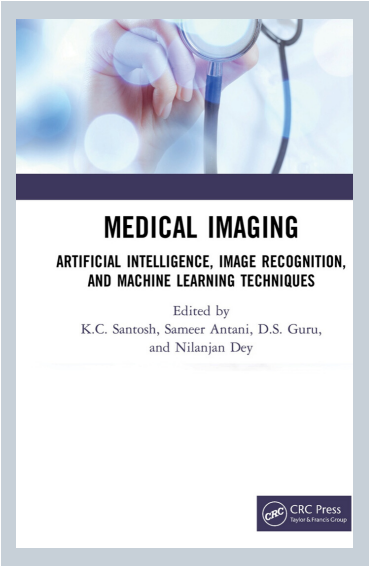
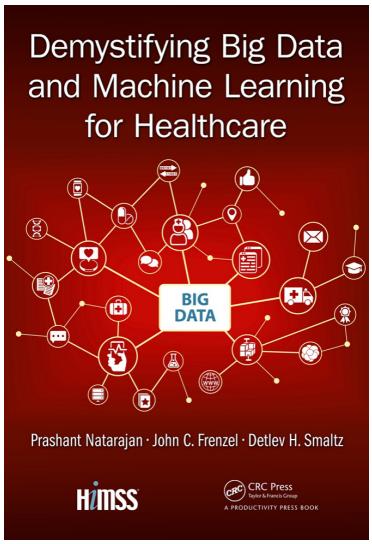
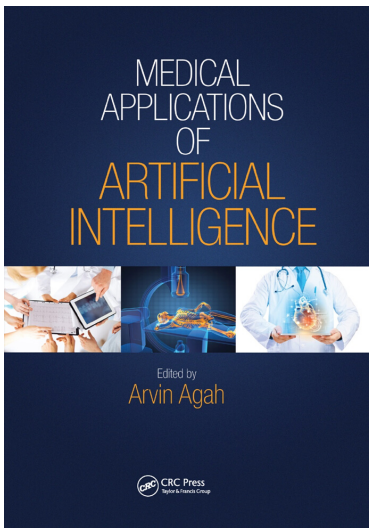
AI in Healthcare

AI is being incorporated into many industries and healthcare is no exception. Discover the current benefits and future possibilites of AI in healthcare.







Contents

The Role of Artificial Intelligence (AI) in Medical Imaging: General Radiologic and Urologic Applications

Diboro Kanabolo and Mohan S. Gundeti

2.1 Introduction to Artificial Intelligence (AI)

Artificial Intelligence (AI) permeates many different sectors of industry, including financial/banking, commerce, social media, and health. Several domains of technology are in active development. These include *audio processing*: speech recognition, music/voice identification; *computer vision*: facial or object recognition; *graph analytics*: as in film recommendations, or mapping directions; *language processing*: for example, machine translation, and machine-based question/answer; as well as *time series*, as in stock forecasting. More sophisticated software programs may use any combination of these domains. A popular example to date is autonomous driving, which may use graph analytics, computer vision, and audio processing [1].

2.1.1 Terminology

The vocabulary below will aid in the understanding of details discussed in the discussion to follow. Of note, this brief glossary will only continue to enlarge with the expansion of AI technology.

In this section, we will define the following terms in alphabetical order: artificial intelligence, computer-aided detection, computer-aided diagnosis, computer-aided triage, classification, deep learning, detection, machine learning, model, neural network, representation learning, segmentation, supervised learning, testing, training, transfer learning, unsupervised learning, and validation.

Artificial intelligence: defined in Merriam Webster's dictionary as the branch of computer science that allows a machine the capability to imitate intelligent human behavior [2]. Two types of AI exist: the first is artificial general intelligence, in which a computer may hypothetically mimic the day-to-day actions of human beings. The second, more appropriate definition to be used in the ensuing discussion is artificial narrow intelligence, in which a computer engages in particular tasks.

- Computer-aided detection (CADe): computer recognition of areas of concern necessitating further evaluation. It does not provide diagnosis [3].
- Computer-aided diagnosis (CADx) is simply a computer's ability to state multiple diagnoses (a differential) or a singular diagnosis, to be followed up by a provider [3].
- Computer-aided triage (CAT) occurs when a computer studies an image and further prioritizes it for review by radiologist, or simply provides a diagnosis, with or without follow up. Triage systems may be utilized commonly in screening protocols, in which the intention is to efficiently reduce the clinical care provided by clinicians [4].
- Classification utilizes the clustering of data points that are of similar properties.
- Deep learning utilizes the concept of deep neural networks, large logic networks organized into three basic layers: input, hidden, and output layers [5, 6]. The input layer processes large amounts of relevant data. The hidden layer tests and compares new data against pre-existing data, classifying and re-classifying in real time, with particular connections weighted with degrees of influence. The output layer utilizes confidence intervals to determine the best outcome from various predicted outcomes [6].
- Detection involves the act of identifying and localizing a finding in an image [7].

- Machine learning: in general, the term refers to the process of teaching a computer to learn from input data without specific programming. As it applies to *radiomics,* the term is used to describe high throughput extraction of quantitative imaging features with the intent of creating minable databases from radiological images [8].
- Modeling is defined as the structural state of a neural logic network, allowing for the transformation of data input value into output [3].
- Neural network is a model or logic network composed of layers with transmission of sequential data inputs consisting of nodes analogous to those of biological neural networks. Representation learning is the subtype of machine learning in which an algorithm may acquire those features that enable it to classify data [3].
- Segmentation is the process of delineating the boundary of a particular finding within an image.
- Supervised learning involves inferring a function from inputs and outputs (training data). This allows for output prediction after any novel input [3].
- Testing involves evaluating the performance of a neural logic network [7].
- Training is the process of selecting the ideal parameters of a neural logic network after sequential, repeated adjustments [7].
- Transfer learning may occur when limited data to solve a novel problem are available, but existing data for a problem of close relevance are plentiful [1].
- Unsupervised learning occurs when outputs are inferred in the absence of input data.
- Validation involves the use of a data subset that is distinct from a training set to adjust parameters of the model [7].

