist few standards for the design and development of automated decision support systems and there have been calls to enhance current functionalities and create tools to avoid automation associated errors [78]. Changes to the regulatory framework have been recommended [79]. Furthermore, as recommendations based on computerized algorithms and decision support systems become part of the practice reality in the medical field, legal structures need to be adapted to allow physicians to base diagnostic and treatment decisions on their individual acumen and expertise, even if in disagreement with machine recommendations, without immediate legal exposure.

Issue 4. Inhibitory medical data protection regulations. While patients have a valid interest in protecting confidential medical data, overly protective limitation to access community health care data thwarts medical research and knowledge development and can harm general public health interests. In the interest of advancing medical knowledge and quality of care it will be necessary to increase access to biomedical information whilst at the same time protecting legitimate individual privacy interests.

Issue 5. Creating a dynamic educational system. It is shocking that the average transfer time for medical knowledge from initial research to widespread implementation in medical practice has been estimated to be between 12 and 17 years [80]. As we increase our ability to fuel computerized clinical decision support systems with real time date, processes need to be developed to extract knowledge regarding diagnoses and optimal treatment and make this available to the medical practitioners. Adjusting to this dynamic decision environment will require a new mind set in programmers, policy makers and practitioners.

10 Summary

Since their early beginnings, more than half a century ago, computer systems have evolved into highly complex data environments that are able to rapidly deliver vast amounts of information. It has been postulated that the computing power of advanced systems will be able to provide medical care to patients in the near future that will be more efficient and of higher quality and lower cost than currently offered by physicians. While this is probably overly enthusiastic, current developments in medical software promise an exciting future for physicians. Needed information will be delivered to our fingertips without delay. Intelligent selection algorithms will allow us to rapidly review case-relevant studies and protocols.

Unusual constellations of signs and symptoms will be screened for rare diseases and suggested for consideration. Our electronic medical records will be smart in prompting us to answer only the questions that are relevant for case-specific decision-making. Graphical user interfaces will make it easy to detect and review even subtle trends and compare symptom constellations of the differential diagnoses under consideration.

Software capabilities have graduated to the professional league of medical care. As the pilots in the diagnostic and therapeutic process, we as physicians are now called to step up to the plate and engage in active conversations with software developers and IT departments. Mustering this initiative will allow us to leverage the unique strengths and capabilities of both information technologies and medical sciences into powerful and effective health care services of the future in which doctors will be able to navigate the complex landscape of a patient's health information similar to how an airline pilot manages a complex flying machine with the assistance of a the sophisticated flight data display of a computerized glass cockpit.

Computers cannot become better medical doctors. Medical doctors can become better medical doctors with the support of smart hospital systems [3]. Information technology and medical sciences are not battling for territory in a zero sum game. If we approach it correctly, everyone wins, most importantly: our patients!

11 Epilogue

As for the triumph of IBM Watson in the *Jeopardy!* game show: the amazing observation, one may argue, is not that Watson won, employing its database of four terabytes, cluster of 90 IBM Power 750 serves each using a 3.5 GHz Power7 eight core processor and able to push the response button within 5 ms, 20 times the human response time. The amazing thing is that the human contestants scored. Just imagine what the two forces combined could achieve [38].

References

1. Moore, G.E.: Cramming more components onto integrated circuits. Electronics **38**(8), 114–117 (1965)
2. Cavin, R., Lugli, P., Zhirnov, V.: Science and engineering beyond Moore's Law. IEEE Proc. **100**(13), 1720–1749 (2012)
3. Holzinger, A., Röcker, C., Ziefle, M.: From smart health to smart hospitals. In: Holzinger, A., Röcker, C., Ziefle, M. (eds.) Smart Health. LNCS, vol. 8700, pp. 1–19. Springer, Heidelberg (2015)

