

Chapter 1

Era of Intelligent Systems in Healthcare



Abstract The aim of this chapter is to prepare the reader for the outstanding trip that she/he embarked when starting reading this book. At first, we shall try to look for answers to some of the most important questions regarding the connection between intelligent systems and healthcare. What are intelligent systems? How can they be used in healthcare? Have they got benefits and prospects? Let us highlight some of the decisive factors for a successful deployment of intelligent systems in healthcare, including intelligent clinical support and intelligent patient management.

1.1 What Are Intelligent Systems?

Mother Nature is the greatest model architect that has ever existed. During its evolution, humanity has often believed that one can create ‘artificial’ models that are not influenced by the surrounding Nature. As time passed by, researchers realized that miming Nature might lead to better models. Thus their interest in this area increased, and problem solving reached another level. The new perspective in seeking inspiration from Nature, both from living systems and from natural sciences, such as physics or chemistry for instance, gave birth to the term “computational intelligence”. Having as starting point biology, researchers designed and developed artificial neural networks, evolutionary algorithms, swarm intelligence, fuzzy logic, etc. From physics or chemistry researchers modeled gravitational optimization, harmony search, chemical reaction optimization, gases Brownian motion optimization, etc.

So, what are intelligent systems (ISs)? This question is rather difficult to answer, as it raises several possible responses. Some say that an IS is a machine that has the capacity to gather and analyze data and communicate with other systems, that can learn from experience, and adapt according to current data, etc. These features can be summarized into: ISs are technologically advanced machines that perceive and respond to the environment.

To have a more comprehensive understanding of this concept, let us go back to the year 1950, and see what A.M. Turing said about the ‘thinking machines’. In Turing [1], Turing had to answer the following question “*Can machines think?*” Analyzing different aspects related to this ‘simple’ question, he concluded that the problem should be reformulated as follows: “*The new form of the problem can be described in terms of a game which we call the ‘imitation game’. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either “X is A and Y is B” or “X is B and Y is A”. The interrogator is allowed to put questions to A and B*”, with the answers, ideally, typewritten; the question is related to sex “*C: Will X please tell me the length of his or her hair?*” In this context, Turing brought up the issue: “*We now ask the question, ‘What will happen when a machine takes the part of A in this game?’ Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, ‘Can machines think?’*”. Summarizing the above, if we assume that we have a person, a machine, and an interrogator, then the object of the ‘imitation game’ for the interrogator is to determine which of the other two is the person, and which is the machine. However, the story does not end here, because “*The question we put ... will not be quite definite until we have specified what we mean by the word ‘machine.’*” Turing concludes that “*We are the more ready to do so in view of the fact that the present interest in ‘thinking machines’ has been aroused by a particular kind of machine, usually called an ‘electronic computer’ or ‘digital computer’*”. Following this suggestion we only permit digital computers to take part in our game.” Starting from this presentation, we come to the phrase “The Turing test”, as a way of answering the question whether machines can think or not. An interesting fact, according to the Stanford Encyclopedia of Philosophy [2], is the idea that the Turing’s test has deeper roots in history. Thus, it is suggested that the Turing test is foreshadowed in R. Descartes’ “*Discourse on the Method of Rightly Conducting one’s Reason and Seeking Truth in the Sciences*”, or, shortly “*Discourse on the Method* (1637)”. Finally, let us mention that there are currently opinions that say the classical Turing test is out of date and that we need to find other ways to measure the machines’ “intelligence” [3–6]. In this context, different approaches have emerged to assess the Turing test capacity to detect the presence of consciousness. We can mention the Chinese room argument [Internet Encyclopedia of Philosophy], which is an experiment of John Searle [7] designed to show that the Turing test is insufficient to detect the presence of consciousness, even if the ‘room’ can behave or function as a conscious mind would (Fig. 1.1).

One might think that ISs are one step away from the Turing test. If we consider that a computer is deemed to have Artificial Intelligence (AI) capabilities, if it can mimic human responses under specific conditions, everything is clear, but (AI) is more than a question of imitation, it is a matter of understanding. There are many ways to describe the concept of AI. According to Encyclopedia Britannica

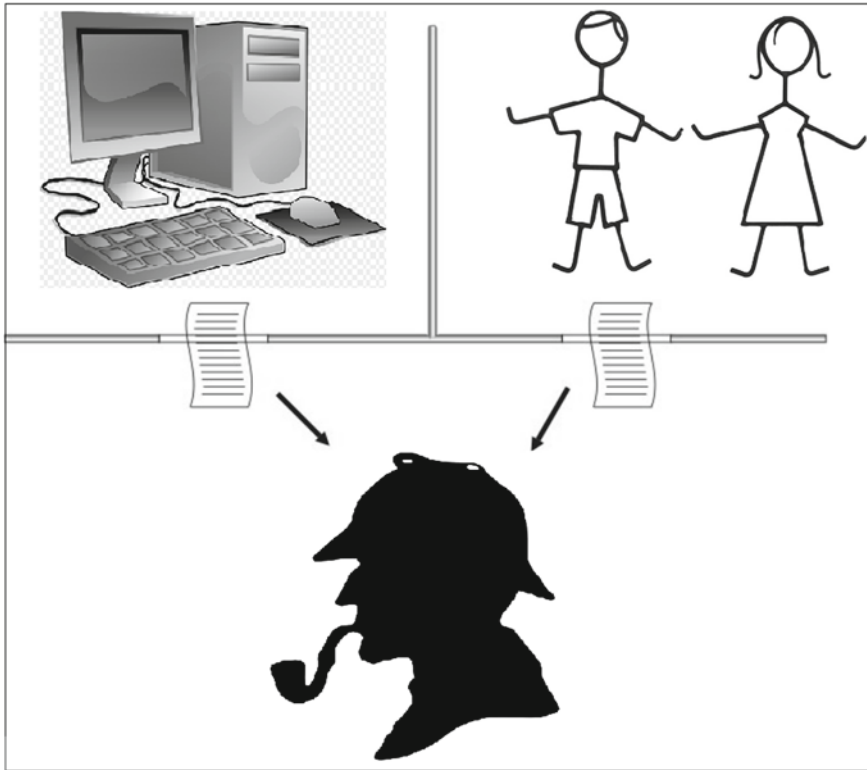


Fig. 1.1 The standard turing test

[Encyclopædia Britannica], AI means “the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings”, having thus the capabilities of “developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.” It is noteworthy that Turing has provided the earliest significant work regarding AI, the first manifesto, all modern computers being considered as “Turing machines”. In an unpublished 1948 report, entitled “Intelligent Machinery”, he introduced many of the central concepts of AI. More details about A. Turing and his substantial contribution to AI, including the report mentioned above, are to be found in Copeland [8].

After this brief history about the beginnings of AI, let us return to our main story regarding ISs. Consistent with the definition of ISs presented at the start, let us also point out that ISs can be found in different forms: from AI models processing huge datasets to AI models controlling robots. The ISs field represents an interdisciplinary research domain bringing together ideas from AI, machine learning (ML), and a range of fields such as psychology, linguistics and brain sciences, connected by many interdisciplinary relationships (Fig. 1.2).



Fig. 1.2 “Intelligent systems”

Nowadays, a wide variety of ISs have been developed, such as:

- expert systems,
- fuzzy systems,
- artificial neural networks,
- evolutionary computation (genetic/evolutionary algorithms, genetic programming, evolutionary strategies),
- support vector machines,
- particle swarm optimization,
- ant colony systems,
- memetic algorithms,
- ant colony optimization,
- clustering,
- Bayesian (learning) model,
- deep learning,
- hybrid models. (neuro-genetic, neuro-fuzzy, fuzzy-genetic, etc.)

The general areas of applications of modern ISs include the following topics:

- Artificial Intelligence
- Pattern Recognition
- Artificial Neural Networks
- Decision Support Systems
- Supervised Semi-Supervised and Unsupervised Learning
- Human-Computer Interaction and Systems

- Soft Computing and Fuzzy Systems
- Computer Vision and Active Vision
- Image and Video Processing and Analysis
- Medical Imaging
- Machine Learning (ML)
- Data Mining (DM)
- Evolutionary algorithms
- Swarm Intelligence
- Ontologies
- Speech Generation and Recognition
- Intelligent Transportation Systems
- Models and Computational Theories of Human Cognition
- Intelligent Robotic Systems
- Knowledge Management and Representation
- Intelligent System Design
- Bayesian Learning
- Evolving Clustering Methods
- Natural Language Processing and Fusion

More details related to ISs in Padhy [9], Shin and Xu [10], Hopgood [11], Grosan and Abraham [12], Wilamowski and David Irwin [13], Pap [14], Kryszkiewicz et al. [15], Martínez de Pisón et al. [16], Bi et al. [17].

With regard to the practical applications of ISs, there is a huge collection of real situations where they show their potential to effectively solve the problems that arise. In what follows, we will briefly address some well-known real-world applications of ISs, so that the reader can get an idea of the potential of these “smart tools” and their application area.

- Fuzzy logic control [18–22];
- Business Intelligence/Management Intelligent Systems [23–26];
- Intelligent Bioinformatics Systems [27–31];
- Intelligent Healthcare Systems [32–36];
- Intelligent Game [37–41];
- Intelligent Multimedia [42–46];
- User Interfaces and Human Computer Interaction [47–50];
- Knowledge-based Software Engineering/ Knowledge Engineering/ Management [51–55];
- Speech recognition [56–60];
- Brain-Machine Interface Systems [61–65];
- Intelligent Robotic Systems [66–70];
- Intelligent Transportation Systems [71–73];
- Medical Imaging [74–78];
- Psychology [79–82, 160];
- Military applications [83–86, 161];
- Engineering problems [87–90, 162];

- Smart cities [91–95];
- Internet of Things (IoT) [96–100].

Besides presenting some basic theoretical and practical concepts of the most known ISs applications, we believe that is far more interesting for the reader to see their impact on everyday life. Basically, we will talk now about the well-known intelligent devices. An intelligent/smart device is any type of equipment, instrument, or machine that has its own computing capability. Moreover, these devices are, generally, connected to other devices via wireless networks. Most of them are now parts of the so-called “smart home”. In this context, the IoT describes how the Internet will link intelligent devices, to allow these endpoints to generate and share data. Examples of intelligent devices are smartphones, phablets, tablets, smart watches, wearable computers, smart home devices (e.g., smart thermostat, smart lighting bulbs, smart security cameras, smart switch, smart speakers, smart AC control, etc.), intelligent vehicle technologies, etc. (Figs. 1.3, 1.4, 1.5).

Besides these everyday in use smart devices, let us not forget to mention the famous virtual personal assistants: Siri, Google Assistant, and Cortana—the intelligent digital personal assistants on various platforms (e.g., iOS, Android, and Windows Mobile)—that help us find useful information when vocally asked (Fig. 1.6).

It is interesting to go through a brief history of AI/IS to see the impressive rhythm of development of this important field of Computer Science in just 60 years from its beginning [101]:

- One considers that AI started at “Dartmouth Artificial Intelligence” conference, held at Dartmouth College, Hanover, New Hampshire, USA, in 1956 (https://www.livinginternet.com/i/ii_ai.htm—accessed November 24, 2017). In this regard, let us mention the “*Dartmouth AI Project Proposal*”, by J. McCarthy (Dartmouth College), M. L. Minsky (Harvard University), N. Rochester (IBM Corporation), and C.E. Shannon (Bell Telephone Laboratories)—August 31, 1955, states that “*We propose that a 2 month, 10 man study of artificial*



Fig. 1.3 Smart devices



Fig. 1.4 Smart home

intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire". It is noteworthy that the term "Artificial Intelligence" was first coined by John McCarthy.

(<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> accessed November 24, 2017).

- A milestone for the use of AI in real-world applications was the idea of creating a 'perfect machine translator', funded by the *Defense Advanced Research Projects Agency* (DARPA). Unfortunately, two reports (Automatic Language Processing Advisory Committee (ALPAC) report—1966, and Lighthill report—1973) "concluded negatively about the possibility of creating a machine that could learn or be considered intelligent".
- Another landmark in AI history was the occurrence of the "expert systems" in the eighties, giving a 'breather' to the field by UK and Japan funding. It is noteworthy that expert systems first emerged in the early 1950s when the Rand-Carnegie team of Newell, Shaw, and Simon developed a General Problem Solver that dealt with theorems proof, geometric problems and chess playing.
- Another revival of the AI field began in 1993 with the MIT "Cog" project, aiming to build a humanoid robot, and with the Dynamic Analysis and Replanning Tool (DART), aiming to optimize and schedule transportation of supplies or personnel and solve other logistical problems, used by U.S. military.

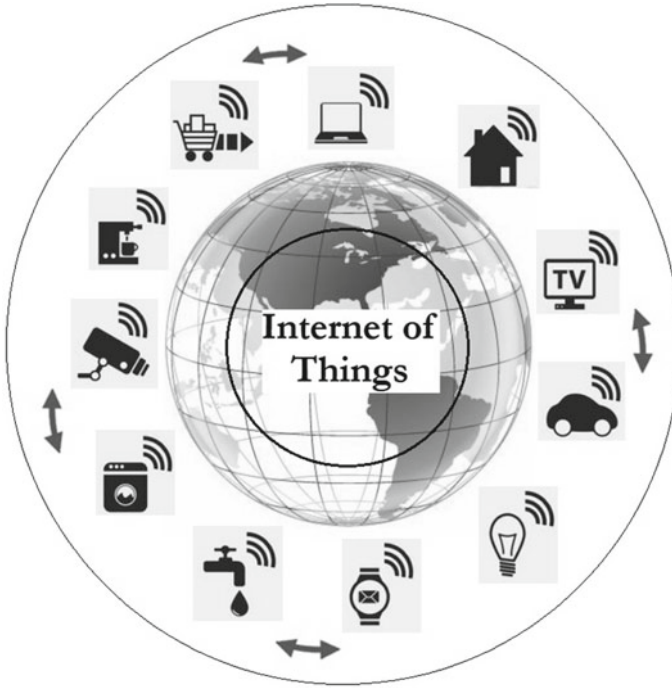


Fig. 1.5 Internet of Things (IoT)

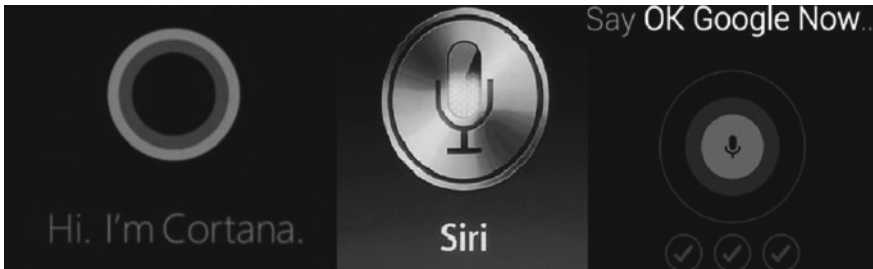


Fig. 1.6 Digital personal assistants

- We now reach February 10, 1996, when “Deep Blue”, the chess-playing super computer developed by IBM, won its first game against a world champion, Garry Kasparov. Overall, Kasparov defeated Deep Blue by a score of 4–2. However, in 1997, after a heavily upgrade, Deep Blue defeated Kasparov in a rematch, the first defeat of a reigning world chess champion by a computer under tournament conditions (score 3.5/2.5). It is noteworthy that C.E. Shannon was the first to think about developing a chess-playing program.

- Then, until 2012, the AI development has materialized mainly through academic research, with no significant practical impact. The milestone was December 4, 2012, when a convolutional neural network was presented at the Neural Information Processing Systems—NIPS 2012 Conference (Advances in Neural Information Processing Systems 25). Since then, the AI development trend has been steadily increasing.

The field of AI went through phases of rapid progress and failures, followed by a cooling in investment and interest often referred to as “AI winters” (inspired by the “nuclear winter”). The first frost occurred in the 1970s, as progress slowed and government funding dried up, and another one in the 1980s because of the failure of the expected commercial impact.

It is interesting to mention an important fact regarding the place and the moment of the occurrence of AI, beyond the official chronology. Thus, according to H. Bruderer (<https://cacm.acm.org/blogs/blog-cacm/222486-the-birthplace-of-artificial-intelligence/fulltext>—accessed November 24, 2017), the birthplace of AI would have been Paris, on the occasion of the Colloques internationaux du Centre national de la recherche scientifique, Paris, Janvier 8-13, 1951, No. 37, Editions du Centre national de la recherche scientifique (CNRS) (Ross, A.W. Ed.), Paris 1953, 475–492, focused on “*Les machines à calculer et la pensée humaine*”. It is noteworthy that, within this colloquium, leading scientists have participated (e.g., N. Wiener, W.S. McCulloch, A.H. Aiken, F.M. Colebrook, etc.), and some functional devices, such as a chess automaton and analog calculator, and the artificial animals (some of the first autonomous robots) have been presented. Under these circumstances, the question regarding the AI birthplace and birth date remains open.

Regarding the companies which operates in the field of computers and computer software/technologies, in the list for the year 2017 of FORTUNE Global 500 largest companies, © 2017 Time Inc. (<http://fortune.com/global500/list/>—accessed November 24, 2017), we have the following situation displayed in Table 1.1.

The World’s biggest public companies (2017 ranking), operating in the Software & Programming field, according to Forbes 2017 GLOBAL World’s Biggest Public Companies 2000, © 2017 Forbes Media LLC, are the following ones: (<https://www.forbes.com/global2000/list/>—accessed November 24, 2017) Table 1.2.