2.1.2 Practical Costs

According to an analysis by the accounting firm Price Waterhouse Coopers (PWC), AI is projected to add $15.7 trillion to global gross domestic product (GDP) by 2030 [9]. The true costs of continued development, secondary to the demand of these potential future economic interests, will only spurn increased investments, with the bulk of such investment grounded in the private sector. As of 2016, the United States federal government annually budgeted $1.1 billion in non-classified AI technology [10]. This is contrasted with the $46 billion investment that the top five original equipment manufacturers in the automotive industry spent in 2015 alone [11].

2.2 Artificial Intelligence in Medicine

A large amount of knowledge must be acquired and tailored to solve complex clinical issues. The early beginnings of AI date to the early 1970s with the explosion of biomedical applications, catalyzed by various seminal conferences around the country. Today, there is potential in medicine to harness "big data," utilizing the large quantities of information generated daily in various settings. Clinical practice is now shifting from episodic analysis of disparate datasets to algorithms relying on consistently updated datasets for improved prediction of patient outcomes [12]. The scope and depth of applications is vast. Machine learning algorithms have already been employed to predict the risk for cardiac arrest in infants, and computer visualization has been utilized for various applications, from cancer detection in radiology to detecting health indicators of mental fatigue [13].

A vast amount of further investment is needed for AI in the application of medicine broadly, and medical imaging technology specifically. These investments include: training datasets with representative images for the purposes of validation for computerized deep neural networks (with periodic updating as necessary); they also include the interoperability framework necessary to uphold these costs—software algorithms for protocoling the analysis and extraction of data with the aid of specialized vendors (in addition to original digital imaging software companies). Transparent stakeholder collaboration is necessary to ensure unified file format, data representation and database architecture.

In a 1970 manuscript, the prominent Professors in Radiology Kurt Rossmann and Bruce Wiley stated that the "central problem in the study of radiographic image quality is to gain knowledge regarding the effect of physical image quality on diagnosis and not necessarily to design 'high-fidelity' systems" [14]. This is a reasonable goal, as costs of development should simply be the minimal sufficient to diagnose lesions of interest, and this has historically happened with the aid of human interpretation. However, the direct costs of this benefit appear to have the capability of meeting Rossmann and Wiley's requirements. Currently, a hospital may install $1,000 graphics processors in its imaging machines in order to increase capacity up to 260 million images per day [15]. These fixed costs are attractive for the consumer (institutions providing medical care) and may indeed be utilized with increasing frequency in the coming years. The institutions are incentivized to expand their investment in technology, given the estimation that up to 10% of costs in US healthcare spending are attributed to radiological imaging. Other stakeholders, including patients, clinicians, regulatory bodies, and hospital administrations may benefit from education on risks, benefits, and limitations of the technology [7]. For these reasons, the true shared costs of development with respect to medical imaging are elusive.

2.3 Artificial Intelligence in Radiology

The task of the radiologist is plural in nature. Radiologists must be able to recognize and interpret patterns in medical images, as well as to consult with physicians from various fields to direct patients' care. Narrow AI, which enables CADe of disease, has been in continual development for decades. Radiologic imaging tasks, for the purposes of CADe, and CADx disease discovery have undergone rapid development in recent years [3]. The field of radiomics is built on the premise of converting images into data from which useful details may be extracted, for the ultimate purpose of providing medical utility. We aim to provide an overview of the various processes and subfields, relating the roles of artificial intelligence in medical imaging.

For the practicing radiologist, optimal interpretation of disease requires optimal image quality. Image interpretation, especially in the context of subtle disease states, can be limited by various factors, both extrinsic and intrinsic to the radiologist. Extrinsic factors may include: the knowledge base of the radiologist, the clinical history of the patient, and his or her detection/characterization thresholds. Intrinsic factors to image quality are based on target object attributes, specifically the geometry and contrast of the targets, and their background/visibility (i.e. gray-scale appearance and detail [local noise and resolution of the image]) [16].

2.3.1 Extrinsic Factors to Image Interpretation

Diagnostic error accounts for approximately 10% of patient deaths, and between 6 and 17% of adverse events occurring during hospitalization. At a total of ~20 million radiology errors per year, and 30,000 practicing radiologists, this is an average of under 700 errors per practicing radiologist [17]. Errors in diagnosis have been associated with clinical reasoning, including: intelligence, knowledge, age, visual psychiatric affect, physical state (fatigue), clinical history of the patient, and gender (male predilection for risk taking). These factors, and the limited access to radiologic specialists for up to 2/3 of the world, encourage a more urgent role for the use of AI in medical imaging, a huge focus of which is machine learning [17].