4. Holzinger, A., Dehmer, M., Jurisica, I.: Knowledge discovery and interactive data mining in bioinformatics - state-of-the-art, future challenges and research directions. BMC Bioinform. **15**(S6), I1 (2014)
5. Alper, B.S., Hand, J.A., Elliott, S.G., Kinkade, S., Hauan, M.J., Onion, D.K., Sklar, B.M.: How much effort is needed to keep up with the literature relevant for primary care? J. Med. Lib. Assoc. **92**(4), 429–437 (2004)
6. Kohn, L.T., Corrigan, J., Donaldson, M.S.: To err is human: building a safer health system. National Academy Press, Washington (DC) (2000)
7. Patel, V.L., Kahol, K., Buchman, T.: Biomedical complexity and error. J. Biomed. Inform. **44**(3), 387–389 (2011)
8. Poole, D.: Explanation and prediction: an architecture for default and abductive reasoning. Comput. Intell. **5**(2), 97–110 (1989)
9. Dong-Hee, S., Min Jae, C.: Ecological views of big data: Perspectives and issues. Telematics Inform. **32**(2), 311–320 (2015)
10. Holzinger, A.: Biomedical Informatics: Discovering Knowledge in Big Data. Springer, New York (2014)
11. Johnson-Laird, P.N.: Deductive reasoning. Ann. Rev. Psychol. **50**, 109–135 (1999)
12. Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A.A., Lally, A., Murdock, J.W., Nyberg, E., Prager, J.: Building Watson: An overview of the DeepQA project. AI Mag. **31**(3), 59–79 (2010)
13. Sackett, D.L., Rosenberg, W.M., Gray, J., Haynes, R.B., Richardson, W.S.: Evidence based medicine: what it is and what it isn't. BMJ. Br. Med. J. **312**(7023), 71 (1996)
14. Guyatt, G.: Evidence based medicine - a new approach to teaching the practice of medicine. JAMA-J. Am. Med. Assoc. **268**(17), 2420–2425 (1992)
15. Holzinger, A.: Trends in interactive knowledge discovery for personalized medicine: cognitive science meets machine learning. Intell. Inf. Bull. **15**(1), 6–14 (2014)
16. Holzinger, A.: Lecture 8 biomedical decision making: reasoning and decision support. In: Biomedical Informatics, pp. 345–377. Springer (2014)
17. Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., Mueller, E.T.: Watson: Beyond jeopardy! Artif. Intell. **199–200**, 93–105 (2013)
18. Holzinger, A., Stocker, C., Ofner, B., Prohaska, G., Brabenetz, A., Hofmann-Wellenhof, R.: Combining HCI, natural language processing, and knowledge discovery - potential of IBM content analytics as an assistive technology in the biomedical field. In: Holzinger, A., Pasi, G. (eds.) HCI-KDD 2013. LNCS, vol. 7947, pp. 13–24. Springer, Heidelberg (2013)
19. Ferrucci, D.A.: Introduction to "This is Watson". IBM J. Res. Develop. **56**, 3–4 (2012)
20. Gondek, D.C., Lally, A., Kalyanpur, A., Murdock, J.W., Duboue, P.A., Zhang, L., Pan, Y., Qiu, Z.M., Welty, C.: A framework for merging and ranking of answers in DeepQA. IBM J. Res. Develop. **56**, 3–4 (2012)
21. Epstein, E.A., Schor, M.I., Iyer, B.S., Lally, A., Brown, E.W., Cwiklik, J.: Making Watson fast. IBM J. Res. Develop. **56**, 3–4 (2012)
22. Lewis, B.L.: In the game: The interface between Watson and Jeopardy! IBM Journal of Research and Development 56(3–4), (2012)
23. Newborn, M.: Deep Blue's contribution to AI. Ann. Math. Artif. Intell. **28**(1–4), 27–30 (2000)
24. Campbell, M., Hoane, A.J., Hsu, F-h: Deep blue. Artif. Intell. **134**(1), 57–83 (2002)
25. Newell, A., Shaw, J.C., Simon, H.A.: Chess-playing programs and the problem of complexity. IBM J. Res. Dev. **2**(4), 320–335 (1958)
26. Ensmenger, N.: Is chess the drosophila of artificial intelligence? A social history of an algorithm. Soc. Stud. Sci. **42**(1), 5–30 (2012)