AI is a research field so vast that we cannot keep track of everything. Among the leading global publishers of AI journals (books) that serve and support the research community, we can quote Springer, Elsevier, Wiley, IEEE, ACM, etc. Because of the substantial significance related to the AI field, both from theoretical and

Table 1.1 FORTUNE Global 500 largest companies

Rank (#)	Revenue (\$M)
9. Apple	\$215,639
69. Microsoft	\$85,320
124. Dell Technologies	\$64,806
280. Oracle	\$37,047
443. SAP	\$24,397

Table 1.2 2017 world ranking—Software & Programming Companies

• #19 Microsoft (USA);	• #70 Oracle (USA);
• #178 SAP (Germany);	• #645 VMware (USA);
• #701 Adobe Systems (USA);	• #857 Symantec;
• #929 Salesforce.com (USA);	• #958 HCL Technologies (India);
• #972 Fiserv (USA);	• #1042 Intuit (USA);
• #1045 Amadeus IT Group (Spain);	• #1155 CDW (USA);
• #1227 CA (USA);	• #1251 Check Point Software (Israel);
• #1321 Dassault Systemes (France);	• #1405 Open Text (Canada);
• #1461 Citrix Systems (USA);	• #1694 Snap (USA);
• #1809 Autodesk (USA);	• #1811 VeriSign (USA);
• #1881 Workday (USA);	• #1907 Amdocs (USA);
• #1935 Red Hat (USA);	• #1967 ServiceNow (USA)

real-world applications point of view, there are thousands of scientific journals currently in publication. Since it is almost impossible to make an objective evaluation (ranking) of the most prestigious international journals in this area (several journal-level metrics being proposed, according to Wikipedia/Journal ranking: https://en.wikipedia.org/wiki/Journal_ranking—accessed November 24, 2017), we have just listed below some of the most popular, and the most influential, currently publishing journals, taking into account, of course, these ranking measures as well.

- IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION (IEEE Computational Intelligence Society);
- IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (IEEE Computer Society);
- INTERNATIONAL JOURNAL OF COMPUTER VISION (Springer);
- IEEE TRANSACTIONS ON FUZZY SYSTEMS (IEEE Computational Intelligence Society);
- IEEE TRANSACTIONS ON CYBERNETICS (IEEE Systems, Man, and Cybernetics Society);
- IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE (IEEE Computational Intelligence Society);
- INTERNATIONAL JOURNAL OF NEURAL SYSTEMS (World Scientific);
- IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (IEEE Computational Intelligence Society);
- NEURAL NETWORKS (Elsevier);
- JOURNAL OF MACHINE LEARNING RESEARCH (JMLR, Inc. and Microtome Publishing);
- ARTIFICIAL INTELLIGENCE (Elsevier);
- PATTERN RECOGNITION (Elsevier);
- KNOWLEDGE-BASED SYSTEMS (Elsevier);
- EXPERT SYSTEMS WITH APPLICATIONS (Elsevier);

- SWARM AND EVOLUTIONARY COMPUTATION (Elsevier);
- EVOLUTIONARY COMPUTATION (MIT Press);
- APPLIED SOFT COMPUTING (Elsevier);
- NEUROCOMPUTING (Elsevier);
- ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY (ACM New York);
- JOURNAL OF THE ACM (ACM);
- DATA MINING AND KNOWLEDGE DISCOVERY (Springer);
- INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS (Wiley);
- ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE (Elsevier);
- ARTIFICIAL INTELLIGENCE REVIEW (Springer);
- JOURNAL OF INTELLIGENT INFORMATION SYSTEMS (Springer);
- JOURNAL OF INTELLIGENT & ROBOTIC SYSTEMS (Springer);
- COMPUTATIONAL INTELLIGENCE (Wiley);
- EXPERT SYSTEMS (Wiley).

Since in the past and, mainly, in the recent years, we have been experiencing an exponential growth of the AI/IS development, every year many prestigious international conferences, covering the most diverse aspects of the field, take place in the most diverse locations worldwide. As examples of such recent events, we mention the following ones:

- IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, Nov 03–08, 2019;
- IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, June 23–26, 2019;
- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, June 15–20, 2019;
- IEEE Intelligent Systems Conference (IntelliSys), United Kingdom, 15–16 November 2018;
- The 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC2018), Miyazaki, Japan, October 7–10, 2018;
- IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Spain, October 1–5, 2018;
- IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, July 8–13, 2018;
- The 19th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2018), Haeundae-gu, Busan, Korea (South), June 27–29, 2018;
- IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Rhodes, Greece, May 25–27, 2018;
- ACM 2018 International Conference on Computing and Artificial Intelligence (ICCAI 2018), Chengdu, China, March 12–14, 2018;

- AAAI Conference on Artificial Intelligence, New Orleans, USA, February 2–7, 2018;
- The 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Boston, USA, November 6–8, 2017;
- IEEE Intelligent Systems Conference 2017 (IntelliSys 2017), London, UK, September 7–8, 2017;
- ACM SIGKDD International Conference on Knowledge discovery and Data Mining, Halifax, Canada, August 13–17, 2017;
- International Conference on Machine Learning and Data Mining (MLDM 2017), New York, NY, July 15–20, 2017;
- The 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Kanazawa, Ishikawa, Japan, June 26–28, 2017;
- IEEE International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, June 15–16, 2017;
- AI and Machine Learning World 2017, London, UK, June 13–15, 2017;
- Applied Artificial Intelligence Conference 2017, San Francisco, CA, May 11, 2017;
- AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, CA, February 4–9, 2017;

Bill Gates, the principal founder of Microsoft, published in 1999 a book (Gates 1999/2001), titled “*Business @ the Speed of Thought*”, in which he made 15 important predictions regarding the future developments of the information technologies. The book contains five chapters and addresses themes like: ‘Information Flow is Your Lifeblood’, ‘Commerce: The Internet Changes Everything’, ‘Manage Knowledge to Improve Strategic Thought’, ‘Special Projects (*No health-care system is an island, Take government to the people, Create connected learning communities*)’, ‘Expect the Unexpected (*Prepare for the digital future*)’. The 15 predictions that Bill Gates wrote in this book can be validated today, after nearly 20 years. We listed below the topics of the predictions of B. Gates, highlighted by World Economic Forum/Business Insider (07 July 2017, Carson, B., Tech reporter —Business Insider, © 2017 World Economic Forum):

- Price-comparison sites;
- Mobile devices;
- Instant payments and financing online, and better healthcare through the web;
- Personal assistants and the internet of things;
- Online home-monitoring;
- Social media;
- Automated promotional offers;
- Live sports discussion sites;
- Smart advertising;
- Links to sites during live TV;
- Online discussion boards;

- Interest-based online sites;
 - Project-management software;
 - Online recruiting;
 - Business community software.
- (<https://www.weforum.org/agenda/2017/07/these-are-the-15-predictions-bill-gates-made-in-1999>—accessed November 24, 2017).

The impetuous development that AI has known, led to the emergence of an intensely debated issue today by well-known scientists as well as ordinary people. The question is: *What scares us when we see the unprecedented development of AI?* Most people worry about the occurrence of the “singularity point”, that is the milestone in history when AI will surpass human intelligence and its consequences. Below we list the opinions of some famous people in this area.

- In “*The Independent*” (May 1, 2014), S. Hawking (Cambridge University), S. Russell (University of California, Berkeley), M. Tegmark (MIT), and F. Wilczek (MIT, 2004 Nobel laureate) talk about the huge potential benefits of AI, but also about the potential jeopardy of wrongly using it. (<http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html>—accessed November 24, 2017);
- The “Tesla” and SpaceX CEO, Elon Musk, renewed his critique of AI, at CNBC (August 11, 2017, J. David—publisher): (<https://www.cnn.com/2017/08/11/elon-musk-issues-a-stark-warning-about-a-i-calls-it-a-bigger-threat-than-north-korea.html>—accessed November 24, 2017);
- and “*The Guardian*” (July 17, 2017. S. Gibbs—publisher): (<https://www.theguardian.com/technology/2017/jul/17/elon-musk-regulation-ai-combat-existential-threat-tesla-spacex-ceo>—accessed November 24, 2017)

To conclude, the AI’s existential risk resides in the possibility that substantial progress in AI could someday result in human extinction, or some other unrecoverable global catastrophe (see also *Wikipedia*: “Existential risk from artificial general intelligence”) (https://en.wikipedia.org/wiki/Existential_risk_from_artificial_general_intelligence#Reactions—accessed November 24, 2017)

1.2 Why and How Intelligent Systems in Healthcare?

Across all the research fields, from engineering, meteorology, and business to healthcare, sociology, and multimedia and so forth, data are being collected and accumulated at a very fast pace. Under these circumstances, there is an urgent need for developing advanced ISs to assist humans in extracting useful information/knowledge from the huge volumes of digital data in order to make real-time accurate decisions. As it is well known, healthcare deals with thorough processes of the diagnosis, treatment and prevention of disease, injury, physical, and mental

impairments. At the same time, it also covers the hospital/patient management. As it rapidly evolved in most countries, the healthcare industry has become the generator of a massive amount of data, including electronic medical records, administrative reports, and other useful knowledge. Figures 1.7 and 1.8 provide a synthetic picture of the way ISs assist and, especially help optimizing the healthcare process, both in terms of computerized/automated medical diagnosis and intelligent patient management.

The healthcare domain includes many industries and companies that are involved in products and services related to health. Among the most important branches of healthcare, one can mention:

- *Pharmaceutical industry*, which is the part of the healthcare sector that deals with medicines. It comprises different fields pertaining to the discovery, development, production, and marketing of pharmaceutical drugs. The pharmaceutical industry worldwide revenue was *grosso modo* estimated over one trillion U.S. dollars in 2014 (Berkrot, B., *Reuters*, #Health News, April 20, 2010).

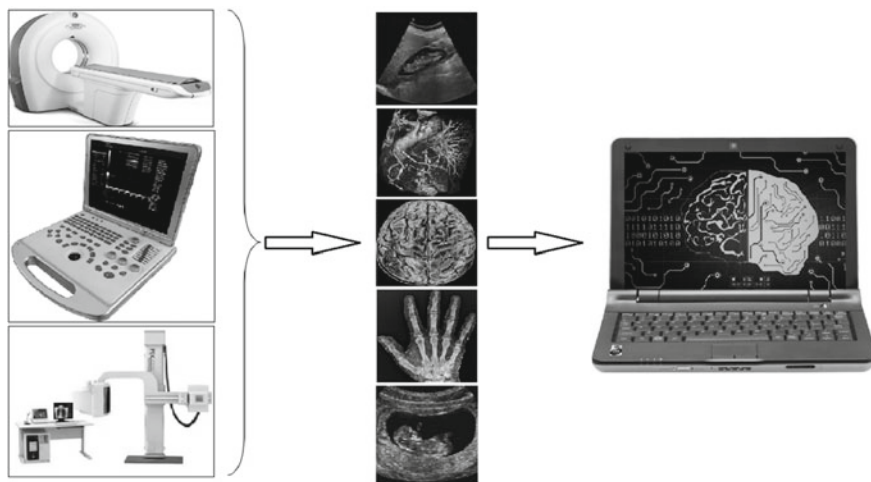


Fig. 1.7 Computer-aided medical diagnosis



Fig. 1.8 Intelligent patient management

- *Biotechnology industry*, which develops, manufactures, and markets novel, patented medicines, basically using the gene engineering technology, For instance, the medical biotechnology revenues exceed \$150 billion annually (Diehl, P., *Biotech Industry*, June 26, 2017). (<https://www.thebalance.com/what-is-biotechnology-375612>—accessed November 24, 2017);
- *Healthcare equipment industry*, which consists of manufacturers of healthcare equipment and medical devices, creating diverse products, such as medical instruments, drug delivery systems, cardiovascular, orthopedic devices, diagnostic equipment, etc. The great importance of this segment is reflected by the fact that the global (durable) medical equipment market is expected to reach \$242.1 billion by 2024, according to GlobeNewswire (Nasdaq), © 2017 GlobeNewswire, Inc. September 6, 2016 (<https://globenewswire.com/>—accessed November 24, 2017), citing a Grand View Research, Inc., © 2017 Grand View Research, Inc.: Market Research Reports & Consulting report. (<http://www.grandviewresearch.com/>—accessed November 24, 2017);
- *Healthcare distribution industry*, which represents an essential part of the healthcare domain, consists of all distributors and wholesalers of healthcare products. To get an idea about the scale of the healthcare distribution market, a simple glance over the pharmaceutical distribution shows that, in 2015, pharmaceutical full-line wholesalers generated a total turnover of €141 billion in the EU (Swiss and Norway included) [102], © Institute for Pharmaeconomic Research, 2017. (<http://www.girp.eu/files/GIRP-IPF%20Study%202016.pdf>—accessed November 24, 2017);
- *Healthcare facilities*, which consist of all types of services the healthcare system offers to patients. These include hospitals, clinics, outpatient care centers, specialized care centers (e.g., birthing centers and psychiatric care centers), etc. The healthcare facilities industry includes the companies engaged in operating hospitals, physician and dental clinics, psychiatric facilities, nursing homes and home healthcare agencies, different healthcare laboratories, etc. For example, if we consider only the medical and diagnostic laboratories in the U.S., then, according to the U.S. Bureau of the Census [103], the total revenue for Q3 2017 has been estimated at \$12.564 billion. (<https://fred.stlouisfed.org/series/REV6215TAXABL144QNSA>—accessed November 24, 2017);
- *Managed healthcare*, which consists of diverse activities, intended to reduce the cost of providing health benefits and improve the quality of healthcare. The managed healthcare plans represent a type of health insurance dealing with contracts with healthcare providers and medical facilities to provide care for members at reduced costs—see U.S. National Library of Medicine (NLM/NIH) (<https://medlineplus.gov/managedcare.html>—accessed November 24, 2017). Because the rate of people to be insured is rising, managed healthcare has a great opportunity of growth. For example, according to the annual financial reports of the Centers for Medicare and Medicaid Services (CMS)—Department of Health and Human Services (HHS) (<https://www.hhs.gov/>), CMS has outlays of approximately \$993.9 billion in fiscal year FY 2017 (net of offsetting receipts

and Payments of the Healthcare Trust Funds). (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CFOReport/Downloads/2017_CMS_FinancialReport.pdf—accessed November 24, 2017)

After this short overview of the current ‘landscape’ in this field, we now focus on the “revolution” brought by AI in healthcare. Let us first take a brief look on the history of using AI tools in medicine.

Let us recall the DENDRAL project (coming from “DENDRitic ALgorithm”), also known as the “grandfather of experts systems”, started at Stanford University, California, in 1965 by Edward Feigenbaum, Bruce G. Buchanan, Joshua Lederberg, and Carl Djerassi, along with a team of research associates and students [104], that was initially intended as a chemical-analysis expert system.

MYCIN, developed by Edward Shortliffe (1972) as a doctoral dissertation for Stanford Medical School [105], was a clinical expert system, derived from DENDRAL, and designed for selection of antibiotics for patients with serious blood infections, hence the name “mycin”—suffix of many antibiotics. Mycin consisted of three components (sub-systems): (a) consultation system, (b) explanation system, and (c) rule acquisition system. It is noteworthy that MYCIN was never publicly used in clinical practice.

Other “intelligent” systems have been proposed in those years in an attempt to support the diagnosis of various diseases. Of the most well-known, we would mention: AAPhelp (AAP = acute abdominal pain) (Leeds University, UK) for the diagnosis of the acute abdominal pain, INTERNIST (University of Pittsburg) for complex diagnosis in general internal medicine, CASNET/glaucoma (CASNET = Causal-Associational Network) (Rutgers University) for the diagnosis and treatment of glaucoma, PIP (= Present Illness Program) (MIT and Tufts-New England Medical Center) for diagnosis of kidney diseases, ABEL (= Acid-Base and Electrolyte Program) (MIT laboratory for Computer Science) for the management of electrolyte and acid base derangements), ONCOCIN (Stanford University School of Medicine) for the treatment of cancer patients receiving chemotherapy, DXplain (Massachusetts General Hospital) for ranking a list of diagnoses based on some clinical findings, etc.

To see how AI can be an important player in the healthcare domain, we need to get a first look inside the physician’s activity.

Let us first briefly analyze the interaction patient-physician. This complex interaction includes empathy, information management, application of expertise in a given context, negotiation with multiple stakeholders, and unpredictable physical response in special situations, such as surgery or post-op, some actions taking place in real time (for instance, patient on the operating table or patient in the ICU (Intensive Care Unit)). In this context, we are wondering: “*Are these important real-life aspects viable AI-applications or not?*” Obviously, these are not AI-applicable functions. Thus, according to a study regarding the way a physician spends his/her time at work (U.S. ambulatory care in four specialties in four states) [106], it resulted that physicians spent 27.0% of their total time on direct clinical face time with patients and 49.2% of their time on EHR (Electronic Health Record),

and desk work during the office day. Under these circumstances, physicians have less time to directly practice medicine, do necessary research, master new medical/AI technology, and improve their skills in order to become better doctors. To conclude, apart from the measures that need to be taken to increase the doctor-patient interaction time against the clerical work, it will be important for physicians and patients to understand and engage the evolution of automation in medicine in order to optimize the patient care. Physicians must be open to the rapidly advancing AI technology, and, more importantly, they should embrace this opportunity rather than fear it. In this sense, there are many concerns of the medical staff on this subject, due to recent news in the media. Thus, in the near future, robots and AI technologies in healthcare could lead to a doctorless hospital [107]. Although hospitals have been slow in adopting AI technologies and robotics, these new approaches took their place gradually, fundamentally changing the landscape. Thus, doctors in near future will have to have new skills vs. today's doctors (e.g., computer skills, robotics), to successfully compete with the new technologies. Let us hope that we will not turn the current hospitals into robot-factory hospitals. It is noteworthy to mention in this sense, that financial pressures will inevitably force the recognition of the fact that the medical robots, powered by the AI technology, will be the only way to help doctors provide a significantly higher medical service than the current one. If driverless cars are going to reduce traffic accidents and congestion, but under human supervision, then we may hope that the intelligent systems grafted in the hospitals of the future will one day save more lives and reduce the cost of healthcare, without removing doctors from the medical activities and decisions (Fig. 1.9).