2.3.2 Intrinsic Factors to Image Quality

2.3.2.1 Geometry

Intrinsic object attributes are arranged in three general geometries: (1) point, (2) line, and (3) extended targets. Point targets (Figure 2.1) are small, with maximum dimensions typically lower than 1 mm. These may include microcalcifications, calculi, or osteophytes. Line targets may have a variable length depending on clinical context. Examples include spicules, septate lines, and

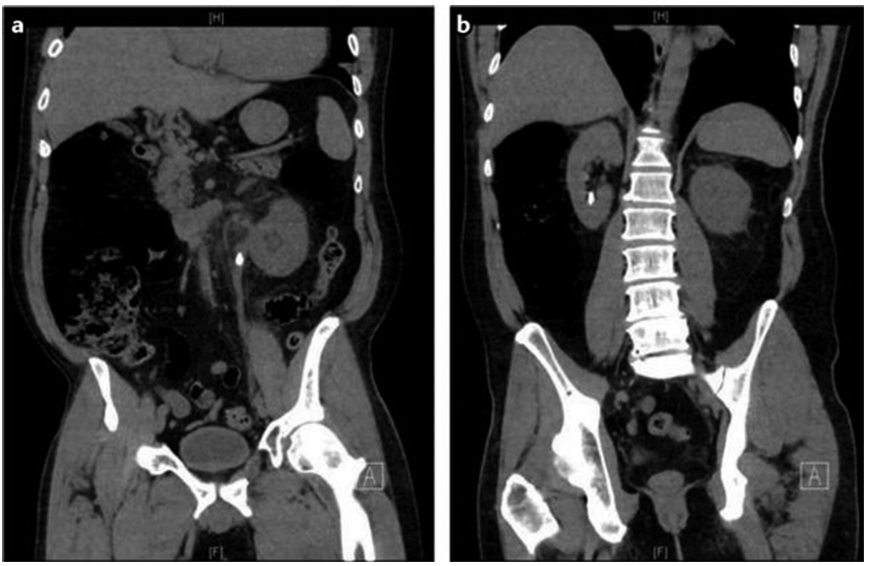


FIGURE 2.1
A coronal demonstration of bilateral 8 mm nephrolithiasis on noncontrast CT, showing point microcalcification. (Reprinted with permission from "An overview of kidney stone imaging techniques" by Brisbane, Bailey, and Sorensen, 2016. *Nature Reviews Urology*, 13, 654–662. 2016.)

lines delineating cortical vs. trabecular bone (Figure 2.2). Extended objects may include tumors, abscesses, and infiltrates [16].

2.3.2.2 Contrast

While in reality, contrast range is on a continuum, we denote here a dichotomy of high and low contrast imaging as per Vyborny, 1997. Examples of high object contrast (Figure 2.3) within the point, line and extended images include a dense microcalcification, tangential pleural calcifications, and calcified granulomas. Fainter microcalcifications, early spicules, and gallstones comprise inherently low contrast point, line, and extended objects, respectively [16]. Note that high object contrast items do not necessarily indicate artificial contrast dye enhancement, though this is a technicality that may vary with the function of other clinical information (renal and liver function, previous allergic responses).

2.3.2.3 Background

The premise of background implies that the radiologic "canvas" on which a particular target finding appears influences its perceptibility. This background may in turn be influenced by not only the gray-scale for which the lesion is displayed, but also the detail of the image itself.

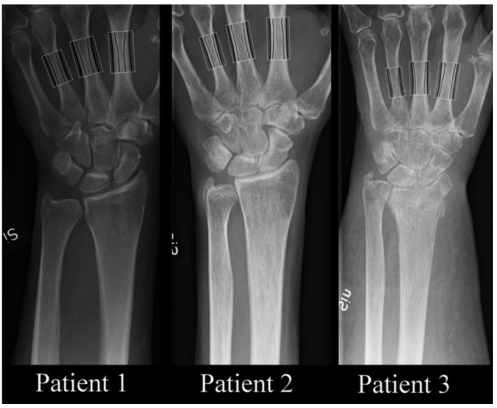


FIGURE 2.2
Delineation of cortical bone thickness (linear). Wrist radiographs used for Dual X-Ray analysis of multiple patients illustrating the cortical bone delineation lines of three patients. (Reprinted with permission from "Digital X-ray radiogrammetry of hand or wrist radiographs can predict hip fracture risk—a study in 5,420 women and 2,837 men" by Wilczek, Kälvesten, Algulin, Beiki, and Brismar. *European Radiology*, 23, 1383–1391. 2012.)

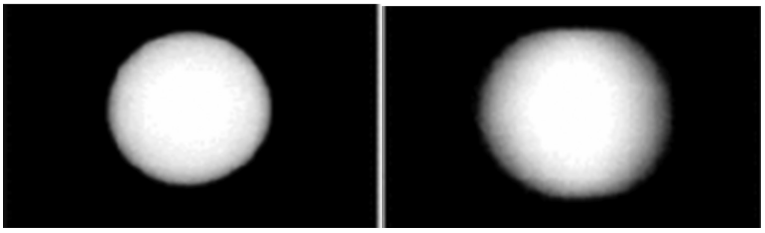


FIGURE 2.3
High vs. low object contrast for circular object. (Cropped and reprinted with permission. "Motion-blur-free video shooting system based on frame-by-frame intermittent tracking" by Inoue, Jiang, Matsumoto, Takaki, and Ishii is licensed under CC BY 4.0.)

Gray-scale components depend on (1) dimensions and anatomical properties, (2) amount of radiation, magnetic resonance, or sound wave frequency to which an area is exposed, and (3) sensitometric characteristics of the recording system [16]. These all determine the relative optical densities of the structures under examination. As stated by Doi et al., 1977, radiomics is challenging in part because of the need to match radiation/magnetic resonance exposure and recording system sensitometric characteristics to the anatomy and targets of interest [18]. When adding the complexity of anatomic structures in immediate proximity, different levels of attenuation may be seen based on the specific qualities of the medium through which the sound waves, radiation, or magnetic resonance are travelling.