27. Navigli, R., Velardi, P.: Structural semantic interconnections: A knowledge-based approach to word sense disambiguation. IEEE Trans. Pattern Anal. Mach. Intell. **27**(7), 1075–1086 (2005)
28. Bhala, R.V.V., Abirami, S.: Trends in word sense disambiguation. Artif. Intell. Rev. **42**(2), 159–171 (2014)
29. Tesauro, G., Gondek, D.C., Lenchner, J., Fan, J., Prager, J.M.: Simulation, learning, and optimization techniques in Watson's game strategies. IBM J. Res. Develop. **56**, 3–4 (2012)
30. Tesauro, G., Gondek, D.C., Lenchner, J., Fan, J., Prager, J.M.: Analysis of WATSON's strategies for playing jeopardy! J. Artif. Intell. Res. **47**, 205–251 (2013)
31. Simon, H.A.: Studying human intelligence by creating artificial intelligence. Am. Sci. **69**(3), 300–309 (1981)
32. Turing, A.M.: Computing machinery and intelligence. Mind **59**(236), 433–460 (1950)
33. Detterman, D.K.: A challenge to Watson. Intelligence **39**(2–3), 77–78 (2011)
34. Rachlin, H.: Making IBM's computer, watson, human. Behav. Anal. **35**(1), 1–16 (2012)
35. Sun, Z.H., Finnie, G., Weber, K.: Abductive case-based reasoning. Int. J. Intell. Syst. **20**(9), 957–983 (2005)
36. Holzinger, A., Jurisica, I.: Knowledge discovery and data mining in biomedical informatics: the future is in integrative, interactive machine learning solutions. In: Holzinger, A., Jurisica, I. (eds.) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. LNCS, vol. 8401, pp. 1–18. Springer, Heidelberg (2014)
37. Chip. http://business.chip.de/news/IBM-Watson-Sein-erster-Job-ist-Arzthelfer_51623616.html
38. Holzinger, A.: Human-computer interaction and knowledge discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? In: Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L. (eds.) CD-ARES 2013. LNCS, vol. 8127, pp. 319–328. Springer, Heidelberg (2013)
39. Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W.: Natural language processing: an introduction. J. Am. Med. Inform. Assoc. **18**(5), 544–551 (2011)
40. Murdock, J.W., Fan, J., Lally, A., Shima, H., Boguraev, B.K.: Textual evidence gathering and analysis. IBM J. Res. Dev. **56**(3–4), 1–14 (2012)
41. Röntgen, W.C.: On a new kind of rays. Science **3**(59), 227–231 (1896)
42. Robert Wood Johnson Foundation (RWJF). http://www.rwjf.org/en/about-rwjf/newsroom/newsroom-content/2013/07/hospitals–physicians-make-major-strides-in-electronic-health-re.html
43. Miller, R.A.: Medical diagnostic decision support systems - past, present and future. J. Am. Med. Inform. Assoc. **1**(1), 8–27 (1994)
44. Techcruch. http://techcrunch.com/2012/01/10/doctors-or-algorithms/
45. Boisot, M., Canals, A.: Data, information and knowledge: have we got it right? J. Evol. Econ. **14**(1), 43–67 (2004)
46. Wickens, C.D.: Engineering Psychology and Human Performance. Charles Merrill, Columbus (1984)
47. Holzinger, A., Stocker, C., Dehmer, M.: Big complex biomedical data: towards a taxonomy of data. In: Obaidat, M.S., Filipe, J. (eds.) Communications in Computer and Information Science CCIS 455, pp. 3–18. Springer, Berlin Heidelberg (2014)
48. Holzinger, A.: On knowledge discovery and interactive intelligent visualization of biomedical data: Challenges in human–computer interaction & biomedical informatics. In: Helferd, M., Fancalanci, C., Filipe, J. (eds.) DATA - International Conference on Data Technologies and Applications, pp. 5–16. INSTICC, Rome (2012)