To conclude this brief glance over a possible scenario of future hospitals' functioning, let us mention the opinion of Naveen Jain, an entrepreneur and philanthropist driven to solve the world's biggest challenges through innovation, founder/CEO of some successful companies, such as Moon Express, BlueDot, Intelius, Talent Wise, Viome, InfoSpace. In a CNBC TV interview (published by Kharpal, A, Fri, 1 Dec 2017, 1:13 AM ET) from the Slush technology conference in Helsinki, Finland, 2017 (<http://www.slush.org/>), where he spoke about the contribution of AI in healthcare, comparing it with a "tsunami", because of its



Fig. 1.9 Robot-factory hospital

considerable role in processing huge amount of data, impossible to be done by humans. The latest AI implications and achievements in healthcare come to justify this visionary opinion. Let us recall, in this regard, only that a Chinese AI-powered robot (Xiaoyi, meaning “little doctor”) became the world’s first machine to pass successfully a medical exam on Nov. 6, 2017 (posted by Mara P. on November 24, 2017, © 2016 TechThe Lead—<http://techthelead.com/robot-chinas-medical-exam/>).

We have seen from the above facts how important AI’s involvement in healthcare can become. There are current discussions about the “Fourth Industrial Revolution” [108], involving new cutting-edge technologies (AI technologies, intelligent robots, driverless cars, nanotechnologies, quantum computing, IoT, etc.), which will affect various domains such as healthcare, economy, and industry, etc. The basic idea is that enormous databases have been collected in recent years, and for the year 2020 being foreseen, for instance, an amount of data equaling about 44 zettabytes (44 trillion gigabytes), according to an IDC Digital Universe study (IDC Analyze the Future), Digital Universe “Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East” (sponsored by EMC², DellEMC), © 2017 KDnuggets (<https://www.kdnuggets.com/2012/12/idc-digital-universe-2020.html>)

As mentioned above, we dispose nowadays of a huge amount of data, from the medical field in our case, which have to be processed somehow to discover useful, often hidden, information in order to make an optimal decision in real time. Let us recall, in this respect, the well-known *UCI Machine Learning Repository*, a collection of databases used by the ML community for the empirical analysis of ML algorithms (<https://archive.ics.uci.edu/ml/datasets.html>). We also recall the *Kaggle datasets* (<https://www.kaggle.com/datasets>), maintaining important datasets as a service to the ML community, the platform for researchers to share their data *Academic Torrents* (<http://academictorrents.com/>) or the dataset repository from *GitHub* (<https://github.com/awesomedata/awesome-public-datasets/blob/master/README.rst>—<https://github.com/awesomedata/awesome-public-datasets/blob/master/README.rst>). Therefore, there are many public or private databases that can be used for different research purposes, e.g., diagnoses, medical procedures, medication, demographics, cost/charge of healthcare services, etc. There are also the efficacious tools provided by AI to process this data. We are, metaphorically speaking, in the situation of Ali Baba in the cave of the forty thieves, searching for valuable information helping to raise the healthcare quality.

In order to better understand the “treasure” of information/knowledge contained in these healthcare databases, let us restrict ourselves only to those concerning medical diagnosis. The huge size datasets containing both the symptoms and the correct diagnosis of a particular illness represent the collection of data from hundreds or thousands of physicians, data gathered over many years. When making a particular diagnosis, the physician processes only the data recorded in his/her memory, data gathered throughout his/her whole career up to that point. The effectiveness and speed of diagnosis-making differ significantly from a resident with a few years of practice under the supervision of an attending physician to an experienced doctor with many years of medical practice, in which he/she has seen numerous cases. How can AI support the decision-making process in this context?

Very simple: by using the powerful data analysis and decision-making mechanisms (the classification technology, in short) applied to the “experience” of hundreds or thousands of doctors, through the corresponding databases, in a blink of an eye. Metaphorically speaking, this computerized data processing is the synergy of information provided by all doctors who have collected data over time and not just a second medical opinion.

Finally, yet importantly, let us talk about the well-known “*curse of dimensionality*”, Bellman [109]. Nowadays, we dispose of very sophisticated and advanced medical equipment, making thus possible the collection of a huge amount of features, more or less important in the medical diagnosis process or patient management. We may have to process datasets with a relatively small number of attributes (e.g., the thyroid disease dataset with 21 attributes from *UCI Machine Learning Repository*—<http://archive.ics.uci.edu/ml/datasets/thyroid+disease>). On the other hand, we have to deal with high dimension datasets containing medical data obtained through DNA microarray analysis (e.g., 24,481 attributes, as in the case of the breast cancer Kent Ridge dataset from *Machine Learning Data Set Repository* (<http://mldata.org/>)). Thus, we have to tackle datasets with a huge number of variables (attributes), therefore becoming “suffocated” by the richness of information. Who gets to decide objectively and effectively which features are important and which are redundant, even in the case of 21 attributes? The answer is very simple: not the human, but the “machine” using its “artificial intelligence” to perform the so-called feature selection and feature extraction techniques.

We have tried above to outline the role of ISs in the medical field, thus answering the question “*Why ISs in healthcare?*” We will now present some of the most important ways in which ISs are involved in this domain, answering the question “*How ISs in healthcare?*” Thus, according to the IBM response to the White House Office of Science and Technology Policy’s Request for information (Summer 2016), IBM [110], systems can advance precision medicine by ingesting patients’ electronic medical history and relevant medical literature, performing cohort analysis, identifying micro-segments of similar patients, evaluating standard-of-care practices and available treatment options, ranking by relevance, risk and preference, and ultimately recommending the most effective treatments for their patients.”

When we think about how ISs are involved in the healthcare decision-making, two approaches that work in tandem come to mind. Firstly, the *natural language processing* technology deals with the information extraction from unstructured data (e.g., clinical notes, medical journals, etc.) in order to improve the machine-readable structured medical data. Secondly, using both structured data contained in medical datasets, or unstructured data ‘translated’ by NLP, cutting-edge *ML/DM algorithms* attempt to extract valuable information/knowledge in order to support efficiently the human decision-making. Let us say a few words about each approach.

- *The Natural Language Processing (NLP) Technology*. Starting with the Turing’s test to assess the machines’ intelligence”, the interaction between human and computers through the natural languages has remained a very important and topical issue even today. In the context of healthcare supported by ISs/AI, we are

interested in how to apply NLP to unstructured data. By unstructured data, or unstructured information, we broadly understand information that either does not have a pre-defined data model and/or does not fit well into relational tables. We can mention as unstructured data: health records, documents, images, audio/video files, analog data, sentiments about a given topic, etc. In the healthcare field, it is very important to have direct information about data or, indirectly, data about data, that is, metadata. For instance, grammar (*syntax*) is metadata. Both structured languages (e.g., Java), and unstructured languages (e.g., the English language) have grammars, describing them (data about data) by clarifying, for example, the relations between words in a sentence. On the other hand, we may be interested by the meaning/interpretation of words or groups of words within a certain medical narrative context, in order to better understand some symptoms declared by the patient (*semantics*). For more details regarding natural language understanding of unstructured data, see Trim [111]. There are different NLP techniques for extracting information/knowledge from unstructured data, depending on their type (e.g., texts, images, audio, sentiments, etc.). In the healthcare case, there are both structured and unstructured data, clinical data, for instance, being either in structured or unstructured form. There are clinical data presented in some templates (e.g., tables), i.e., structured data, which can be processed directly by different ML techniques. There are also unstructured clinical data in the form of free text narratives (e.g., observations and/or thoughts regarding the patient, obtained during the doctor-patient dialogue) that can only be addressed with the NLP techniques [92]. The traditional approach to support clinical decision-making by NLP in the case of free text narratives is a two-fold process. Firstly, there is text processing enabling the discovery of disease-relevant keywords in the unstructured data. Secondly, the validated keywords through a sensitivity analysis focused on their effects on the classification of the normal and abnormal cases are used to supplement the available structured data. A well-known example of a computer system, which has been directly applied in NLP for healthcare, is the famous “IBM Watson” (<https://www.ibm.com/watson/>). IBM announced in 2013 the application of IBM Watson at Memorial Sloan Kettering Cancer Center, New York City. *Watson Oncology* is a cognitive computing system designed to support the oncology community of physicians as they consider treatment options with their patients. Concretely, it interprets cancer patients’ clinical information and identifies individualized, evidence-based treatment options (<https://www.mskcc.org/about/innovative-collaborations/watson-oncology>). Let us also mention, in this context, the natural language application-programming interface (API) from Google—*Google Cloud Natural Language* (<https://cloud.google.com/natural-language/>). Of the most popular NLP applications we mention just the following ones:

- *Medicine*, supporting physicians to extract and summarize information of any symptoms, drug dosage, and response data. Consequently, one can identify possible side effects of any medicine while highlighting or flagging significant items in data.

- *Machine translation*, with the focus on keeping the meaning of sentences intact along with grammar and tenses.
 - *Text categorization*, used for assignation of different documents to predefined categories or indices.
 - *E-mail spam filtering*, using a set of protocols to determine which of the incoming messages are spams and which are not.
 - *Information extraction*, dealing with the extraction of entities, such as names, places, events, dates, times and prices, increasing thus both accuracy and efficiency of a directed search.
 - *Summarization of information*, used to summarize huge amount of the data while keeping the meaning intact.
- *Classical and Modern ML/DM Algorithms*. ML is a very important topic of Computer Science that study theories, algorithms, and related real-life applications that give computers the ability to learn like humans, in other words, acting without being explicitly programmed. While DM can be seen as the science of exploring big data (large or even huge datasets) for extracting implicit, previously unknown and potentially useful information, ML represents the underlying technology to accomplish this task. Depending on the learning paradigm, ML may be broadly classified into *learning with a teacher* and *learning without a teacher*. Briefly, learning under the supervision of a ‘teacher’, or *supervised learning*, seen as a ‘past experience’ of the model, means the process of establishing a correspondence between an input and an output, using a training dataset. The purpose of supervised learning is to predict the output value for any new input after completion of the training process, done under the supervision of a ‘teacher’. Metaphorically speaking, this is the case of a student who learns from his/her teacher through a question-answer process. The (supervised) classification process represents a classical example of the supervised learning. Unlike supervised learning, in learning without a teacher there is no ‘teacher’ to monitor the learning process. There are two categories of such a learning paradigm. Firstly, *reinforcement learning* which aims to connect situations to actions by maximizing a reward (reinforcement) signal. The learning of an input-output mapping is performed by repeated interactions with the environment, in order to maximize the performance, and a “supervisor” does not conduct the student’s learning process. Secondly, *self-organized learning* or *unsupervised learning* which operates with no external ‘teacher’ (or referee) to monitor the learning process. In the unsupervised learning, the model is adapted to observations, being distinguished by the fact that there is no a priori output (the learner is fed with only questions without answers, therefore, no teacher). Classical examples of unsupervised learning are the clustering process, or outlier detection. In this context, we cannot fail to *mention the statistical learning*. Briefly, statistical learning (SL) refers to various tools to model and understand complex datasets by using both statistics techniques and ML algorithms. As in the ML case, there is both supervised and unsupervised SL. While supervised SL builds a statistical model for predicting/estimating an output based on one or more inputs,

unsupervised SL deals with inputs without supervising output. More about DM, ML, and SL can be found, for instance, in Gorunescu [112], James et al. [113], Sugiyama [114]. Below we will provide more details about the mechanisms of DM/ML/SL, along with some healthcare problems they can efficiently solve.

- *Classification/Decision-Making*. First, let us explain from a theoretical point of view the notion of classification/decision-making, also known as pattern recognition, class prediction or discriminant analysis. Thus, (supervised) classification represents a specific regression analysis from the statistical modeling point of view. Based on one or more independent variables (called *predictor variables*), this particular type of regression has to predict a categorical dependent variable (called *grouping variable*). From the AI/ML point of view, the predictor variables represent the object's features, while the grouping variable represents the class label. To conclude, given a set of features, one has to guess the class label, this 'guess' being called a *decision*, and the process itself is called *decision-making*. In this context, the process of automatic diagnosis, supported by ISs, represents practically a classification issue. Taking into account the signs and symptoms of a patient, he/she must be classified into one of the possible classes corresponding to the different types of possible diseases which can cause these signs/symptoms. The process of classification is based on four basic components:
 - *Class*, which is the dependent (categorical) variable of the model, representing the 'label' (diagnosis type) put on the object (patient) after its classification. An example of such a class (label) is "type 2 diabetes".
 - *Predictors*, which are the independent variables of the model, representing the characteristics/attributes (results of the medical tests) of the objects (patients) that have to be classified (diagnosed). For example, the medical test needed to diagnose type 2 diabetes is the glycated hemoglobin (A1C) test, or its alternative test, consisting of random blood sugar test, fasting blood sugar test, and oral glucose tolerance test (for specific details, see, for instance, (<https://www.mayoclinic.org/>)).
 - *Training dataset*, which is the set of data containing values for both the class and the corresponding attributes. It is used in the "*learning phase*" for 'training' the model to recognize the appropriate class, based on available predictors. Examples of such sets are the publicly available datasets provided on Internet for people who work with data. One of the best known and used by the researchers is UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.html>), which contains a rich collection of medical data on different types of cancer, heart diseases, dermatology, hepatitis, HIV, molecular biology, Parkinson, diabetes, and so on.
 - *Testing dataset*, which contains new data, formed by attributes only, without the corresponding class. Based on its previous learning experience, obtained through the training process, the classification model will choose the most appropriate class, and the classification accuracy (model performance) can be thus evaluated in real conditions.

Summarizing, a (supervised) classification process is characterized by:

- *Input*, which is formed by the training dataset containing objects with attributes;
- *Output*, which is the assigned label for each object, based on the its attributes;
- *Classifier*, which is fed with input and used to predict the class of a new object (output) with unknown label.

We have illustrated in Fig. 1.10 the design stages of building a classification model.

Among the most popular classification models (algorithms) today, we could mention:

- *Classification/Decision Tree. Random Forest.* The classification/decision tree (CT/DT) paradigm can be summarized by the following idea. Having a data space, we want to divide it into distinct classes by using a classification rule. Such a rule has the simplified form: “**If** the value of the attribute A_1 is lower than a_1 , **then if** the value of the attribute A_2 is higher than a_2 , ..., **then** the object with attributes A_1, A_2, \dots belongs to class C_i .” The decision boundaries are a_1, a_2, \dots , and the values of the attributes are compared with these thresholds. Once the whole data space is well divided into data sub-spaces as homogeneous as possible, the classification rule is built upon the optimal values of the decision boundaries obtained during the *tree induction*, (the “tree growth” through the training (learning) process), and it will be used to classify new objects. The tree induction is characterized by:
 - Each internal node of the tree (i.e., non-terminal node) expresses the testing based on a certain attribute;
 - Each ‘branch’ of the tree is the test’s result;
 - The ‘leaf’ nodes (i.e., terminal nodes) represent the (decision) classes.

Let us just say a few words about how to split the tree’s branches (i.e., split criteria), considered as measures of node ‘impurity’. Among the most known

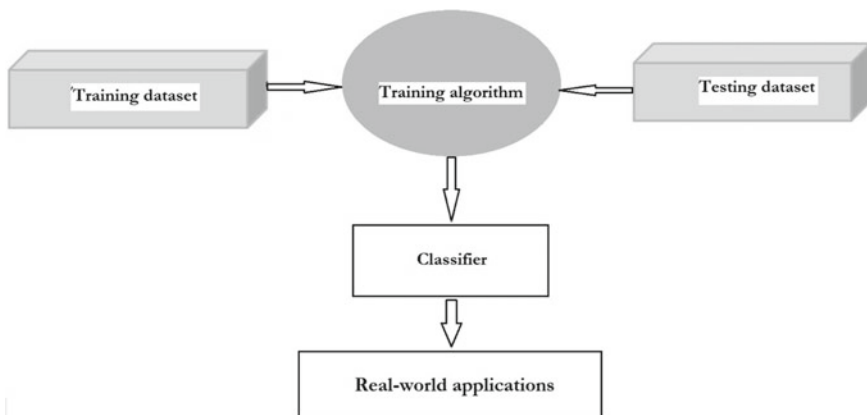


Fig. 1.10 Model of (supervised) classifier

measures we can mention: GINI index, entropy/information gain, misclassification/chi-square/G-square measures, variance reduction, etc. Notable types of decision trees include: ID3 (Iterative Dichotomiser 3), C4.5, C5.0, CART (Classification And Regression Tree), CHAID (Chi-squared Automatic Interaction Detector).