Intuitively, details from the perspective of an observer increase in visibility with proximity to an image. As this distance becomes infinitesimally

smaller, the effects of local noise become evident. During analysis of images in clinical practice, it is often the case that noise seen by this blurring is essential to the recognition of a pathological phenomenon. These might include pulmonary edema seen on chest X-ray, or invasion of cerebral ventricle by tumor. Thus, these effects are conscious to the observer. The inherent lack of detail caused by sub-optimal resolution or unsharpness (local noise) of a target object on an image pose two problems: (1) The size of the target may be overestimated. (2) The contrast enhancement seen between the image and background will be sub-optimal. This is especially true for point target lesions. Lack of sharpness of an object can lead to an overestimation in size secondary to line spread function, which causes broadening due to lack of inherent contrast between an object and its background. The quantitative measures of the influence of detail in image contrast have been extensively studied. The effect of local noise on line border overestimation is quantified as the line spread function and has been assessed in angiograms (line targets of high contrast) [19]. The possibility of minimal contrast enhancement between background and image due to lack of delineation may be a point of considerable tragedy. In clinical practice, this may perhaps be true of point lesions in particular. For example, in routine chest X-rays, the possibility of missing a diagnosis of lung cancer in its inchoate stage poses more risk than overestimating a tumor's size on an initial interpretation. Detail is of the utmost importance, and it is for this reason that the automation of image analysis with AI should not exclude human interpretation.

With the intrinsic characteristics of imaging described above, we see that a significant proportion of the technical factors involved in proper automated interpretation of an image can be overcome with proper allocation of resources to their development. For example, research funds for the role of machine learning in point and line spread function analysis for historically low-resolution imaging, such as radiographs or computed tomography.

Still, for common lesions, it appears that CADe may be of benefit. Independent studies suggest that women receiving regularly scheduled mammograms over a period of 10 years have a 50 to 63% chance of receiving a false positive diagnosis. It is worth noting that in up to 33% of occurrences, two or more radiologists inspecting the same image may disagree on the findings seen on mammogram. However, with the same empirical methods, visual pattern recognition software is at least 5–10% more accurate than physicians, further supporting the roles of experience and bias in radiologists' reporting [20].

Further techniques for CADe have been studied extensively, including image preprocessing techniques to enhance the image quality followed by an adaptive segmentation. Other techniques utilized include voting-based combination utilizing different classifiers: Bayesian network, multilayer perception neural networks, and random forest for classification of lung region symmetry. Other developments of interest include: angular relational signature-based chest radiograph for classification of frontal and

lateral X-rays, edge map analysis for abnormal tissue, generalized line histogram technique for rotational artifact detection, and cross-correlation of circle-like elements using a few normalized templates and unsupervised clustering techniques [21–27].

2.3.3 Specific Technical Example of AI in Medical Imaging

The Kohonen self-organized map (KSOM) may be helpful in the understanding of AI's capacity and potential. It is ideal for utilizing artificial bias and sensory experience to enhance accuracy of CADe and CADx when applied to medical imaging. Briefly, it is an artificial neural network (ANN) developed to decrease complexity by representing multidimensional vectors into as little as one dimension. Yet, data is stored in such a way as to maintain topological relationships [28]. This vector quantization may be illustrated with computerized color discrimination in which individual colors are visualized on a spectrum (Figure 2.4, Figure 2.5a). The KSOM is a form of unsupervised feature extraction and classification. Unsupervised feature extraction utilizes images and clinical narrative texts to allow for high throughput application in the analysis of clinical data. The steps in this process include disease detection, lesion segmentation, diagnosis, treatment selection, response assessment via repeat imaging, and using data about a patient to create clinical predictions in regards to potential patient outcomes [29]. In contrast, supervised techniques typically consist of vector pairs and

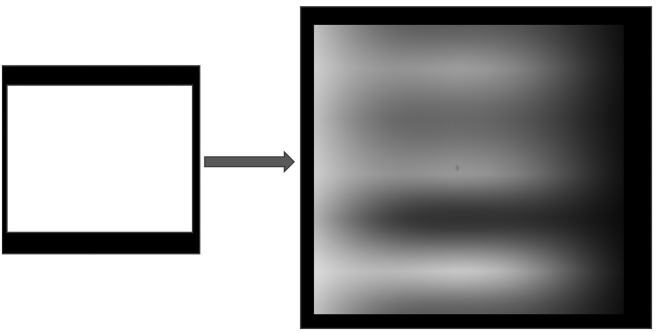


FIGURE 2.4
Color discrimination analogy for illustrating KSOM. The white color spectrum is broken down into its varying wavelengths in a topographically intact manner. It is important to note that regions of the spectrum containing similar properties (wavelengths) are clustered together. Each distinct coordinate in the map has its own component red, green, and blue vectors. When vectors (colors) are randomly selected from a training dataset, all distinct coordinate nodes are examined to assess that which is the most similar to the input vector from the training set. This coordinate will be known as the BMU. A radius for which this best value unit applies is calculated and is known as the BMU neighborhood. The neighboring coordinates are then adjusted (weighted with respect to the BMU), and the process of input vector presentation is repeated. This mechanism is an ideal illustration for understanding unsupervised learning in AI.