49. Marinescu, D.C.: Classical and Quantum Information. Academic Press, Burlington (2011)
50. von Neumann, J.: Electronic methods of computation. Bull. Am. Acad. Arts Sci. **1**(3), 2–4 (1948)
51. Adelson, E.H.: Checkershadow illusion. http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html (last Accessed 24 November 2014) (1995)
52. Geisler, W.S., Kersten, D.: Illusions, perception and Bayes. Nat. Neurosci. **5**(6), 508–510 (2002)
53. von Ahn, L., Blum, M., Hopper, N.J., Langford, J.: CAPTCHA: Using hard AI problems for security. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656. Springer, Heidelberg (2003)
54. Rusu, A., Govindaraju, V., Soc, I.C.: Handwritten CAPTCHA: Using the difference in the abilities of humans and machines in reading handwritten words. IEEE Computer Society, Los Alamitos (2004)
55. Belk, M., Germanakos, P., Fidas, C., Holzinger, A., Samaras, G.: Towards the personalization of CAPTCHA mechanisms based on individual differences in cognitive processing. In: Holzinger, A., Ziefle, M., Hitz, M., Debevc, M. (eds.) SouthCHI 2013. LNCS, vol. 7946, pp. 409–426. Springer, Heidelberg (2013)
56. Weiss, Y., Simoncelli, E.P., Adelson, E.H.: Motion illusions as optimal percepts. Nat. Neurosci. **5**(6), 598–604 (2002)
57. Leunens, G., Verstraete, J., Vandenbogaert, W., Vandam, J., Dutreix, A., Vanderschueren, E.: Human errors in data transfer during the preparation and delivery of radiation treatment affecting the final reasult - garbage in, garbage out. Radiother. Oncol. **23**(4), 217–222 (1992)
58. Kahneman, D.: Remarks on attention control. Acta Psychologica **33**, 118–131 (1970)
59. Shiffrin, R.M., Gardner, G.T.: Visual processing capacity and attention control. J. Exp. Psychol. **93**(1), 72 (1972)
60. Duncan, J.: Selective attention and the organization of visual information. J. Exp. Psychol. Gen. **113**(4), 501–517 (1984)
61. Anderson, C.H., Van Essen, D.C., Olshausen, B.A.: Directed visual attention and the dynamic control of information flow. In: Itti, L., Rees, G., Tsotsos, J.K. (eds.) Neurobiology of Attention, pp. 11–17. Academic Press/Elservier, Burlington (2005)
62. Patel, V.L., Arocha, J.F., Zhang, J.: Thinking and reasoning in medicine. Cambridge Handb. Think. Reason. **14**, 727–750 (2005)
63. Alfaro-LeFevre, R.: Critical Thinking, Clinical Reasoning, and Clinical Judgment: A Practical Approach. Elsevier Saunders, Philadelphia (2013)
64. Rice, S.P., Boregowda, K., Williams, M.T., Morris, G.C., Okosieme, O.E.: A Welsh-sparing dysphasia. Lancet **382**(9904), 1608 (2013)
65. Croskerry, P.: The importance of cognitive errors in diagnosis and strategies to minimize them. Acad. Med. **78**(8), 775–780 (2003)
66. Berner, E.S., Graber, M.L.: Overconfidence as a cause of diagnostic error in medicine. Am. J. Med. **121**(5), S2–S23 (2008)
67. Shortliffe, E.H., Buchanan, B.G.: A model of inexact reasoning in medicine. Math. Biosci. **23**(3), 351–379 (1975)
68. Shortliffe, E.H., Buchanan, B.G.: Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading (1984)
69. Washington Technology. http://washingtontechnology.com/articles/2011/02/17/ibm-watson-next-steps.aspx

70. Stocker, C., Marzi, L.-M., Matula, C., Schantl, J., Prohaska, G., Brabenetz, A., Holzinger, A.: Enhancing patient safety through human-computer information retrieval on the example of german-speaking surgical reports. In: TIR 2014 - 11th International Workshop on Text-based Information Retrieval, pp. 1–5. IEEE (2014)
71. Simpson, E.H.: The interpretation of interaction in contingency tables. J. R. Stat. Soc. Ser. B-Stat. Methodol. **13**(2), 238–241 (1951)
72. Croskerry, P., Nimmo, G.: Better clinical decision making and reducing diagnostic error. J. R. Coll. Physicians Edinb. **41**(2), 155–162 (2011)
73. Djulbegovic, B., Hozo, I., Beckstead, J., Tsalatsanis, A., Pauker, S.: Dual processing model of medical decision-making. BMC Med. Inform. Decis. Mak. **12**(1), 94 (2012)
74. Huppertz, B., Holzinger, A.: Biobanks – A source of large biological data sets: open problems and future challenges. In: Holzinger, A., Jurisica, I. (eds.) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. LNCS, vol. 8401, pp. 317–330. Springer, Heidelberg (2014)
75. Downing, G., Boyle, S., Brinner, K., Osheroff, J.: Information management to enable personalized medicine: stakeholder roles in building clinical decision support. BMC Med. Inform. Decis. Mak. **9**(1), 1–11 (2009)
76. Rosenbloom, S.T., Denny, J.C., Xu, H., Lorenzi, N., Stead, W.W., Johnson, K.B.: Data from clinical notes: a perspective on the tension between structure and flexible documentation. J. Am. Med. Inf. Assoc. **18**(2), 181–186 (2011)
77. Lenfant, C.: Clinical research to clinical practice — lost in translation? N. Engl. J. Med. **349**(9), 868–874 (2003)
78. Love, J.S., Wright, A., Simon, S.R., Jenter, C.A., Soran, C.S., Volk, L.A., Bates, D.W., Poon, E.G.: Are physicians' perceptions of healthcare quality and practice satisfaction affected by errors associated with electronic health record use? J. Am. Med. Inf. Assoc. **19**(4), 610–614 (2011)
79. Bowman, S.: Impact of electronic health record systems on information integrity: quality and safety implications. Perspect. Health Inf. Manage. **10**, 1–19 (2013)
80. Morris, Z.S., Wooding, S., Grant, J.: The answer is 17 years, what is the question: understanding time lags in translational research. J. R. Soc. Med. **104**(12), 510–520 (2011)