Starting from a (classification) tree, why not use an entire forest? This is, actually, the basic idea of considering the *random forest* (RF) model. The idea behind this paradigm is to build many trees using random objects sampled from a dataset. Seen as an extension of CART, RF builds a multiple CART model with different sample and different initial variables. A random vector is generated for the k th tree, independently from the random vectors corresponding to the $(k - 1)$ past trees, and the remaining data are used for class prediction. In this way, we can consider RF like bootstrapping algorithm with CART model. Among the main advantages of RF let us mention the ability of managing a large number of variables (features), performing a sensitivity analysis (i.e., ranking variables according to the impact on the class variable), providing information about the relation between the variables and the classification, handling missing values, no overfitting, detecting interactions between variables, etc. Instead of growing a single tree, RF grows many trees. To classify a new (still unclassified) object, corresponding to an input vector, we assign the vector to each of the trees in the forest. Then, just like in a voting system, each tree “votes” for a certain class and RF chooses the classification having the most votes over all the trees in the forest [115], (<https://www.stat.berkeley.edu/~breiman/RandomForests/>).

- *Bayesian Classifier.* Let us first mention the roots of this paradigm of great importance in the modern theory of decisions. Recall that the English Presbyterian minister and mathematician Thomas Bayes discovered the following famous result [116], particularly important through its applications. Simply, Bayes’ theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. Formally, the theorem can be described by the conditional probability formula:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)},$$

where H stands for a hypothesis and E represents the corresponding evidence. In the Bayesian terminology, $P(H|E)$ is called *posterior probability*, $P(H)$ is called *prior probability*, $P(E)$ is the *evidence*, and $P(E|H)$ is called *likelihood*. Thus, the Bayes’ formula may be written as:

$$\text{Posterior probability} = \frac{\text{Likelihood} \times \text{Prior probability}}{\text{Evidence}}.$$

The Bayesian classifier is based on the minimization of the expected risk when making a decision regarding the choice of a certain class. When assigning an object to a class, one makes a decision, thus we often speak either of Bayesian

classification or of Bayesian decision. The Bayesian classification/decision rule can be summarized as follows. Let D_k be the decision rule regarding the ‘natural’ state A_k , and let $P(\text{error}|x) = 1 - P(A_k|x)$ be the error related to A_k , given measurement x . The goal of the Bayesian decision is to minimize $P(\text{error}|x)$, by making the optimal choice of D_k . In most cases, when there is no real danger of a wrong decision, all the states/attributes A_k are considered to be independent, hence we speak about the *naïve Bayes* decision/classification.

- *Artificial Neural Networks.* Artificial neural networks (ANNs), or, simply, neural networks (NNs) may be regarded as non-programmed (non-algorithmic) adaptive information processing systems. NNs learn from examples and behave like the human brain. While the network through the learning process acquires knowledge, the (synaptic) weights, quantifying the intensities of the inter-neuron connections, are used to store the gained knowledge. Once the learning/training process is completed and the synaptic weights are set up, the network is used to classify new objects without known class/label. NN is composed of a large number of interconnected processing elements (artificial neurons), working in parallel to solve a specific problem. Basically, NN consists of the *input units* fed with information from the environment, the *computational units* within the network (*hidden neurons*), controlling the actions in the network, and the *output units*, which synthesize the network’s response. Perhaps the most known and used type of NN is the *multi-layer perceptron* (MLP), consisting of multiple layers of computational units, usually interconnected in a feed-forward way. Each neuron in one layer is directly connected to the neurons of the subsequent layer, and the computational units use a certain *activation function*, usually a *hyperbolic tangent function*. It is noteworthy that, according to the *universal approximation theorem* for NNs, just a one hidden layer MLP (3-layer MLP) can approximate arbitrarily closely every continuous function, mapping intervals of real numbers to some output interval of real numbers. The way NN is used for classification is relatively simple. In the training phase, the correct class for each record is known (supervised training), and the output nodes can assign correct values or not. Then, one compares the network’s computed values for the output nodes to the target values, after that it calculates an error term for each node. The errors are then propagated back through the network. The errors are used to adjust the synaptic weights so that, during the next iteration, the output values will be closer to the target values. The iterative learning process consists in the presentation of the training samples to the network one at a time, and the weights associated with the input values being adjusted each time, this process being often repeated a certain number of epochs. The idea behind this learning paradigm is that the network is trained by adjusting the synaptic weights in order to predict the correct class label of input samples. Once the NN architecture has been designed (structure complexity and parameters), the network is ready to be trained. To start this process, the initial weights are chosen randomly and the training process begins. Once the learning phase has been completed and the synaptic weights have been adjusted, the

network is ready to classify new (unlabeled) cases. Advantages of NNs include the high tolerance to noisy data, as well as the ability to classify patterns on which they have not been trained (unsupervised NNs). For more details, see Gorunescu [112], Haykin [117].

- *Deep Learning*. NNs, developed in the 1950s, attempted to simulate the way the brain worked in greatly simplified form due to the computational constraints at that moment. It has been demonstrated that “shallow” NNs, i.e., feedforward NNs with a single hidden layer of finite size have the capacity to approximate continuous functions, result subsequently generalized to feed-forward multi-layer architectures (see universal approximation theorem). As it is well known, the basic idea underlying the continuing development of NNs was to build an algorithm attempting to mimic the activity in layers of neurons in the neocortex, a part of the cerebral cortex where the thinking occurs (higher-order brain functions, such as sensory perception, cognition, spatial reasoning and language, etc.). However, NNs have led to many disappointments as compared to other breakthroughs. Rapid development of modern computing tools, and improvements in mathematical formulas, enable computer scientists to build up NNs with a large number of layers than ever before, which is infeasible for classical ones. Under these circumstances, deep learning (DL) is a modern extension of the classical NNs paradigm. Simplifying, one can view DL as a NN with many layers, in an attempt to mimic better the human neocortex. The other reason for the recent popularity of DL is the increase of the data volume and complexity (big data). Among the most popular DL applications we can mention automatic speech recognition, image recognition, natural language processing, image restoration, healthcare, etc. Medical records such as doctors’ reports, test results and medical images are a gold mine of health information for DL. A clear majority of DL techniques is used in imaging analysis, which makes sense given that images are naturally complex and high volume. Medical images such as MRIs, CT scans, and X-rays are among the most important tools doctors use in diagnosing diseases ranging from spine injuries to heart disease and cancer. Analyzing medical images can often be a difficult and time-consuming process requiring thus predominant use of DL. In the medical applications, the commonly used DL algorithms include convolution neural network (CNN), recurrent neural network, deep belief network, and deep neural network. CNN has been developed as a much better alternative to the classical ML algorithms when handling high dimensional data, i.e., data with a large number of traits. In the classical approach for high-dimensional image analysis, the solution is to perform dimension reduction. Firstly, preselect a subset of pixels as features, and, secondly, perform the ML algorithms on the resulting lower dimensional features. However, the usual feature selection procedures may lose information from the images. Alternatively, the inputs for the CNN are the properly normalized pixel values on the images. CNN then transfers the pixel values in the image through weighting in the convolution layers and, alternatively, sampling in the sub-sampling layers. The final output is a recursive function of the weighted input values. The weights are trained to minimize the average error

between the outcomes and the predictions [118]. The implementation of a CNN has been included in popular software packages such as Caffe from Berkeley AI Research, CNTK from Microsoft, TensorFlow from Google, Keras. https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software.

- *Support Vector Machines*. Suppose we dispose of a set of objects, and we want to separate them into two distinct classes. Imagine a pasture with many sheep grazing, some white, some black, mixed together. The shepherd who looks after them wants, at some point, to divide them by color. If, ideally, the two types of sheep are well separated by an imaginary straight line (i.e., the white ones placed on one side, the black ones placed on the other side), then the (linear) separation issue for the shepherd is completely solved. Unfortunately, in the real-world case, they are well mixed, and the shepherd's problem becomes much more complicated. The support vector machines (SVMs) bring an efficient solution to this problem, i.e., the linear separation of non-linear mixed objects [112, 117, 119, 120]. In order to understand this solution, the shepherd should know what it means to 'teleport' an object in another space, as in the "Star Trek" movies. The idea behind the SVM paradigm is to consider a *separation hyperplane* (in a higher dimensional space than the one containing the original objects), such that the margin of separation between different types of objects (e.g., white sheet and black sheet) is maximized. To summarize, if the original objects are linearly separable, then there exist a (linear) hyperplane, which separates them. If, as in real-world situations, the original objects are not linearly separable, then the kernel SVM is used based on a separation hyperplane in a high-dimension (kernel space). Thus, SVM employs the *kernel trick*, i.e., a clever solution to a non-linear separation problem by mapping the original non-linearly separable points into a higher-dimensional space, where a linear classifier exists. For the shepherd's problem, the sheep are 'teleported' into a high-dimensional space where they can be (linearly) separated. In brief, a non-linear classification in the original space is equivalent to a linear classification in the new space, so the main problem remains the linear separation of objects. Under these circumstances, assume the objects are linearly separable. Consequently, there is a linear hyperplane, which separates the two types of objects (e.g., white sheep and black sheep). There could be infinitely many possible such hyperplanes in the absence of additional constraints. For a given hyperplane, the separation between it and the closest object is called *margin of separation*. The goal of SVM is to find the particular hyperplane, which maximize the margin of separation, the so-called *optimal hyperplane* (or *optimal decision boundary*). Let us finally see where the name of SVM comes from. Thus, the "support vectors" are those objects that lie closest to the optimal hyperplane. It can be said that they are 'on the border', so they are difficult to classify, but, on the other hand, they play an essential role for SVM.
- *k-Nearest Neighbors*. Imagine that we are on the edge of a pond and we see more birds on the water. How do we figure out which ones are ducks? A handy solution: using the inductive reasoning of the 'duck test': "If it looks like a duck, swims like a duck, and quacks like a duck, then it probably is a duck". Returning

to the k -nearest neighbors (k -NN) algorithm for pattern recognition, the classification is performed by labeling a new object taking into account the resemblance with the k closest neighboring objects. Given a training dataset and a new object to be classified, a certain similarity “distance” between the new object and the training objects is first computed, and the nearest (most similar) k objects are then chosen. The new object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors according to the similarity. Unlike other classifiers, there is no training phase for k -NN. For this reason, k -NN is considered a *lazy learner* because it does not learn a discriminative function from the training data, but just “memorizes” the training dataset instead. In addition, k -NN is a non-parametric classifier because it does not make any assumptions on the underlying data distribution. We only need three elements to use it: (a) a set of stored records (training dataset), (b) a similarity ‘measure’ to compute the resemblance between objects, and (c) the value of k , representing the necessary number of (neighboring) objects for comparison, based on which we will achieve the classification of a new object. Next, computing the ‘similarity’ between all the training records and the new object, we identify the k nearest objects (most similar k neighbors), and assign the label that is most frequent among the k training records nearest to that object (“majority voting”). k -NN algorithm is among the simplest of all ML algorithms, being insensitive to outliers, versatile in applications (both classification and regression), having (relatively) high accuracy, and non-parametric feature (no assumptions about data). On the other hand, it is computationally expensive, requires high memory, is sensitive to irrelevant features and the scale of the data, and the prediction speed might be slow for large k . k -NN is particularly well suited for multi-modal classes as well as applications in which an object can have many class labels.

- *Genetic Algorithms for Classification/Feature Extraction.* The genetic algorithms (GAs) are based on the modern theory of evolution, with roots both in the Darwin’s natural selection principle presented in the famous “*Origin of species*” [121], and in the Mendel’s genetics regarding the discreet nature of the hereditary factors transferred from parents to children (“*Versuche uber Pflanzenhybride*”—“*Research about plant hybrids/Experiments in plant hybridization*” (1865)). GAs are considered as the most popular case of evolutionary algorithms (EAs). They represent a metaheuristic optimization algorithm, based on a population of potential solutions and using specific mechanisms inspired by the biological evolution, such as: *chromosomes, reproduction, mutation, recombination, selection, and survival of the fittest*. GAs consist of a population of chromosomes, and multiple operators: selection according to fitness, crossover to produce new offspring, and random mutation of new offspring. A GA algorithm has the following steps. *Step 1*: the data are encoded in a vector form and the recombination and mutation rates are picked; *Step 2*: the population, consisting of a certain number n of chromosomes, is chosen; *Step 3*: the fitness function is computed for each chromosome; *Step 4*: the iteration takes place through selection, crossover and mutation until

n chromosomes have been generated; *Step 5*: the current population is replaced by the new one; *Step 6*: a termination criterion is used to stop the evolutionary process. When GAs are used for a classification task, they usually attempt to find either (decision) boundaries between classes (dividing hyperplanes), or sets of classification rules, or feature extraction/selection, etc.

- *Rule-Based Classification*. A rule-based classifier uses a set of IF-THEN rules to classify objects in different classes. The process of rule-based classification uses a training dataset of labeled objects from which classification rules are extracted in order to build the classifier. The set of rules are then used in a given order to classify new (unlabeled) objects. The IF part of the rule is called rule *antecedent* or *precondition*. The THEN part of the rule is called rule *consequent*. The antecedent part consists of one or more attribute tests that are logically connected by AND. The consequent part consists of class prediction. Mathematically speaking, a rule is an implication of the form $X \rightarrow Y$, where X represents the rule antecedent (a conjunction of attributes values), and Y represents the rule consequent, representing the class label. Rule-based classifiers have some advantages as compared to other classification algorithms: high expressiveness and performance, easy to interpret, easy to generate, and high classification speed.
- *Logistic Regression*. Logistic regression (LR) is a statistical classification model, which aims to determine a certain outcome given by a dichotomous variable with two possible states, based on a dataset containing one or more independent predictor/explanatory variables. To understand the logistic regression, let us briefly recall the regressive model. The basic model is the *simple linear regression* (SLR). SLR is represented mathematically by a linear equation connecting the *response* (dependent) variable Y with the *predictor* (independent) variable X ($Y = b \cdot X + a$). It is interesting to mention in this context that the regression paradigm has its roots in the work of the famous geneticist Sir Francis Galton regarding “*regression towards the mean*” (“Regression towards mediocrity in hereditary stature”, 1886). According to this theory, the characteristics in the offspring regress towards a ‘mediocre’ point (the mean, actually), opposed to the assumption that extreme characteristics in parents are passed on completely to their offspring. Instead of analyzing bivariate data as in the case of SLR, in most real-life issues the multivariate data are being analyzed. Thus, the *multiple linear regression* (MLR) model is represented mathematically by a linear equation connecting the *response* (dependent) variable Y with the *predictor* (independent) variables X_1, X_2, \dots, X_k , ($Y = b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_k \cdot X_k + a$). However, there are many research areas, including healthcare, economics, physics, meteorology, astronomy, biology, etc., in which the dependent variable Y is no longer a continuous variable, but a binary (categorical) one, which can be simply encoded as $Y = 0$ and $Y = 1$ (*logistic regression*). In order to transform the value of the dependent variable into a binary one we shall use the following formula: $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_m \cdot X_m$, where p is the proportion of objects that have a certain characteristic

After reviewing some well-known classification algorithms, it is natural to ask ourselves the question “which ones are the best?” As is to be expected, the answer to such a question depends on the problem at hand, as well as the data we are working with. A subjective choice would be, irrespective of the (subjective) ranking: Naïve Bayes Classifier, Support Vector Machine, Artificial Neural Networks/Deep Learning, Logistic Regression, Random Forests, Decision Trees, k-Nearest Neighbors. However, the problem remains open, due to the fact that even at this time such algorithms are being developed throughout the world.

Besides the classification models, there are other ML algorithms that need to be briefly discussed:

- *Survival Regression Models.* In different real-world circumstances, such as clinical investigations, there are many situations where several known quantities potentially affect a certain outcome (e.g., patient prognosis). Statistical models are frequently used to analyze survival with respect to several factors simultaneously, providing at the same time the effect size for each factor. In many medical, biological or engineering researches a common issue is to determine whether or not certain continuous (independent) variables are correlated with the survival or failure times. In this context, the first idea that comes to mind is to apply a multiple regression model, but we will end up with two major issues. On the one hand, the dependent variable of interest (i.e., survival/failure time) is most likely not normally distributed, violating thus an assumption for ordinary least squares multiple regression. On the other hand, some observations will be incomplete, the so-called *censored data*. One of the most known and used survival regression models is the popular *Cox’s proportional hazard model (Cox PH model)*. It is the most general survival regression model because it is not based on any assumptions concerning the nature or shape of the underlying survival distribution. No assumptions are made about the nature or shape of the hazard function, the model assuming that the underlying hazard rate, rather than survival time, is a function of the independent variables (*covariates*). The standard Cox PH model assumes time independent variables, i.e., variables not changing over time (e.g., sex). A more general approach requires an extension to time dependent variables. One of the advantages of the Cox model is its ability to encompass covariates that change over time. The extension of the Cox model to the case of covariates that change over time is known as the *Cox’s Proportional hazard model with time-dependent covariates*.
- *Cluster Analysis/Clustering.* Cluster analysis or clustering, also called segmentation or taxonomy analysis, is an exploratory data analysis that aims to identify structures within the data. By clustering we mean the technique of dividing a set of data into several groups, called *clusters*, based on certain predetermined similarities. It is noteworthy that the classification process is different from the clustering process. When we classify a certain object, we assign a certain label to that object, while in clustering we group a set of objects in a certain number of clusters, such that objects in the same cluster are more

similar to each other than to those in other clusters. Given a measure of similarity and a set of objects, each of them characterized by a set of attributes, the clustering issue is how to divide them into groups (clusters) such that:

- Objects belonging to a cluster are more similar to one another;
- Objects in different clusters are less similar to one another.