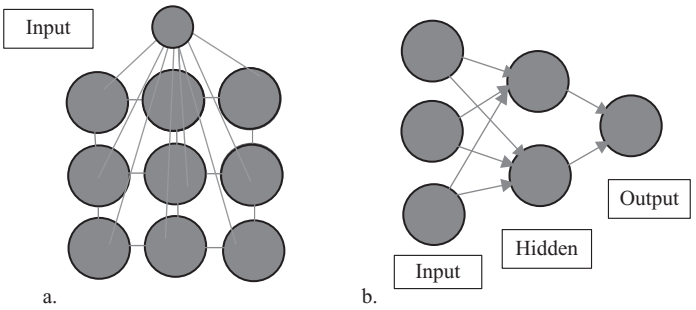


FIGURE 2.5
(a) Kohonen network. Each node of the 3 x 3 lattice represents a vector with a distinct coordinate and contains a particular corresponding weight with respect to the input vectors. (b) An ANN. Each node represents a neuron, and arrows represent specific synapses between them. The input layer feeds are processed by a hidden layer of functions before being expelled into a known result, or output.

include input and output (goal) vectors. When an input is presented to the neural network, the output of the network is compared with this goal vector. When the two contrast, the weights of the networks neurons are adjusted to reduce the error in this output. This process may be seen in neural networking systems with the sole purpose of validating the neural network developed by imaging software of interest. This form of learning, the self-organizing map, can be applied broadly to other neuronal networks.

The KSOM bears a slight similarity to artificial neural networks, in that rather than being a logic network, it is one of a topographical relationship (Figure 2.5b). Training each node is a complex matter. In order for the node to recognize its similarity to a particular element of the input image, it must calculate the weights of its inherent elements. The node with the most similarity to a distinct portion of the input element is known as the best matching unit, or BMU. Counterintuitively, this seems to contradict the necessity of imaging, as each lattice of nodes may contain multiple coordinates with similarly containing weights. To resolve this concern, we must note that the entire lattice may be considered as a neighborhood of elements. At each iteration, the radius of the BMU within the neighborhood—otherwise described as the area within the lattice in which BMU is most similar to the weighted input element—is determined. The weights are then adjusted and the radius of the nodes continues to shorten as more BMUs are described with time.

With an appreciation for this complexity, we must note that the conduction of peer review reads, as in radiology-aided clinical interpretation with hope of CADx, is in its infancy. Before this process is popularized, it is a popular conception that the technology be used for the purposes of screening at the population level. For example, detection of diabetes-induced macular edema is currently not financially practical, as 20 million diabetics in the country would need to be screened for a yield of approximately 5%.

2.4 Urologic Applications

As alluded to previously, the automobile has undergone extreme enhancement. From initial prototypes in the early 20th century to the present, with autonomous vehicle development, we have seen significant evolution. Some posit that the development of the da Vinci robotic surgical system developed by Intuitive Surgical Sunnyvale CA (USA) will be no different. Used throughout a variety of urologic procedures, the da Vinci robot is highly valued in surgeries for deep and narrow fields, and when micro-suturing/fine dissection are needed [30]. Automated surgical development with the robot may one day be of immense interest. The difference to be overcome between driving and surgery presumably is the emergent non-linear changes that may occur in a surgical suite—for example, hemorrhage, bowel perforation, etc. Still, the technology is under development and remains full of potential. An automated robot known as the Smart Tissue Autonomous Robot (STAR) has already been shown capable of performing surgical procedures using ex-vivo porcine tissue and in living pigs. STAR's accuracy is due in part to integration of near-infrared fluorescent imaging, as well as 3D quantitative plenoptic imaging. The former is currently used for intra-operative detection of sentinel lymph nodes, but has several applications for intra-operative imaging. This is due to the minimal absorption by abundant molecules, and consequent absence of autofluorescence in the near-infrared range [31]. The latter, plenoptic imaging elaborately computes a 3D point for each pixel in an image through a microlens array and image sensors. While these technologies are under development, they bolster the potential of urologic surgery.

AI also has the potential to enhance skills assessment during real-time cases, providing immediate feedback with or without the aid of live human proctors. The technology has already been used to analyze movement trajectory, for eye tracking, and for gaze mapping data. It may prove to have a role in surgeon credentialing, already being able to categorize surgeons on a gradient from novice to expert [32, 33].

Deep learning has been found to be superior to non-deep learning forms of artificial intelligence, including various image recognition technologies [34]. The technology, feedforward probabilistic neural networks, has been used to analyze prognosis of bladder cancer recurrence by histopathology using morphological and textural nuclear features. It shows precision, with accuracy of 72.3% for recurrent tumors, and 71.1% for no recurrence [35]. Neural networks have also been used by many urologists with favorable results in the interpretation of prostatic and renal images [36]. One such study by Tewari et al. utilized an ANN to predict recurrence of prostate cancer following surgery and radiotherapy. The model used conventional parameters to assess 1,400 patients after radical prostatectomy, accurately predicting PSA progression in 76% of cases. As early as 2003, ANN grading and classification of urothelial carcinomas showed promise. Utilizing the Adaptive Stochastic

On-Line method with respect to morphological and textural nuclei, Grade I, II, and III tumors showed a classification accuracy of 90%, 94.9%, and 97.3%, respectively [37].

Research on AI-guided prostate cancer detection published in 2018 is promising as well, suggesting a 92% sensitivity and 82% specificity when using likelihood maps trained by support vector machine learning (a form of supervised learning technology) using a combination of T2 weighted, apparent diffusion coefficients, and diffusion weighted imaging from 14 patients [38]. This compares to 89% sensitivity and 73% specificity for radiologist-detected lesions [30]. This technology may prove of high impact, as the challenge for urologists remains to distinguish benign from malignant lesions without invasive biopsy. In 1995, Moul and colleagues used a KSOM and back propagation programs to stage non-seminomatous testicular germ cell tumor staging, with a sensitivity of 88% and specificity of 96% [39]. There is a paucity of data published in the last two decades prior to the present on further development of AI as it specifically relates to urologic imaging. A significant proportion of AI urological imaging research involves prostate cancer detection or staging. Abbod et al. reported that 60% of published articles found were related to prostate cancer detection, staging, and prognosis. This compares to 32% for bladder tumors.