From a ML point of view, the clustering process is a form of unsupervised learning. We can mention three types of cluster structure: (a) a single cluster against the rest of the data, (b) the segmentation/partition of the data in a certain number of clusters, and (c) a (nested) hierarchy of clusters. Classically, there are two major approaches to the clustering process:

- *Partitional/flat* clustering, which represents a division of a set of objects into non-overlapping subsets (clusters), such that each object is in exactly one subset (*hard clustering*). Let us also mention the *soft clustering* in which an object has fractional membership in several clusters;
- *Hierarchical* clustering, which represents a set of nested clusters that are organized as a tree.

Remark. Although partitional clustering is efficient and conceptually simple, it has a number of drawbacks, such as: an unstructured set of clusters; the requirement of a prespecified number of clusters as input; and the techniques are nondeterministic. On the other hand, the hierarchical clustering produces a (nested) hierarchy that is a more informative structure than the flat clustering, does not require a prespecified the number of clusters, and most techniques used are deterministic.

Beyond the classic approach above, we will mention some typical cluster models (https://en.wikipedia.org/wiki/Cluster_analysis), such as connectivity models, centroid models, distribution models, density models, subspace models, graph-based models, neural model, etc.

The clustering process involves three main steps:

1. Defining a similarity measure;
2. Defining evaluation criteria measures for the clusters building process;
3. Building an algorithm to construct clusters based on the chosen evaluation criterion.

The clustering paradigm is based on the similarity between objects. A similarity measure indicates how similar two objects are. The choice of a specific measure essentially depends on the problem at hand. In other words, the resemblance depends on the point of view of the intended purpose (e.g., segmentation according to gender, age, disease type, symptoms, etc.). Let us remark that the choice of a measure of similarity must be always in accordance with the type of available data (e.g., numerical, categorical, rank, fuzzy, etc.).

When we segment a certain dataset into clusters, it is naturally to consider the concept of a good clustering. Obviously, there is no universal definition for what is

a good clustering. Nevertheless, several evaluation criteria have been developed in literature. We will mention below some important criteria in this direction:

- *Internal validation*, which evaluates the clustering result using only quantities and features inherited from the dataset, without respect to external information (e.g., sum of squared error (SSE), other minimum variance criteria, scatter criteria, etc.).
- *External validation*, which consists in comparing the obtained clustering with other segmentation approaches (e.g., mutual information based measure, precision-recall measure, etc.).
- *Relative validation*, which compares two different clustering models or clusters (e.g., statistical testing, using SSE, etc.).

As mentioned above, there are two major classes of clustering models: partitioning methods and hierarchical methods. For the methods in the first category, we mention error-minimization algorithms (e.g., k -means algorithm) and graph-theoretic clustering. Among the hierarchical methods, we mention agglomerative hierarchical clustering (*bottom-up*), divisive hierarchical clustering (*top-down*), and conceptual.

For more information regarding the clustering models, see Gorunescu [112], Rokach and Maimon [122], Mirkin [123].

- *Feature Selection/Feature Extraction*. In Bellman [124], has first introduced the term of “*curse of dimensionality*”, referring as a phenomenon which often occurs when applying ML algorithms to high-dimensional data. In this situation, “the number of samples needed to estimate an arbitrary function with a given level of accuracy grows exponentially with respect to the number of input variables (i.e., dimensionality) of the function” [125]. In the ML practice, there is a maximum number of features in the database, above which the algorithm’s performance will decrease rather than increase. To avoid such a situation, one can use the dimensionality reduction technique, obtaining a reduced amount of time and memory required by data processing, better visualization, elimination of irrelevant features and possible noise reduction. As techniques for dimensionality reduction, we can mention typical multivariate exploratory techniques such as factor analysis, principal components analysis, multidimensional scaling, cluster analysis, canonical correlation, etc. There are usually two different approaches for dimensionality reduction: (a) *feature selection (FS)*, and (b) *feature extraction (FE)*. FS uses different methods for selecting a subset of the existing features, the most significant ones for the proposed purpose (e.g., classification, regression). In this way, the selected subset of variables from the input can efficiently describe the input data. As a result, the effects from noise or irrelevant variables are reduced, and this approach provides prediction at least as good, but with a much lower computing effort. Instead, FE performs a projection of the existing features to a lower dimensional space, altering thus the original representation of the data. FE creates new variables as combinations of others to reduce the dimensionality of the selected features. The literature on FS and FE

techniques is very vast, covering ML and pattern recognition algorithms. Since each underlying algorithm behaves differently for different data, a comparison between FS/FE algorithms can only be done using a single dataset. Of the well-known FS techniques, we recall (a) *filter methods*, (b) *wrapper methods*, (c) *embedded methods*, (d) *ensemble FS*, etc. Regarding FE, we can mention (a) *principal component analysis*, (b) *linear discriminant analysis*, (c) *factor analysis*, (d) *multidimensional scaling*, (e) *correspondence analysis*, etc.

So far along, we may think that IS in healthcare is a real miracle. But can we actually trust its results? Is it possible to measure the machine’s performance and if so, can we take comfort in the fact that the measurements are accurate? Fortunately, since 1800 we have Statistics to rely on. Data scientists use statistical methods for comparing classifiers. This is not an easy task to tackle. As stated before, there is no single classifier that works best on all given real-world problems. No classifier offers a ‘short-cut’. We can say that this phenomenon is related to David Wolpert and William G. Macready theorem ‘No-free-lunch’. Each classification algorithm plays the role of a ‘restaurant’, providing a certain ‘dish’ at a certain ‘price’, thus its up to us to determine the ‘smart-deal’ [126].

Because the issue at hand is the healthcare system, a very important part is represented by the associated cost regarding correct and incorrect classification. Depending on the case the cost varies. For a better understanding of the process, let us imagine the following situation: let’s suppose that we have a classifier that establishes whether a tumor is benign or malignant in a certain type of aggressive cancer. The cost of misclassifying a malignant tumor as benign is greater than misclassifying a benign tumor as malignant? Both situations are serious, whether a person believes he/she is healthy and does not start the treatment (surgery, chemotherapy, radiotherapy, etc.) or the person is healthy and starts an unnecessary treatment [127].

The classification evaluation counts the correct and incorrect predicted objects. These values are being tabulated in a *confusion matrix*—see Table 1.3. In addition, a *cost matrix* can be created in order to maximize benefit or minimize cost—see Table 1.4. Depending on the case, sometimes a more accurate classification is desired for certain classes rather than others.

Each cost is assigned according to the problem. The overall cost and accuracy are computed using the following formulas:

$$\begin{aligned} \text{Cost} &= p \times a + q \times b + r \times c + s \times d \\ \text{Accuracy} &= \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN} \end{aligned}$$

Example 1.1 Let us consider the example mentioned above: a case of an aggressive form of cancer. We have two classifiers that C_1 and C_2 . Below we have three tables. The first one, Table 1.5, presents the cost matrix established a priori in this sensitive issue. Tables 1.6 and 1.7 present the confusion matrix for the two classifiers.

Keeping the formulas in mind, we will obtain the following numbers measuring the performance:

Table 1.3 Confusion matrix (2 class model)

Classification	Predicted class		
Observed class		Class = YES	Class = NO
	Class = YES	<i>a</i> (true positive = <i>TP</i>)	<i>b</i> (false negative = <i>FN</i>)
	Class = NO	<i>c</i> (false positive = <i>FP</i>)	<i>d</i> (true negative = <i>TN</i>)

Table 1.4 Cost matrix (2 class model)

Cost matrix	Predicted class		
Observed class		Class = YES	Class = NO
	Class = YES	<i>p</i>	<i>q</i>
	Class = NO	<i>r</i>	<i>s</i>

$$\begin{aligned}
 Cost_{C_1} &= -2 \times 134 + 50 \times 45 + 1 \times 76 + 0 \times 217 = 4408 \\
 Cost_{C_2} &= -2 \times 203 + 50 \times 60 + 1 \times 100 + 0 \times 109 = 2694 \\
 Accuracy_{C_1} &= \frac{351}{472} = 74\% \\
 Accuracy_{C_2} &= \frac{312}{472} = 66\%
 \end{aligned}$$

Now comes the real problem. We have both the cost and the accuracy of each classifier. These pairs of parameters that determine the measure of performance must be taken into consideration when choosing the best performing model. This remains an open question, since we must find a balance between cost and accuracy.

Besides the cost and the accuracy, two other statistical concepts must be added to the checklist for finding the ‘smart-deal’ of IS: *sensitivity* and *specificity*. It must be noted that these concepts are used for binary classification. Both concepts are equally important, so a classification must be sensitive and specific. *Sensitivity* measures the proportion of ‘true positives’ that are correctly classified, whereas *specificity* the proportion of ‘true negatives’ correctly classified.

Table 1.5 Cost matrix

Cost matrix	Predicted class		
Observed class		Class = YES	Class = NO
	Class = YES	-2	50
	Class = NO	1	0

Table 1.6 Confusion matrix for model C_1

Classification C_1	Predicted class		
	Observed class		Class = YES
Class = YES		134	45
Class = NO		76	217

Table 1.7 Confusion matrix for model C_2

Classification C_2	Predicted class		
	Observed class		Class = YES
Class = YES		203	60
Class = NO		100	109

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

In order to find the probability that the model will give the correct diagnosis we need to compute two other concepts: *positive predictive value* (PPV) and *negative predictive value* (NPV). Basically, PPV computes the proportion of ‘positive’ cases that are classified correct, and NPV computes the proportion of ‘negative’ cases that are classified correct.

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

If one interprets these parameters in our example, 100% sensitivity means that all cancer patients are recognized as having the disease, whereas 100% specificity means all healthy people will be recognized as being healthy. Unfortunately, 100% sensitivity and 100% specificity is just a beautiful dream, impossible to be reached.

For even a more thorough analysis we can use other four indicators:

- *false positive rate* (FP rate) = $FP / (FP + TN) = 1 - specificity$
- *false negative rate* (FN rate) = $FN / (TP + FN) = 1 - sensitivity$
- *Likelihood ratio positive* (LR+) = $sensitivity / (1 - specificity)$
- *Likelihood ratio negative* (LR-) = $(1 - sensitivity) / specificity$

In practice, the medical personnel prefers to use high false positive rate, rather than false negative rate.

Another interesting way of assessing the prediction results is the *Receiver Operating Characteristic Curve* (ROC). The history of the ROC curves is

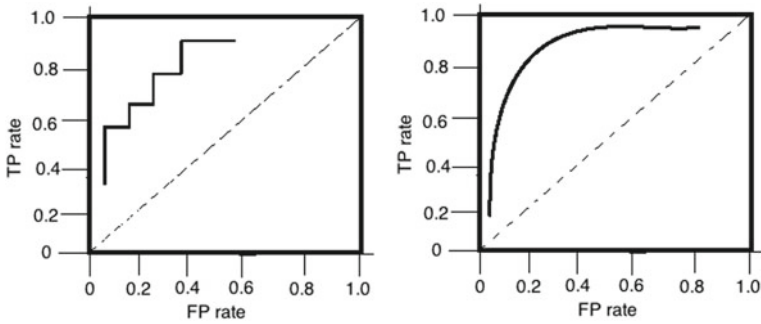


Fig. 1.11 ROC curve for the discrete and continuous case

fascinating. For this, we must time travel back on December 7, 1941. We are located on the island of Oahu, Hawaii. We are during the *World War II*, in Pearl Harbor. The radar picked up some unusual behavior, described as a flock of birds. The famous reply concerning this situation was: “Don’t worry about it”. The primitive radar failed to differentiate between enemy aircrafts and birds. 2,403 Americans died that day. 1,178 were wounded. Electrical and radar engineers started developing ROC curves. Since then, they have been used in various applications in medical research [127–132].

For a better understanding we shall present in Fig. 1.11 below two types of ROC curves, discrete and continuous depending on the output of the classifiers. A ROC curve is a two-dimensional graph, which plots on the X-axis the FP rate, and on the Y-axis the TP rate. This way we can find the trade-off between the benefits and costs.

Now that we see how a ROC curve looks like, we need to interpret it. In order to obtain a good interpretation, we can use the following rule: one point in space is better than another if it is situated to the northwest of the square. Practically, if the TP rate is high and FP rate is low, or both at the same time, the prediction is better. In fact, using the *line of no-discrimination*, one can say that if the points are above the line, the classification is good, otherwise is poor.

For some, the ROC curve is not as suggestive as a number, so the solution is using the *Area under the ROC curve* (AUC). Obviously, a picture is worth a thousand words, whereas a number is a number. Using the AUC diminishes the information regarding the pattern of trade-offs of the classifier.

The translation guide of the AUC is:

- 0.90–1.00 = excellent classification
- 0.80–0.90 = good classification
- 0.70–0.80 = fair classification
- 0.60–0.70 = poor classification
- 0.50–0.60 = failure.

Ok, so what about the values below 0.50? If you random guess the outcome you produce the diagonal line between 0.0 and 1.0. Obviously, any number below 0.50 is a total failure.

But what is the philosophy behind the AUC? If one is familiar to the Mann-Whitney U test or the Wilcoxon test of ranks, the interpretation of the AUC is equivalent. AUC is the probability that the classifier will assign a higher score to a positive example rather than to the negative one. More information about AUC can be found in Fawcett [133], Hanley and McNeil [134].

If a data scientist wants to validate its model, unfortunately the ROC curve and AUC are not sufficient. Let us suppose that again we have two classifiers C_1 and C_2 that we want to compare. If both classifiers are tested on the same database, there shouldn't be any issues, but what if the first one, C_1 has an accuracy of 72% on a sample size of 4500 items, and the second C_2 has an accuracy of 88% on a sample size of 50. Can you actually say the second classifier performs better?

Fortunately, the z -test, also known as the “*difference between two proportions*” test will provide the correct answer. The test computes the p -value (which determines how much was the hazard involved in the result). If $p > 0.05$ there is no statistical significant difference between the two models.

At this point, we answered the questions why and how IS in healthcare. Our journey continues to the next level: what are the actual benefits of intelligent healthcare?

1.3 What Are the Benefits of Intelligent Healthcare?

Intelligent healthcare... The name resonates with a science fiction movie or book. We shall once again state that is our strong belief that there will be no doctorless hospitals. Even if healthcare without AI can no longer exist, the human sixth sense shall never be replaced. Nevertheless, intelligent healthcare is happening today.

It should be noted, that unfortunately 100% accuracy in diagnosis or patient management is just a dream. People fail, machines fail. Statistics is just numbers that give you an insight on what may or may not happen. Data scientists ‘play’ with numbers, but unfortunately, those numbers are people, people’s characteristics. No one knows exactly what will happen if you pinpoint a certain person, but intelligent healthcare and statistics can give you and the doctors a probability. If that person is part of that probability or not it is not up to us to know or foresee.

In healthcare there are three major concerns: a disease, a disorder and a syndrome. The definition of a disease was published in the British Medical Journal [135] in 1900: “*resulting from a pathophysiological response to external or internal factors*”. A disease has signs and symptoms (e.g. cancer, cardiovascular disease). The definition of a disorder is “*abnormal physical or mental condition*” [136] (e.g. arrhythmia—abnormal heartbeat). A syndrome is a collection of symptoms that suggests a certain disease (e.g. Down’s syndrome, auto-immune syndrome, acute respiratory distress syndrome). If for a disease or a disorder there is

treatment, for a syndrome there is no such thing. That is why it is very difficult for the medical personnel to identify and treat a certain syndrome. Here the trial-error approach is used.

The success of using AI in medicine is going up. Billions of dollars are pushed in the system. According to a recent report of Frost and Sullivan (<https://ww2.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare>—accessed May 4, 2018) the market for AI will reach 6 billion dollars in 2021. AI goes towards personalized medicine gathering information in order to tailor personalized treatments and to monitor patient's response to it.

Next we will present the benefits of using intelligent healthcare in diagnosing and treating diseases and disorders.

The most frightening C word is Cancer. It is everywhere. Let's admit it, each one of us has one or even more loved ones who are/were diagnosed with this disease. Some are still with us struggling each day, praying and hoping for new therapies to be discovered, others lost the battle. The statistic is bad. According to Cancer Research UK (<http://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer#heading-Zero>—accessed May 4, 2018) in 2012, 14.1 million new cases of cancer occurred worldwide. In 2008 an estimated 169.3 million years of health life were lost globally due to cancer. It is an estimate that by 2030 there will be 23.6 million new cases each year.

In the last episode of Dr. House aired on May 21, 2012, Dr. House tells his best friend, an oncologist who had just been diagnosed with cancer that: "Cancer is boring". Sadly, it isn't true. In Pleasance [137], the authors state that in fact cancer is actually hundreds or thousands of rare diseases, and that each tumor is to some extent unique. Nothing is boring about that. In some cancers (e.g. esophageal) the chemotherapy protocol has not change in years. If the genomic and outcome data is recorded each unique tumor might have its protocol.

But what is cancer? Cancer happens when a cell mutates and starts to reproduce but has no longer control of its growth. Even in the same type of cancer, cells can grow rapidly and that makes the cancer aggressive or can grow slowly. So, even if it is of the same type, cancer is an individualized disease, so it must be addressed accordingly McDermott et al. [138].

So now that we have established an example of a devastating disease, let's take a look on how would the healthcare system benefit from the use of AI.

Due to smoking and polluted air, the rates of lung cancer in China have gone up. Only 15% of people that are diagnosed survive. Fortunately, due to the fact that Chinese citizens are required to get regular lung screenings, the disease is caught early. On May 16, 2017, Forbes magazine presented Infervision Inc., a company that focuses on intelligent healthcare (<https://www.forbes.com/sites/jenniferhicks/2017/05/16/see-how-artificial-intelligence-can-improve-medical-diagnosis-and-healthcare/#129562306223>—accessed May 5, 2018.). Infervision uses Deep Learning to learn from images produced by XRays, MRIs, CTs and even pathology, in order to identify abnormal tissues in the lungs. The technology is being used in Shanghai Changzheng Hospital.

At the Mayo clinic, the genomic information of brain tumors is discovered without a biopsy, by the use of Deep learning. At Stanford, an ANN is trained to recognize skin cancer.

“Deep-patient” is an unsupervised representation that predicts from EHR, the future of patients from 78 diseases from schizophrenia, severe diabetes to various cancers [139].

Even if cancer is scary, it is the second leading cause of death worldwide. According to a report of the American Heart Association—<http://www.heart.org/HEARTORG/>—accessed May 5, 2018—that compiled data from more than 190 countries, heart diseases took 17.3 millions lives each year. By 2030 the number is expected to be 23.6 million. Comparing cancer and heart disease deaths in expected 2030, we can observe a tie.

About 47% of sudden cardiac deaths occur outside a hospital. Heart problems are currently diagnosed by monitoring the timing of the heartbeats in scans, but cardiologists are not always accurate. During this procedure one in five patients suffers a heart attack or undergoes an unnecessary surgery. Out of 60000 scans performed each year, 12000 are misdiagnosed. The estimated cost is \$812 millions.

An AI system that was developed at the John Radcliffe Hospital discovers in scans information that the human eye cannot see. The system has been tested in six cardiology units in clinical trials and the results are expected to be published soon.

Deep learning can find patterns in heterogeneous syndromes and image recognition. Intelligent healthcare can classify new genotypes and phenotypes for: heart failure, Takotsubo cardiomyopathy, hypertrophic cardiomyopathy, hypertension, and coronary artery disease. Just like in cancer, the therapy can be tailored accordingly.

Data Scientists from GOOGLE and Verily have developed a system that assesses a person’s risk of heart disease using AI. The deep learning algorithm analysis scans of the back of a patient’s eye and gathers information related to the individual’s age, blood pressure and even if the person is a smoker or not Poplin et al. [140]. 300 000 patients were used in this study. The fundus of the eye is full of blood vessels that reflect the overall health. Analyzing the fundus important predictors of cardiovascular diseases are found.

In an article that was published by the Journal of the American Medical Informatics Association [141], the researchers modeled a recurrent neural network in order to learn temporal relations among events in EHR. Thus, the algorithm anticipates early stages of heart failure, leading to a tailored prevention plan. In a recent study, which was published in January 2018 [142], a smart phone equipped with an AI algorithm measures arterial stiffness. Arterial stiffness appears when the arteries become rigid, increasing blood and pulse pressure. This process causes the tension to travel to the peripheral vasculature, causing organ damage (kidney, brain, etc.). The only current alternatives to this smart phone app are MRI and tonometry.

The 7th leading cause of death in the USA is diabetes according to the American Diabetes Association—<http://www.diabetes.org/diabetes-basics/statistics/>—accessed May 16, 2018. The cost of caring for diabetes patients is up to \$245 billion per year [143]. In 2015, 84.1 million Americans age 18 and older had prediabetes.

Unfortunately this disease is turning into an epidemic. According to Neocleous et al. [144] 28.5% of diabetics have diabetic retinopathy that often leads to blindness. A deep neural network was developed for detection of diabetic retinopathy in 2016, Pomprapa [145].

Diagnosing diseases and disorders is a difficult task. An even more difficult issue is diagnosing syndromes. Syndromes have no protocols when it comes to diagnostic or treatment. Doctors use the trial error approach in most cases, often leading to misjudge.

But, here comes AI to the rescue (at least in some syndromes...)! AliveCor—<https://www.alivecor.com/>—accessed May 14, 2018—together with Mayo Clinic started in 2016 to test whether the long QT syndrome can be identified in EKG results. Long QT syndrome causes almost 4000 sudden deaths per year in children and young adults in the US. According to Knaus et al. [146] the syndrome appears in 1 in 2000 live births. AliveCor sells portable EKG sensors that can be attached to smartphones and an EKG embedded in an Apple Watch band. The system developed by AliveCor uses a deep neural network that learns from the EKG results from over 1000 patients in order to find relevant features of the long QT, leading to an early diagnosis that might prevent sudden deaths.

AI is improving Down syndrome diagnosis also. The results published in *Ultrasound in Obstetrics and Gynecology* in 2017 [147], by a research team from Netherlands, Cyprus and the UK, use an ANN to diagnose the syndrome. Current non-invasive procedures compute an estimate of the maternal age, blood tests and ultrasound to detect the presence or absence of fetal nasal bone. Still, the estimates have a considerable false-positive rate. Another non-invasive test is the cell-free DNA, but the costs are very high. The same trained ANN can diagnose Turner syndrome also.

An automatic ventilation system using AI was developed for acute respiratory distress syndrome (ARDS). ARDS is hard to diagnose, because it has to be differentiated from pneumonia or congestive heart failure. ARDS has no cure and can leave the survivors with diminished functional capacity, mental illness and low quality of life; medical personnel just provide life support for the patient until the lungs start functioning again. The aim of the study [148] is to develop an automatic control system for mechanical ventilation using AI. Currently this approach can be applied in animal trials, no approval for humans has been granted yet.

AI is used to diagnose rare diseases, such as the Mabry syndrome [149]. The Mabry syndrome is a rare condition that is triggered just by the mutation of a single gene, and it causes mental retardation. The scientists use AI to find in the photographs of 91 patients specific cell surface changes, such as: a narrow tent-shaped upper lip, broad bridge of the nose and wide-set eyes, etc. These features may be obvious or not, thus complicating the differential diagnosis.

If these technologies will be produced at a reasonable cost so that every primary care office could afford them, the detection could be made early and it could save lives.

Besides diagnosing diseases, the benefits of intelligent healthcare are found in:

- *Managing Medical Records.* As stated earlier in this chapter, only 27% of the workday, the doctor has face time with the patient. An intelligent management of medical records improves that ratio. An intelligent system may collect, store, search data like medical history of a patient, etc. much faster.
- *Digital Consultation.* According to Winters et al. [150] a John Hopkins' study made in 2012 in the US, 40500 patients die per year in the ICU due to misdiagnosis. An UK-based startup, Babylon, uses AI to give medical consultation. The user reports his/her symptoms into the app, which uses speech recognition to compare the symptoms to a database of illnesses. Babylon recommends even treatments. Hopefully, it is not like a GOOGLE symptom search that scares you to death. Another tool is <https://symptoms.webmd.com/#/info> (accessed May 5, 2018) symptom checker.
- *Virtual Nurses.* Yes, there are two AI nurses, Molly—<http://www.sensely.com/>—accessed May 5, 2018—developed by the Startup Sense.ly and KidsMD—a new skill created for Amazon Alexa—enabled devices. Molly is a digital nurse that monitors patient's condition and follow up with treatments between patient's visits to the doctor. It is specialized in chronic illnesses. KidsMD is developed by Boston Children's Hospital and answers questions like “*My child has 101.3 fever. What should I do?*” The app answers which medications should be taken and if a doctor's visit is required. Somehow similar, when GOOGLE is told that you have a fever, it points on GOOGLE Maps the closest hospitals.
- *Reading X-rays, CT Scans, EKG, EEG, etc.* Analyzing these images is time consuming and tiring. Using AI for this, the doctor has the time to focus on the more complicated cases.
- *Medication Management.* AICure app—<https://aicure.com/>—accessed May 5, 2018, is developed by the National Institutes of Health that monitors the way the patient uses medication. AI and the smartphone's webcam observe and confirm the fact that the patient has taken his/her medication. Users are participants in clinical trials, treatment completion for tuberculosis, hepatitis C. Patients that suffer from dementia or Alzheimer diseases could benefit for it.
- *Drug Creation.* The time needed for developing drugs through clinical trails is long, it takes months or even years, and the cost of this process reaches billions of dollars. AI for drug discovery—<http://www.atomwise.com/>—accessed May 5, 2018, was used to scan existing medicines that could be redesigned to fight the Ebola virus. The speed of the program was amazing, finding two drugs that may reduce Ebola infectivity in just one day.
- *Precision Medicine.* AI detects from body scans mutations and abnormal information in DNA. This helps predict diseases early based on genetics.
- *Monitoring Ones Health.* Apple, Fitbit, Spire, Prana, etc. have developed wearable health trackers that monitor ones health (heart rate, pulse, breathing, stress levels, posture, etc.). The system alerts problems to the user and even shares the information to doctors. Feebris—<http://feebris.com/>—accessed May 5, 2018- analyses data through digital stethoscopes to fight pneumonia.

CloudUPDRS—<http://www.updrs.net/>—accessed May 21, 2018—is a smart-phone app that assesses Parkinson’s disease symptoms through a “finger tapping” test. AsthmaMD—<http://www.asthmamd.org/>—accessed May 21, 2018—assesses lung performances through a hand-held meter. The app records other symptoms and medications.

The origin of medicine comes from the Latin “Ars medicina”—“the art of healing”. Beautiful said, right? It is an art after all. Let’s sail a little throughout history:

- In ancient Egypt the first known by name physician was Imhotep, 2600 BC. He was able to describe and treat 200 diseases;
- In 500 BC Alcmaeon of Croton was able to differentiate arteries from veins;
- The first aspirin was prescribed in 400 BC by Hippocrates;
- Smallpox was diagnosed in 910 AD in Persia by Rhazes;
- The spectacles were invented by Roger Bacon in 1249;
- Leonardo da Vinci dissects corpses in 1489;
- The microscope was invented in 1590 by Zacharius Jannssen;
- The birth of cardiology was in 1628 by the published study of William Harvey, *An anatomical Study of Motion of the heart and of the blood in animals*;
- In 1656, took place the first canine blood transfusion experimented by Sir Christopher Wren;
- The first appendectomy is performed by Claudius Aymand in 1763;
- Edward Jenner developed the first vaccine in 1796;
- After 162 years from the first canine blood transfusion, the first successful blood transfusion on humans takes place. James Blundell performs it;
- In 1849, the first woman gains her medical degree from the Geneva Medical College in New York. She was Elizabeth Blackwell;
- Wilhelm Conrad Roentgen discovered the X-rays in 1895;
- In 1901 the blood types were classified into A, B, AB and 0 by Karl Landsteiner;
- The first time an electrocardiograph was used was in 1913. Dr. Paul Dudley White was the pioneer’
- Sir Alexander Fleming discovered penicillin in 1928;
- The first paper on medical ultrasonic-ultrasound was published in 1942 by Dr. Karl Theodore Dussik;
- John Hopps invented the first cardiac pacemaker in 1950. In 1952 the pace-maker was developed by Paul Zoll;
- The first kidney transplant took place in 1954. The doctor who performed the surgery was Dr. Joseph E. Murray;
- The first human heart transplant was performed in 1967 by Dr. Christian Bernard;
- In 1975 Robert S. Ledley invented CAT-Scans.

Medicine has come a long way from bloodletting for every disease to personalized medicine. To refresh one’s memory let us mention that for headaches the

cure was pressing a hangman's rope to your head, or for gout the treatment was applying a superb mixture of worms, pig's marrow and herbs, boiled together with the carcass of a red-haired dog.

We hope we have convinced you of the grand benefits that AI brings to healthcare. Let us continue our journey to see what is next for healthcare, what prospects do intelligent systems bring?

1.4 What Are the Prospects of Using Intelligent Systems in Healthcare?

Healthcare... The final frontier... These are the voyages of the Intelligent Systems. Their continuing mission: To explore strange new diseases... To seek out new treatments; new drugs... to boldly go where no one has gone before! This is an adaptation of Captain's Jean-Luc Picard quote from the Star Trek series *The Next Generation* (1987) that fits perfectly in the context regarding the prospects of Intelligent Systems in healthcare!

A hard task is to identify which idea is pure speculation or not. A well-informed opinion can be found in the report delivered in December 2017 by JASON [151]. JASON is a scientific group, which has been advising the US government since 1960. The group's concerns are of sensitive nature. The report on McNamara Line electronic barrier strategy employed in 1966–1968 during the Vietnam War brought the group's notoriety. Even if the group is military focused, JASON now redirects its interest in AI in healthcare.

There are two major actors in this play. First, there is the academia, which is interested in developing new ML techniques. The other one is the private sector, which has expressed an enormous interest in AI in health applications. Currently there are 106 listed startups from 15 different countries (US, UK and Israel, etc.). The question that intrigues everybody: *Is this just a step in the long road of intelligent systems applied in healthcare or their end is just around the corner?*

The quality and cost of healthcare all over the world have enabled the scientists and the industry to open their minds and seek new solutions. JASON's report states that this time around, AI applied in healthcare will be a blast because the confluence of three forces:

- The current medical system is unsatisfactory
- The ubiquity of smart devices in our lives
- The success of home services provided through Amazon and others.

One of the prospects of intelligent systems in healthcare is the development of clinical applications. For this idea to be put into practice one major concern regards the availability of relevant data [152]. Finding quality data is a real challenge. Medical datasets have privacy issues and are expensive to collect because of longitudinal studies (repeated observations) and clinical trials. Once researchers collect

the data, they tend to keep it to themselves. Other issues concern the lack interoperability of EHR systems.

After the data is collected, it must be labeled. The labeling is done by independent professional assessments of each case, whether we are talking about images or other types of data.

A very interesting fact pointed out by JASON, that we totally support and agree with, is that the academia's focus should be concentrated also on creating rigorous testing and validation approaches, not only on researching for new methods. The problems in implementation must be identified and ameliorated [153, 154], in order to build up confidence. Modifying existing protocols for diagnosis and treatment can be achieved only if the medical community trusts the intelligent systems. For this, the intelligent systems:

- Should address an important clinical need;
- Should perform at least as well as existing approaches;
- Should be statistically tested and validated in order to be trusted;
- Should provide improvements in patient outcomes, life expectancy, quality of life;
- Should reduce costs.

mHealth (mobile or digital health) apps are becoming nested by smart technologies [156]. The apps provide answers to a full spectrum of health issues. Some questions have risen regarding the fact whether the medical community should integrate mHealth into healthcare? Some of the mHealth apps have been included in clinical trials. The American Medical Association (AMA) has recently embraced a set of principles to promote effective and safe mHealth apps. AMA encourages the medical personnel and users to use the apps, associated devices, sensors and trackers.

According to a survey released by AMA in 2016, 31% of doctors see the potential of digital tools in medicine and half of them believe in the improvement brought by those (<https://www.ama-assn.org/sites/default/files/mediabrowser/specialty%20group/washington/ama-digital-health-report923.pdf>). The AMA report states the fact that academia and the industry must fit the IDSS in the existing medical systems and practices by including data privacy assurance, linkage between EHR and billing, reimbursements.

The use of mHealth apps is two-fold. On the one hand the users can monitor their health, and, on the other hand, large dataset can be created and then used for training the AI applications. For example, using the Asthma MD app, the user can upload anonymously data into a Google database for research purposes. The data is analyzed in order to find correlations between asthma and environmental factors, triggers, and climate change. Another example is the mHealth app for managing Parkinson disease [155]. The app was available for free in the US through the Apple App Store. 48 k people downloaded the app, and 78 of the 25% eligible individuals accepted to share their personal data. The numbers show that people are open to building medical databases for research purposes.

More and more research proposal grants on IDSS applied in healthcare are written in order for the academia to receive sponsoring from government agencies. Still, it is our strong belief that in some countries there is still not enough funding in this domain.

Assessment through smartphone apps cannot be done in cases like measuring metrics of health (minerals, vitamins, hemoglobin, cholesterol, etc.), viruses, bacteria, and cancer or heart disease biomarkers. Hopefully, one day, maybe these will be evaluated through small blood samples, and then the dream of monitoring them through mHealth will become true.

All the ideas presented above sound too good to be true. Sadly, like everything else in this world, this internet/smart device healthcare has its Achilles' heel. Because of the enormous sums of money that are invested in this technology, there are and will be many predatory companies that scam people. So the question goes down to this: how does someone separate the wheat from the chaff? How can we know which app to trust?

There are many sites that offer treatments for certain diseases at an unreasonably expensive price. For example, we shall present a case mentioned in the JASON report. It regards a mutation of the Methylene tetrahydrofolate reductase (MTHFR) gene. Doing a simple Google search on this MTHFR gene mutation, one discovers some sites that state that the "conventional" medical doctors ignore this gene as source of your health problems. Some *naturopaths* have obviously the solution to one's problem. Firstly, you will read all the health issues that this mutation produces: anxiety, fatigue, brain fog, dysplasia and of course cancer, heart attacks, stroke, Alzheimer, miscarriage, autism, etc. A genetic test is then required in order to see whether the gene is mutated or not. The cost of a test plus four-month treatment plan is \$3000. A "small" price for finding out what is wrong with you, and for enabling you to take control of your life and health. The mutation will be solved through a "natural remedy plan".

To avoid these scams, you should read carefully the sites you visit. For example, if the site is out of date for several years, that should indicate a problem. Another warning sign is the expert's CV. Check his/hers ISI Web of science, Scopus and Google scholar profile. If his/her citations come only from himself/herself, you should start having doubts. Check out where he/she published her articles. Are the journals trustworthy, or are they predatory?