Ogiela and Tadeusiewicz have used structural pattern recognition to represent upper urinary tract lesions with the aid of nominal features. In the course of analysis, they utilized expansive tree grammar to segment and filter the images, skeletonize, and then transform them into 2D images with diagrams showing the various contours of the straightened organ. In this way, they applied grammar rules to the urograms of the renal pelvis and calyces together with skeletons obtained using a skeletonizing algorithm. This tree grammar algorithm, $G_{edt} = (\Sigma; \Gamma; r; P; Z)$ has definitions for each component. The entire algorithm, G_{edt}, works in stages. The first stage works to define the renal pelvis, followed by larger renal calyces, then smaller calyces, and finally the renal papillae. The short branches of the renal papillae are found to be concave with respect to the smaller calyces after skeletonization. If the papillae are convex/shortened, an abnormality may be present and detected [40].

2.5 Benefits vs. Disadvantages

The benefits of artificial intelligence are numerous. The first of these includes increased productivity, due in part to absence of need for natural breaks in a 24-hour workday, enabling images to be read continuously. This allows results to be returned to patients quickly, and aids medical decision making.

Secondly, the lack of humanistic implications may be a strength as well. The various biases, lack of knowledge, or clerical errors made in the process of observing an image are minimized with computerization. Third, the cost of instituting a new graphics processor or imaging software is fixed. Over time, not only does this benefit allow for saving of human resources, it facilitates a margin of profit for any healthcare administration that only grows with time. Finally, artificial intelligence will continue to progress in its innovative capacities. Because many hospitals apply marketing strategies, AI may enable hospitals to market to their stakeholders, including potential employees and patients [41].

The disadvantages of artificial intelligence must also be considered. First, AI imaging algorithms and software may be able to compete for existing human labor more cost-effectively—it is estimated that AI will eliminate 1.8 million jobs by 2020. This disadvantage is tempered by the increase in employment—in 2017, it was estimated that AI will generate 2.3 million jobs by 2020, and will have a net gain of 2 million jobs by 2025 [42]. Secondly, the personal connection of reviewing an image will continue to be a necessary component of radiology; however, with further development and validation of imaging and interpretation software, there is a reasonable fear concerning the loss of the human review. This is the case of the development of STAR surgery. In the ensuing decades of its development, enhancements of imaging technology may facilitate the automation of surgery. In time, the personal trust a patient places in a surgeon's skills may grow increasingly tenuous.

2.6 Future Considerations

Looking to the future, AI's continued development will likely follow Satya Nadella's three phases for the many technological breakthroughs that have preceded it. The first, invention and design of the technology itself; the second, retrofitting (e.g. engineers receive new training, traditional radiologic equipment is redesigned and rebuilt); the third, navigation of the dissonance, distortion, and dislocation, where challenging novel questions are raised [36]. These may include: what the function of the physician will be when radiomics enables the detection of trends in particular illnesses, or whether CADe or CADx may guide management without the aid of the radiologist.

Many agree that AI should augment, rather than replace human ability. Amongst other considerations previously discussed, this must also be infused with the applicable protections for transparency, as well as privacy and security. This can be accomplished, but it will require a concerted effort amongst all stakeholders, including the requisite regulatory bodies governing the integration and incorporation of these powerful tools into practice.

References

1. S. De Brule, "What is artificial intelligence?" *MachineLearnings*, 2017. Available at: https://machinelearnings.co/how-to-prepare-your-career-for-artificial-intelligence-driven-automation-1bb153759b3b.
2. Merriam-Webster Definition of Artificial Intelligence, Merriam Webster, 2018. Available at: https://www.merriam-webster.com/dictionary/artificial%20intelligence. Accessed May 26, 2018.
3. M. L. Giger, "Machine learning in medical imaging," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, 2018.
4. FDA, US, "Guidance for Industry and Food and Drug Administration Staff: Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data—Premarket Notification [510 (k)] Submissions," 2012.
5. S. Nadella, *Hit Refresh: The Quest to Rediscover Microsoft's Soul and Imagine a Better Future for Everyone*, HarperCollins Publishers, New York, 2017.
6. S. Kim, "AI Coming to a Portfolio Near You." Barron's, 2018.
7. A. Tang *et al*. "Canadian Association of Radiologists white paper on artificial intelligence in radiology," *Canadian Association of Radiologists Journal*, 2018.
8. National Academies of Sciences, Engineering, and Medicine, *Improving Diagnosis in Health Care*. National Academies Press, 2016.
9. Pwc, "Sizing the prize: what's the real value of AI for your business and how can you capitalise?" *PwC*. Available at: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. Accessed May 26, 2018.
10. G. Brockman, "The dawn of artificial intelligence," Testimony before U.S. Senate Subcommittee on Space, Science, and Competitiveness, November 30, 2016.
11. Richard Viereckl *et al*. "Connected car report 2016: opportunities, risk, and turmoil on the road to autonomous vehicles," *PwC*, 2016. Available at: http://www.strategyand.pwc. com/reports/connected-car-2016-study.
12. Z. Obermeyer and E.J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *The New England Journal of Medicine*, vol. 375, no. 13, pp. 1216, 2016.
13. Y. Yamada and M. Kobayashi, "Detecting mental fatigue from eye-tracking data gathered while watching video: evaluation in younger and older adults," *Artificial Intelligence in Medicine*, 2018.
14. K. Rossmann and B. E. Wiley, "The central problem in the study of radiographic image quality," *Radiology*, vol. 96, no. 1, pp. 113–118, 1970.
15. M. Molteni, "If you look at X-Rays or Moles for a Living, AI is coming for your job," Wired, 2017. Available at: https://www.wired.com/2017/01/look-x-rays-moles-living-aicoming-job/
16. C. J. Vyborny, "Image quality and the clinical radiographic examination," *Radiographics*, vol. 17, no. 2, pp. 479–498, 1997.
17. World Health Organization, *Baseline Country Survey on Medical Devices 2010*, WHO Press, Geneva, Switzerland, 2011.
18. K. Doi, K. Rossmann, and A. G. Haus, "Image quality and patient exposure in diagnostic radiology," *Photographic Science and Engineering*, vol. 21, no. 5, pp. 269–277, 1977.