This example can be easily extrapolated to IDSS applications. Imagine all the dermatological diseases, including skin cancer. Online skin cancer services exist: <https://www.directderm.com/>—accessed May 21, 2018, <https://www.firstderm.com/>—accessed May 21, 2018, <https://www.skinvision.com/>—accessed May 21, 2018. Skinvision is a new company that enables users to send a picture through the app and receive feedback on it. There is little information about the algorithms behind the analysis, and nothing on how was this application validated. In fine print we can read that the website acknowledges that "our solution is not a diagnostic device".

Do not get fooled by these websites. There are websites that can be trusted such as: Mayo Clinic—<https://www.mayoclinic.org/>—accessed May 21, 2018, WebMD—<https://www.webmd.com/>—accessed May 21, 2018.

Another prospect of intelligent systems in healthcare should be the encouragement of “crowdsourcing” movement. AI competitions stimulated the creation of large corpuses of data, used afterwards in healthcare applications. Some AI competitions are: ImageNet <http://www.image-net.org/challenges/LSVRC/>—accessed May 21, 2018, Kaggle, The Booz Allen Hamilton Data Science Bowl—<http://www.datasciencebowl.com/totheclinic/>—accessed May 21, 2018, Zooniverse—<https://www.zooniverse.org/>—accessed May 21, 2018.

We appeal again to the famous Dr. House. If you have watched the series, you know that whenever a difficult case was presented to him, Dr. House asked his team to go and search the person’s house, neighborhood, work place, etc. This is personalized medicine. Everything we do, our education, our neighborhood, our economic status, our diet, our social context defines and determines our health. If we corroborate all of these with genetic information we can determine certain diseases and elaborate treatment plans.

The Human Genome Project was an international research plan with the goal of determining and mapping all the genes in the human genome. In 1984, the US government started this plan, but the project was launched in 1990 and completed in 2001. In 2003, the human genome sequence, 3 billion DNA bases, was determined. The ultimate goal of the project was to identify correlation between certain gene variations and human diseases. The scientists were looking for the answers to questions such as: Why do some people have cancer or Alzheimer? If you could have your human genome analyzed in order to foresee which disease you will suffer from during your lifetime, would you want to know? Surprisingly enough, it has been found that human disease is rarely linked to specific genetic mutations [157, 158]. This led to the idea that personalized medicine is a combo box of genetics, behavior, environment, family history, life and treatment experiences.

The Project *All of Us*—<https://allofus.nih.gov/>—accessed May 21, 2018—brings together National Institutes of Health (NIH) and Precision Medicine Initiative (PMI). The goal of the project is to collect a massive amount of individual health data from genetics to social behavior. The data gathered will contain:

- Basic information on medical history and lifestyle (habits and overall health);
- Physical measurements (blood pressure, pulse, height, weight, hip/waist circumference);
- Biosamples of blood and urine;
- Optional DNA testing;
- EHR (health care visits, diagnosis, procedures, treatments, vital signs, laboratory tests).

All this data will be collected over 10 years. Future goals include EHR and data collected from wireless sensor technologies (mobile/wearable devices) and geospatial and environmental data. All the data will be available to participants,

researchers and the public. On the bright side, this means that the academia and the industry can have a large dataset that they can use to improve their IDSSs, but unfortunately the privacy protection of the participants is not guaranteed. PMI has recognized this from the start of the project. The *All of Us* project is trying to develop privacy, trust principles, data security policy principles and framework.

For the database to be complete, the environmental data must be monitored and recorded (e.g. pathogen exposure). Let us stop for a minute and ask the following question: when you were in your doctor's office or in the E.R., were you asked in what environment you live in? Our guess is that you weren't. Not because the doctor missed or forgot this question, but because this question is not on the standard protocol questionnaire. Environmental exposure can trigger a disease such as cancer or autism. The things that need to be measured are: chemical components of air pollution, allergens, noise, UV intensity, lead, asbestos, radiation, and human pathogens. All this data must be captured and added to the training set, in order for the IDSS to perform better.

Another matter that needs to be addressed is Robotic surgery, O'Sullivan et al. [163]. We do not know whether you are a fan of the series *Grey's Anatomy* or *The Resident*, or any other new medical drama show, but for sure you have heard about the *da Vinci* robot from Intuitive Surgical. The Food and Drug Administration (FDA) cleared *da Vinci* for surgery in 2000.

The *da Vinci* robot is a combo box that contains a console and four robotic arms. The surgeon is able to move the robotic arms through the console. The arms can hold objects, and act as scalpels, scissors, bovies or graspers. The *da Vinci* has better motion than the human hand, and it can also reduce the tremor and refine the surgeon's hand movements. The *da Vinci* system allows the surgeon to operate while sitting, opposed to the conventional laparoscopy. A procedure using *da Vinci* is minimally invasive, it has less pain associated with it, less blood loss and thus less need for blood transfusions.

Not all the robotic surgeries involve the use of AI, but some of them use computer vision to identify distances, specific body parts, etc. In the future using AI the robotic surgery might determine the margins of a tumor, separating normal tissue from malignant tissue, etc.

If it can be autonomous driving, why not even autonomous robot surgery? The Smart Tissue Autonomous Robot (STAR) performed a surgery on a pig's small intestines using its own vision tools and artificial intelligence. The mind-blowing fact is that STAR outperformed the human surgeons that were given the same task; the robot's stitches were more consistent and resistant to leaks, Shademan et al. [159]. STAR performed the surgery both in *ex vivo* tissue in the lab and on *in vivo* tissue on an anesthetized pig. In 40% of the trials, human intervention was needed.

Current medical systems all over the world need to be improved. Medical care is expensive and these costs must be reduced. AI plays and will play an important role in healthcare, whether it will regard early diagnosis and tailored treatment, protection from insurance fraud or reduction of costs.

One question arises though. When you start driving a car on a trip and you use a navigation app, the app tells you that the driver holds the whole responsibility. So,

who holds the responsibility when it comes to IDSS in healthcare? If a machine makes a fatal or near fatal error who pays for it? The hospital that bought the system? The manufacturer? The doctors that did not use it properly?

No matter the advances in the field, the old saying stands still: *Medicine is not mathematics*. The sixth sense of the human doctor shall never be replaced by a machine, no matter how that machine was trained, which algorithm it has been implemented with, and its statistical validation results. We once again strongly state that intelligent systems in healthcare will provide at best a recommendation.

References

1. Turing, A.M.: Computing machinery and intelligence. *Mind* **59**, 433–460 (1950)
2. Oppy, G., Dowe, D: The Turing Test. The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.) (2016). <https://plato.stanford.edu/archives/spr2016/entries/turing-test/>
3. Aron, J.: Forget the Turing test – there are better ways of judging AI. *New Sci* (2015). <https://www.newscientist.com/article/dn28206-forget-the-turing-test-there-are-better-ways-of-judging-ai/>
4. Ball, P.: The truth about the turing test. *BBC Future*, 24 July 2014
5. Lewis, T.: Brainsy machines need an updated IQ Test, experts say. *Live Science* (2014). <https://www.livescience.com/47296-turing-test-needs-an-update.html>
6. Mols, B.: Judge Weighs in on Chatbot’s turing test performance. *Commun. ACM* (2014). <https://cacm.acm.org/news/175744-judge-weighs-in-on-chatbotsturing-test-performance/fulltext>
7. Searle, J.: Minds, Brains, and Programs. *Behav. Brain Sci.* **3**, 417–424 (1980)
8. Copeland, B.J. (ed.): *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*. Oxford University Press, Oxford (2004)
9. Padhy, N.P.: *Artificial Intelligence and Intelligent Systems*, 1st edn. Oxford University Press, Oxford (2005)
10. Shin, Y, Xu, C.: *Intelligent Systems: Modeling, Optimization, and Control*. CRC Press (2008)
11. Hopgood, A.: *Intelligent Systems for Engineers and Scientists*, 3rd edn. CRC Press (2011)
12. Grosan, C., Abraham, A.: *Intelligent Systems: A Modern Approach* (Intelligent Systems Reference Library Series). Springer (2011)
13. Wilamowski, B.J., David Irwin, J.D. (eds.): *Intelligent Systems*, 2nd edn. CRC Press (2011)
14. Pap, E. (ed.): *Intelligent Systems: Models and Applications*. Topics in Intelligent Engineering and Informatics Series. Springer, Heidelberg (2013)
15. Kryszkiewicz, M., Appice, A., Slezak, D., Rybinski, H., Skowron, A., Ras, Z.W. (eds.): *Foundations of Intelligent Systems*. Springer (2017)
16. Martínez de Pisón, F.J., Urraca, R., Quintián, H., Corchado, E. (eds.): *Hybrid Artificial Intelligent Systems*. Springer, Cham (2017)
17. Bi, X., Kapoor, Y., Bhatia, S. (eds.): *Intelligent Systems and Applications*. Springer (2018)
18. Bai, Y., Zhuang, H., Wang, D. (eds.): *Advanced Fuzzy Logic Technologies in Industrial Applications*. Springer, London (2006)
19. Melin, P., Castillo, O., Aguilar, R., Kacprzyk, J, Pedrycz, W. (eds.): *Foundations of Fuzzy Logic and Soft Computing*. Springer, Heidelberg (2007)
20. Feng, G.: *Analysis and Synthesis of Fuzzy Control Systems: A Model-Based Approach*. CRC Press (2010)

21. Siddique, N.: *Intelligent Control: A Hybrid Approach Based on Fuzzy Logic Neural Networks and Genetic Algorithms*. Springer, Dordrecht (2014)
22. Mendel, J., Hagaras, H., Tan, W.-W., Melek, W., Ying, H.: *Introduction to Type-2 Fuzzy Logic Control: Theory and Applications*. Wiley, Hoboken (2014)
23. Pareek, D.: *Business Intelligence for Telecommunications*. Auerbach (Taylor & Francis), New York (2007)
24. Casillas, J., Martínez-López, F., Juan Manuel Corchado Rodríguez, J.M.C. (eds.): *Management Intelligent Systems: First International Symposium*. Springer, Cham (2012)
25. Aufaure, M-A., Zimányi, E. (eds.): *Business Intelligence*. Springer, Heidelberg (2012)
26. Sherman, R.: *Business Intelligence Guidebook: From Data Integration to Analytics*. Morgan Kaufmann (Elsevier) (2015)
27. Keedwell, E., Narayanan, A.: *Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems*. Wiley, Chichester (2005)
28. Sperschneider, V.: *Bioinformatics: Problem Solving Paradigms*. Springer (2008)
29. Masulli, F., Peterson, L., Tagliaferri, R. (eds.): *Computational Intelligence Methods for Bioinformatics and Biostatistics*. Springer (2010)
30. Huang, D-S., Han, K., Gromiha, M. (eds.): *Intelligent Computing in Bioinformatics*. Springer, Cham (2014)
31. Angelini, C., Rancoita, P., Rovetta, S. (eds.): *Computational Intelligence Methods for Bioinformatics and Biostatistics*. Springer (2016)
32. Silverman, B., Jain, A., Ichalkaranje, A, Jain, L. (eds.): *Intelligent Paradigms for Healthcare Enterprises: Systems Thinking*. Springer, Heidelberg (2005)
33. Sordo, M., Vaidya, S., Jain, L. (eds.): *Advanced Computational Intelligence Paradigms in Healthcare - 3*. Springer, Heidelberg (2008)
34. McClean, S., Millard, P., El-Darzi, E., Nugent, C. (eds.): *Intelligent Patient Management*. Springer, Heidelberg (2009)
35. Brahmam, S., Jain, L. (eds.): *Advanced Computational Intelligence Paradigms in Healthcare 5*. Springer, Heidelberg (2010)
36. Oldenburg, J., Chase, D., Christensen K., Tritle, B. (eds.): *Engage! Transforming Healthcare Through Digital Patient Engagement*. HIMSS Press (2013)
37. Buckland, M.: *Programming Game AI by Example*. Jones & Bartlett Learning (2004)
38. Baba, N., Handa, H. (eds.): *Advanced Intelligent Paradigms in Computer Games*. Springer, Heidelberg (2007)
39. Millington, I., Funge, J.: *Artificial Intelligence for Games*, 2nd edn. CRC Press (Taylor & Francis), New York (2009)
40. Kirby, N.: *Introduction Game AI*. Cengage Learning PTR (2010)
41. Mandziuk, J.: *Knowledge-Free and Learning-Based Methods in Intelligent Game Playing*. Springer, Heidelberg (2010)
42. Nourani, C.: *Intelligent Multimedia Computing Science: Business Interfaces, Wireless Computing, Databases and Data Mining*. American Scientific Publishers, (2005)
43. Hassanien, A-E., Abawajy, J., Abraham, A., Hagaras, H. (eds.): *Pervasive Computing: Innovations in Intelligent Multimedia and Applications*. Springer, London (2009)
44. Chen, C.W., Li, Z., Shiguo Lian, S. (eds.): *Intelligent Multimedia Communication: Techniques and Applications*. Springer, Heidelberg (2010)
45. Yan, L., Ma, Z. (eds.): *Intelligent Multimedia Databases and Information Retrieval: Advancing Applications and Technologies*. IGI Global (2011)
46. Atrey, P., Kankanhalli, M, Cavallaro, A. (eds.): *Intelligent Multimedia Surveillance: Current Trends and Research*. Springer, Heidelberg (2013)
47. Jacko, J. (ed.): *Human-Computer Interaction Handbook*, 3rd edn. CRC Press (2012)
48. Kurosu, M. (ed.): *Human-Computer Interaction. Novel User Experiences, Part III*. Springer (2016)
49. Kurosu, M. (ed.): *Human-Computer Interaction. User Interface Design, Development and Multimodality, Part I*. Springer (2017)

50. Yamamoto, S. (ed.): *Human Interface and the Management of Information: Information, Knowledge and Interaction Design, Part I*. Springer (2017)
51. Kendal, S., Creen, M.: *An Introduction to Knowledge Engineering*. Springer, London (2007)
52. Virvou, M., Matsuura, K. (eds.): *Knowledge-Based Software Engineering*. IOS Press (2012)
53. Kravets, A., Shcherbakov, M., Kultsova, M., Iijima, T. (eds.): *Knowledge-Based Software Engineering*. Springer, Cham (2014)
54. North, K., Kumta, G.: *Knowledge Management: Value Creation Through Organizational Learning*. Springer, New York (2014)
55. Bolisani, E., Handzic, M. (eds.): *Advances in Knowledge Management: Celebrating Twenty Years of Research and Practice*. Springer, Cham (2015)
56. Rajput, N., Nanavati, A.A.: *Speech in Mobile and Pervasive Environments*. Wiley (2012)
57. Schmitt, A., Minker, W.: *Towards Adaptive Spoken Dialog Systems*. Springer, New York (2013)
58. Bjorn, S., Batliner, A.: *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. Wiley, Hoboken (2013)
59. Yu, D., Deng, L.: *Automatic Speech Recognition: A Deep Learning Approach*. Springer, London (2015)
60. Zakaria Kurdi, M.Z.: *Natural Language Processing and Computational Linguistics 1: Speech, Morphology and Syntax*. Wiley, New York (2016)
61. Dayan, P., Abbott, L.: *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems (Computational Neuroscience Series)*. MIT Press, Cambridge (2005)
62. Wolpaw, J., Wolpaw, E.W. (eds.): *Brain-Computer Interfaces: Principles and Practice*, 1st edn. Oxford University Press, Oxford (2012)
63. Rao, R.: *Brain-Computer Interfacing: An Introduction*, 1st edn. Cambridge University Press, Cambridge (2013)
64. Allison, B.Z., Dunne, S., Leeb, R., Del, R. Millán, J., Nijholt, A. (eds.): *Towards Practical Brain-Computer Interfaces: Bridging the Gap from Research to Real-World Applications*. Springer, Heidelberg (2013)
65. Clerc, M., Laurent Bougrain, L., Lotte, F. (eds.): *Brain-Computer Interfaces 1: Methods and Perspectives (Cognitive Science)*, 1st edn. Wiley-ISTE (2016)
66. Bekey, G.: *Autonomous Robots*. MIT Press, Cambridge (2005)
67. Liu, D., Wang, L., Tan, K.-C. (eds.): *Design and Control of Intelligent Robotic Systems*. Springer, Heidelberg (2009)
68. Siegwart, R.: *Introduction to Autonomous Robots*. MIT Press (2011)
69. Vargas, P., Di Paolo, E., Inman Harvey, I., Phil Husbands, P. (eds.): *The Horizons of Evolutionary Robotics*. MIT Press, Cambridge (2014)
70. Craig, J.: *Introduction to Robotics Mechanics and Control*. Pearson (2018)
71. Adeli, H., Xiaomo Jiang, X.: *Intelligent Infrastructure: Neural Networks, Wavelets, and Chaos Theory for Intelligent Transportation Systems and Smart Structures*. CRC Press Taylor & Francis, Boca Raton (2008)
72. Gentile, G., Noekel, K. (eds.): *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*. Springer, Cham (2016)
73. Hasan, S.F., Siddique, N., Chakraborty, S.: *Intelligent Transportation Systems: 802.11-based Vehicular Communications*, 2nd edn. Springer (2018)
74. Schaefer, G., Hassanién, A., Jiang, J.: *Computational Intelligence in Medical Imaging: Techniques and Applications*, 1st ed. Chapman and Hall/CRC Press, London (2009)
75. Wöhler, C.: *3D Computer Vision: Efficient Methods and Applications*. Springer, Heidelberg (2009)
76. Tavares, J., Jorge, R.N. (eds.): *Computational Vision and Medical Image Processing*. Springer, Dordrecht (2011)
77. Farinella, G.M., Battiato, S., Cipolla, R. (eds.): *Advanced Topics in Computer Vision*. Springer, London (2013)