19. K. Rossmann, A. G. Haus, and G. Dobben, "Improvement in the image quality of cerebral angiograms," *Radiology*, vol. 96, no. 2, 361–366, 1970.
20. R. Pearl, Artificial intelligence in healthcare: Separating reality from hype. *Forbes*. 2018.
21. K. C. Santosh and Laurent Wendling, "Angular relational signature-based chest radiograph image view classification," *Medical & Biological Engineering & Computing*, pp. 1–12, 2018.
22. K. C. Santosh and Sameer Antani, "Automated chest x-ray screening: can lung region symmetry help detect pulmonary abnormalities?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1168–1177, 2018.
23. S. Vajda *et al*. "Feature selection for automatic tuberculosis screening in frontal chest radiographs," *Journal of Medical Systems*, vol. 42, no. 8, p. 146, 2018.
24. F. T. Zohora, Sameer Antani, and K. C. Santosh, "Circle-like foreign element detection in chest x-rays using normalized cross-correlation and unsupervised clustering," *Medical Imaging 2018: Image Processing*, vol. 10574. International Society for Optics and Photonics, 2018.
25. F. Tuz Zohora and K. C. Santosh, "Foreign circular element detection in chest x-rays for effective automated pulmonary abnormality screening," *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 7, no. 2, pp. 36–49, 2017.
26. K. C. Santosh *et al*. "Edge map analysis in chest X-rays for automatic pulmonary abnormality screening," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 9, pp. 1637–1646, 2016.
27. K. C. Santosh *et al.* "Automatically detecting rotation in chest radiographs using principal rib-orientation measure for quality control," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, no. 02, p. 1557001, 2015.
28. T. Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, pp. 52–65, 2013.
29. D. Rubin, "Frontiers of AI in medical imaging for clinical decision making." 2017 Human AI Collaboration: A Dynamic Frontier Conference. Accessed Apr 07, 2018.
30. M. Honda *et al.* "Current status of robotic surgery in urology," *Asian Journal of Endoscopic Surgery*, vol. 10, no. 4, pp. 372–381, 2017.
31. R. G. Pleijhuis *et al.* "Near-infrared fluorescence (NIRF) imaging in breast-conserving surgery: assessing intraoperative techniques in tissue-simulating breast phantoms," *European Journal of Surgical Oncology (EJSO)*, vol. 37, no. 1, pp. 32–39, 2011.
32. M. J. Fard *et al.* "Automated robot-assisted surgical skill evaluation: predictive analytics approach," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 1, p. e1850, 2018.
33. W. R. Boysen, M. G. Gundeti, "Use of robot-assisted surgery in pediatric urology—have we reached the limits of technology?" *AUANews*, vol. 23, no. 6, pp. 19–20, 2018.
34. Xinggang Wang *et al.* "Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning," *Scientific Reports*, vol. 7, no. 1, p. 15415, 2017.
35. P. Spyridonos *et al.* "A prognostic-classification system based on a probabilistic NN for predicting urine bladder cancer recurrence," in *Digital Signal Processing*, 2002. *2002 14th International Conference on Vol. 2*. IEEE, 2002.

36. M. F. Abbod *et al.* "Application of artificial intelligence to the management of urological cancer," *The Journal of Urology*, vol. 178, no. 4, pp. 1150–1156, 2007.
37. D. K. Tasoulis *et al.* "Urinary bladder tumor grade diagnosis using on-line trained neural networks," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 2003, Springer, Berlin, Heidelberg, Germany.
38. Y. Oishi *et al.* "Automated diagnosis of prostate cancer location by artificial intelligence in multiparametric MRI," *European Urology Supplements*, vol. 17, no. 2, e888–e889, 2018.
39. J. W. Moul *et al.* "Neural network analysis of quantitative histological factors to predict pathological stage in clinical stage I nonseminomatous testicular cancer," *The Journal of Urology*, vol. 153, no. 5, pp. 1674–1677, 1995.
40. M. R. Ogiela and Ryszard Tadeusiewicz, "Artificial intelligence structural imaging techniques in visual pattern analysis and medical data understanding," *Pattern Recognition*, vol. 36, no. 10, pp. 2441–2452, 2003.
41. G. A. Okwandu, "Marketing strategies of hospital service organizations in Nigeria: a study of selected privately owned hospitals in Port Harcourt," *Journal of Hospital Marketing & Public Relations*, vol. 14, no. 1, pp. 45–57, 2002.
42. Yen Nee Lee, "Robots 'are here to give us a promotion,' not take away jobs, Gartner says." *CNBC*, 2017. Available at: https://www.cnbc.com/2017/12/18/artificial-intelligence-will-create-more-jobs-than-it-ends-gartner.html?__source=facebook%7Ctech.