78. Wilson, R., Hancock, E., Bors, A., Smith, W. (eds.): *Computer Analysis of Images and Patterns*. Springer, New York (2013)
79. Boden, M.: *A History of Cognitive Science*, vol. 2. Oxford University Press, Oxford (2006)
80. Edelman, S.: *Computing the Mind*. Oxford University Press, Oxford (2008)
81. Friedenberg, J.: *Artificial Psychology: The Quest for What It Means to Be Human*, 1st edn. Psychology Press (2008)
82. Rescorla, M.: The Computational Theory of Mind. *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), (Zalta, E. ed.). Springer (2017). <https://plato.stanford.edu/archives/spr2017/entries/computational-mind/>
83. Nath, V., Levinson, S.: *Autonomous Military Robotics*. Springer (2014)
84. Bergen, P., Rothenberg, D. (eds.): *Drone Wars: Transforming Conflict, Law, and Policy*. Cambridge University Press, New York (2014)
85. Galliot, J.: *Military Robots: Mapping the Moral Landscape (Military and Defense Ethics)*, 1st edn. Routledge (2015)
86. Wilson, D. (ed.): *Popular Mechanics Robots: A New Age of Bionics, Drones & Artificial Intelligence*. Hearst (2015)
87. Koeppe, M., Schaefer, G., Abraham, A. (eds.): *Intelligent Computational Optimization in Engineering: Techniques & Applications*. Springer, Heidelberg (2011)
88. Prokopczuk, M.: *Energy Pricing Models: Recent Advances, Methods, and Tools*. Palgrave Macmillan (2014)
89. Kahraman, C., Cevik Onar, S. (eds.): *Intelligent Techniques in Engineering Management: Theory and Applications*. Springer, Cham (2015)
90. Karampelas, P., Ekonomou, L. (eds.): *Electricity Distribution: Intelligent Solutions for Electricity Transmission and Distribution Networks*. Springer, Heidelberg (2016)
91. Batty, M.: *The New Science of Cities*. MIT Press, Cambridge (2013)
92. Townsend, A.: *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. W. W. Norton & Company (2013)
93. Pethuru R., Raman, A.: *Intelligent Cities: Enabling Tools and Technology*. CRC Press (Taylor & Francis), Boca Raton (2015)
94. Komninos, N.: *The Age of Intelligent Cities: Smart Environments and Innovation-for-all Strategies*. Routledge, New York (2015)
95. Song, H., Srinivasan, R., Sookoor, T., Jeschke, S.: *Smart Cities: Foundations, Principles, and Applications*. Wiley, Hoboken (2017)
96. McEwen, A., Cassimally, H.: *Designing the Internet of Things*, 1st edn. Wiley, Hoboken (2013)
97. Holler, J., Tsiatsis, V., Mulligan, C., Avesand, S., Karnouskos, S., Boyle, D.: *From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence*, 1st edn. Academic Press (2014)
98. Greengard, S.: *The Internet of Things*. MIT Press (2015)
99. Maciej, K.: *Building the Internet of Things: Implement New Business Models, Disrupt Competitors, Transform Your Industry*. Wiley, New York (2016)
100. Sathi, A.: *Cognitive Internet of Things: Collaboration to Optimize Action*. Palgrave Macmillan, New York (2016)
101. Corea, F.: *A Brief History of Artificial Intelligence*. © 2017 KDnuggets (2017). <https://www.kdnuggets.com/2017/04/brief-history-artificial-intelligence.html>
102. Walter, E., Lazic-Peric, A.: *Distribution profile and efficiency of the European pharmaceutical full-line wholesaling sector (Technical Report)*. Institute for Pharmoeconomic Research Vienna, February, X (2017). <http://www.girp.eu/files/GIRP-IPF%20Study%202016.pdf>
103. U.S. Bureau of the Census, *Total Revenue for Medical and Diagnostic Laboratories, Establishments Subject to Federal Income Tax [REV6215TAXABL144QNSA]*, retrieved from FRED, Federal Reserve Bank of St. Louis, 23 November 2017. <https://fred.stlouisfed.org/series/REV6215TAXABL144QNSA>

104. Feigenbaum, E., Buchanan, B.: DENDRAL and Meta-DENDRAL roots of knowledge systems and expert system applications. *Artif. Intell.* **59**, 233–240 (1993)
105. Shortliffe, E.: *Computer-Based Medical Consultations: MYCIN*. Elsevier (1976)
106. Sinsky, C., Colligan, L., Li, L., Prgomet, M., Reynolds, S., Goeders L., Westbrook, J., Tutty, M., Blike, G.: Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann. Intern. Med.* **165**(11), 753–760 (2016). <https://doi.org/10.7326/m16-0961>
107. Jaiprakash, A., Roberts, J., Crawford, R.: Robots in health care could lead to a doctorless hospital). *The Conversation—Copyright © 2010–2017. The Conversation Trust (UK) Limited*, 8 February 2016. <https://theconversation.com/robots-in-health-care-could-lead-to-a-doctorless-hospital-54316>. Accessed 3 Dec 2017
108. Schwab, K.: *The Fourth Revolution*. World Economic Forum (2016)/Crown Business (2017)
109. Bellman, R.E.: *Dynamic Programming*. Princeton University Press, Princeton (1957)
110. IBM: Preparing for the future of artificial intelligence. © Copyright IBM Corporation (2016). http://research.ibm.com/cognitive-computing/cognitive-horizons-network/downloads/IBM_Research_CSOC_OSTP_Response%20v2.pdf. Accessed 12 Dec 2017
111. Trim, C.: Natural Language Understanding of Unstructured Data. IBM developerWorks@. https://www.ibm.com/developerworks/community/blogs/nlp/entry/natural_language_understanding_of_unstructured_data?lang=en. Accessed 13 Dec 2017
112. Gorunescu, F.: *Data Mining: Concepts Models and Techniques*. Springer, Heidelberg (2011)
113. James, G., Witten, D., Hastie, T., Tibshirani, R.: *An Introduction to Statistical Learning (With Application in R)*. Springer, New York (2013)
114. Sugiyama, M.: *Statistical Machine Learning*. Morgan Kaufmann (Elsevier) (2016)
115. Yang, Z.R.: *Machine Learning Approaches to Bioinformatics*. World Scientific (2010)
116. Bayes, T.: An essay towards solving a problem in the doctrine of chances, *Phil. Trans.* **53**, 370–418 (1763). <https://doi.org/10.1098/rstl.1763.0053>
117. Haykin, S.: *Neural Networks. A Comprehensive Foundation*, 2nd edn. Prentice Hall (1999)
118. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Wang, Y., Dong, Q., Shen, H., Wang, Y.: Artificial intelligence in healthcare: past, present and future. *BMJ Stroke Vasc. Neurol.* 1–14, e000101 (2017). <https://doi.org/10.1136/svn-2017-000101>
119. Stoean, C. Stoean, R.: *Support vector machines and evolutionary algorithms doe classification. Single or Together?* Springer, Cham (2014)
120. Murty, M.N., Raghava, R.: *Support Vector Machines and Perceptrons. Learning, Optimization, Classification, and Application to Social Networks*. Springer (2016)
121. Darwin, C.: *On the origin of species* (1859)
122. Rokach L., Maimon O.: Clustering methods. In: Maimon, O., Rokach, L. (eds.) *Data Mining and Knowledge Discovery Handbook*. Springer, Boston (2005)
123. Mirkin, B.: *Clustering: A Data Recovery Approach*, 2nd edn. Chapman & Hall/CRC Press (2013)
124. Bellman, R.E.: *Adaptive Control Processes*. Princeton University Press, Princeton, NJ (1961)
125. Chen, L.: Curse of dimensionality. In: Liu, L., Ozsu, M.T. (eds.) *Encyclopedia of Database Systems*. Springer, Boston (2009)
126. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**, 67 (1997)
127. Pepe, M.S.: *The Statistical Evaluation of Medical Tests For Classification and Prediction*. Oxford University Press, Oxford (2004)
128. Zweig, M.H., Campbell, G.: Receiver-operating characteristic (ROC) – plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* **39**(8), 561–577 (1993)
129. Zur, R.M., Pesce, L.L., Jiang, Y.: The effect of two priors on Bayesian estimation of “Proper” binormal ROC curves from common and degenerate datasets. *Acad. Radiol.* **17**(8), 969–979 (2010)

130. Roemer, V.M., Walden, R.: Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios for electronic fetal heart rate monitoring using new evaluation techniques. *Z. Geburtshilfe Neonatol.* **214**(3), 108–118 (2010)
131. Jin, H., Lu, Y.: A non-inferiority test of areas under two parametric ROC curves. *Contemp. Clin. Trials* **30**(4), 375–379 (2009)
132. Lasko, T.A., Bhagwat, J.G., Zou, K.H., Ohno-Machado, L.: The use of receiver operating characteristic curves in biomedical informatics. *J. Biomed. Inform.* **38**(5), 404–415 (2005)
133. Fawcett, T.: An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**, 861–874 (2006)
134. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29036 (1982)
135. Macilwaine, S.W., Lond, L.R.C.P, Eng, M.R.C.S.: What is a disease? *Br. Med. J.* 1703–1704 (1900)
136. Merriam-Webster's Collegiate Dictionary, 11th edn. Springfield, MA; Merriam-Webster Inc (2003)
137. Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordonez, G.R., Bignell, G.R., Ye, K., Alipaz, J., Bauer, M.J., Beare, D., Butler, A., Carter, R.J., Chen, L., Cox, A.J., Edkins, S., Kokko-Gonzales, P.I., Gormley, N.A., Grocock, R.J., Haudenschild, C.D., Hims, M.M., Hames, T., Jia, M., Kingsbury, Z., Leroy, C., Marshall, J., Menzies, A., Mudie, L.J., Ning, Z., Royce, T., Schulz-Trieglaff, O.B., Spiridou, A., Stebbings, L.A., Szajkowski, L., Teagye, J., Williamsom, D., Chin, L., Ross, M.T., Campbell, P.J., Bentley, D.R., Futreal, P.A., Stratton, M.R.: A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010)
138. McDermott, U., Downing, J.R., Stratton, M.R.: Genomics and the continuum of cancer care. *N. Engl. J. Med.* **364**(4), 340–350 (2011)
139. Miotto, R., Li, L., Kidd, B.A., Dudley, J.T.: Deep patient: an unsupervised representation to predict the future of patients from electronic health records. *Sci. Rep.* **6** (2016). <https://doi.org/10.1038/srep26094>
140. Poplin, R., Varadarajan, A., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., Webster, D.R.: Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018). <https://doi.org/10.1038/s41551-018-0195-0>
141. Sun, J., McNaughton, C., Zhang, P., Perer, A., Gkoulalas-Divanis, A., Denny, J.C., Kirby, J., Lasko, T., Saip, A., Malin, B.A.: Predicting changes in hypertension control using electronic health records from a chronic disease management program. *J. Am. Med. Inform. Assoc.* **21**(2), 337–344 (2013)
142. Taballali, P., Razavi, M., Pahlevan, N.M.: Artificial intelligence estimation of carotid-demoral pulse wave velocity using carotid waveform. *Sci. Rep.* **8** (2018). <https://doi.org/10.1038/s41598-018-19457-0>
143. Schwartz, P.J.: Stramba-Badiale, M., Crotti, L., Pedrazzini, M., Besana, A., Bosi, G., Gabbarini, F., Goulene, K., Insolia, R., Mannarino, S., Mosca, F., Nespola, L., Rimini, A., Rosati, E., Salice, P., Spazzolini, C.: Prevalence of the congenital long-QT syndrome. *Circulation* **120**(18), 1761–1767 (2009). <https://doi.org/10.1161/circulationaha.109.863209>
144. Neocleous, A.C., Syngelaki, A., Nicolaides, K.H., Schizas, C.N.: Two-stage approach for risk estimation of fetal trisomy 21 and other aneuploidies using computational intelligence systems. *Ultrasound Obstet. Gynecol.* (2017). <https://doi.org/10.1002/uog.17558>
145. Pomprapa, A., Muanghong, D., Kony, M., Leonhardt, S., Pickerodt, P., Tjarks, O., Schwaiberger, D., Lachmann, B.: Artificial intelligence for closed-loop ventilation therapy with hemodynamic control using the open lung concept. *Int. J. Intell. Comput. Cybern.* **8**(1), 50–68, <https://doi.org/10.1108/IJICC-05-2014-0025> (2015)
146. Knaus, A., Pantel, J.T., Pendziwiat, M., Hajjir, N., Zhao, M., Hsieh, T.C., Schubach, M., Gurovich, Y., Fleischer, N., Jager, M., Kohler, S., Muhle, H., Korff, C., Moller, R.S., Bayat, A., Calvas, P., Chassaing, N., Warren, H., Skinner, S., Louie, R., Evers, C., Bohn, M., Christen, H.J., van den Born, M., Obersztyn, E., Charzewska, A., Endziniene, M., Kortum,

- F., Brown, N., Robinson, P.N., Schelhaas, H.J., Weber, Y., Helbig, I., Mundlos, S., Horn, D., Krawitz, P.M.: Characterization of glycosylphosphatidylinositol biosynthesis defects by clinical features flow cytometry, and automated image analysis. *Genome Med.* **10**(1). (2018). <https://doi.org/10.1186/s13073-017-0510-5>
147. Kvedar, J.C., Fogel, A.L., Elenko, E., Zohar, D.: Digital medicine's march on chronic disease. *Nat. Biotechnol.* **34**, 239–246 (2016)
 148. Abramoff, M.D., Folk, J.C., Han, D.P., Walker, J.D., Williams, D.F., Russell, S.R., Massin, P., Cochener, B., Gain, P., Tang, L., Lamard, M., Moga, D.C., Quelled, G., Niemeijer, M.: Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol.* **131**, 351–357 (2013)
 149. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Winder, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J., Webster, D.R.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016)
 150. Winters, B., Custer, J., Galvagno Jr., S.M., Colantuoni, E., Kapoor, S.G., Lee, H., Goode, V., Robinson, K., Nakhasi, A., Pronovost, P., Newman-Toker, D.: Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. *BMJ Qual. Saf.* **21**(11) (2012). <http://dx.doi.org/10.1136/bmjqs-2012-000803>
 151. JASON report: Artificial Intelligence for Health and Health Care. JSR-17-Task-002 (2017). https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf
 152. Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.M., Zietz, M., Hoffman, M.M., Xie, Q., Rosen, G.L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A.E., Shrikumar, A., Xu, J., Cofer, E.M., Lavender, C.A., Turaga, S.C., Alexandari, A.M., Lu, Z., Harris, D.J., DeCaprio, D., Qi, Y., Jundaje, A., Peng, Y., Wiley, L.K., Segler, M.H.S., Boca, S.M., Swamidass, J., Huang, A., Gitter, A., Greene, C.S.: Opportunities and obstacles for deep learning in biology and medicine. *R. Soc. Interface* (2018). <http://dx.doi.org/10.1101/142760>
 153. Greene, C.S., et al.: Opportunities and obstacles for deep learning in biology and medicine. (2017). <http://dx.doi.org/10.1101/142760>
 154. Lazer, D., Kennedy, R., King, G., Bospignani, A.: The parable of Google flu: traps in Big data analysis. *Science* **343**, 1203 (2014)
 155. Doerr, M., Truong, A.M, Bot, B.M., Suver, C., Mangravite, L.M.: Formative evaluation of participant experience with mobile eConsent in the app-mediated Parkinson mPower Study: a mixed methods study. *JMIR mHealth uHealth* **5**(2) (2017)
 156. Silva, B.M.C., et al. Mobile-health: a review of current state in 2015. *J. Biomed. Inform.* **56**, 265–272 (2015)
 157. Yang, Q., Khoury, M.J., Friedman, J.M., Little, J., Flanders, W.D.: How many genes underlie the occurrence of common complex diseases in the population? *Int. J. Epidemiol.* **34**, 1129–1137 (2005)
 158. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., Hirschhorn, J.N.: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev.* **9**, 356 (2008)
 159. Shademan, A., Decker, R.S., Opfermann, J.D., Leonard, S., Krieger, A., Kim, P.C.W.: Supervised autonomous robotic soft tissue surgery. *Sci. Transl. Med.* **8**(337), 337–364 (2016). <https://doi.org/10.1126/scitranslmed.aad9398>
 160. Lieto, A., Cruciani, M. (eds.): Artificial intelligence and cognition. In: Proceedings of 1st International Workshop AIC 2013, Turin, Italy (2013). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.5575&rep=rep1&type=pdf>
 161. Spiegeleire, S., Maas, M., Sweijts, T.: Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers. The Hague Centre for Strategic Studies (2017)

162. Gobetto, M.: *Operations Management in Automotive Industries: From Industrial Strategies to Production Resources Management, Through the Industrialization Process and Supply Chain to Pursue Value Creation*. Springer, Dordrecht (2014)
163. O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Holzinger, K., Holzinger, A., Sajid, M.I., Ashrafian, H.: Legal, regulatory and ethical frameworks or standards for AI and autonomous robotic surgery. *The Int. J. Med. Robot. Comput. Assist. Surg.* (2018). <https://doi.org/10.1002/rcs.1968>