A Systematic Review of 3D Imaging in Biomedical Applications

8.1 Introduction

Scientific visualization is a method of scientific computing. It converts acquired symbolic data into a geometric form to convey silent information of underlying data and to see the unseen structure which is beneficial for comprehension, analysis, and interpretation [1, 2]. The field of scientific visualization has been widely explored in the last three decades. The National Science Foundation Visualization created the field of scientific visualization in 1987 by presenting a paper in a scientific computing workshop [3]. The phenomenal growth in scientific visualization is possible due to advances in data acquisition and computational technologies. The acquired data can be two-, three-, or even more dimensional, and is used to convey detailed information about complex phenomena/processes such as the flow of gases and fluids, biological processes, space, and earth sciences. The acquired data are voluminous and exclusively in numerical format; it is impossible for the human brain to analyze and interpret thosedata. Most of the valuable information may be lost in the manual process. Thus scientists are motivated to adopt technological advancements [4].

Nowadays, scientific visualization is effectively used to simulate physical processes in order to achieve a better and precise understanding of our universe. This is used to study natural phenomena which are extremely large or small, unduly quick or slow, or might be too harmful or dangerous to observe directly. In addition to this, scientific visualization helps researchers to extract meaningful information as well as to discover new information from complex and voluminous datasets by using interactive computer graphics and imaging techniques. Scientific visualization is the basis behind the growth of several fields, such as computer graphics, image processing, computer vision, signal processing, cognitive sciences, computational geometry, user interfaces, and computer-aided design. Moreover, visualization is the backbone of a wide variety of applications such as defense (computer-generated forces and advanced distributed simulation applications, for instance), engineering (computer-aided design and computer architecture for instance), computational fluid dynamics, and computer graphics applications [1].

In addition to this, visualization plays a vital role in most computerized medical applications such as computer-aided diagnosis (CAD), computer-assisted surgery (CAS), and simulator development. In CAD applications, visualization is essential for accurate disease diagnosis and prognosis. For instance, it is possible to visualize the actual anatomical structure of bones or several soft tissues, and their responses for particular situations. Moreover, experts can identify the effects of procedures before treating patients. CAS visualization plays a crucial role in the development of custom prostheses and anatomic models. Other than this, visualization-based CAS systems are successfully developed for neurosurgery, image-guided surgery, custom

anatomic atlas, robotic assistance, and surgical planning [1]. Virtual reality (VR)-based simulator development is an emerging field in medicine, where (3D) visualization plays a vital role. Most of the VR-based simulators are developed to increase surgical competence and reduce runtime complications. Especially in the orthopedic field, such types of VR-based trainer simulators are in demand. With the help of simulators novice trainees can do the practice of various orthopedic psychomotor skills at no cost once it has been developed [5]. In addition to this, simulators are helpful for 3D model visualization of complex bone anatomy.

This chapter aims to provide detailed information on several volume visualization principles, techniques, and algorithms that are widely used in the medical field. Explanations on volume visualization techniques used in the rest of the fields discussed above are beyond the scope of this chapter. Section 8.2 discusses some preliminaries of volume visualization. It includes the explanation of volumetric data and their types, several grid structures, and the data acquisition process. Sections 8.3 and 8.4 respectively cover the detailed information on the various indirect and direct volume rendering techniques used in the literature to render medical data. Section 8.5 discusses the challenges in volume rendering on conventional computational devices and explores the recent advances in volume visualization. The primary focus is given toexploring advances in hardware-based volume visualization and transfer functions (TFs). This section also covers the advantages of GPU utilization in both direct and indirect volume rendering techniques. Section 8.6 liststhe various commonly used tools and libraries for volume visualization. Section 8.7 provides the conclusion and future directions.

8.2 Volumetric Data

Before proceeding to a detailed explanation of several volume visualization approaches in the biomedical field, it is necessary to explore some basic concepts that are required to understand the logic of visualization algorithms. This section covers the required basic terminologies and various data acquisition alternatives available in the medical field to acquire volume data. Different ways of representing volume data, various grid (mesh) structures, and general steps in the volume visualization process are explained in detail in this section.

8.2.1 Data Acquisition

Volume datasets can be generated by using sampling techniques (stochastic method for instance), simulation, modeling techniques, curving the material, or hand painting in 3D. Other than these, voxelizing geometric description of

objects and writing programs are the most popular alternatives to generate volume datasets [6]. However, in the medical field, volume datasets are often acquired by scanning the material-of-interest using 3D scanning techniques such as magnetic resonance imaging (MRI), computer-aided tomography (CT), positronemission tomography (PET), and/or sonogram machines. In addition to this, laser scan confocal and other high-power microscopes are also used to acquire data [7]. Depending upon the concern of inspection, different 3D scanning techniques are used. For instance, to inspect soft tissues, MRI is preferable, whereas for inspecting bone diseases and trauma CT is widely used.

Due to the evolution of medical imaging modalities (such as CT, MRI, and PET for instance) after 1970 the terminologies in medical imaging have shifted from 2D projection to fully isotropic 3D images [8]. Commonly, these medical imaging modalities scan material-of-interest from several angles and generate anumber of 2D slices. The count of slice per stack is highly dependent upon the amount of portion one needs to scan, the thickness of each slice, and the distance between two slices. The radiographer sets these parameters during the time of inspection. Then the sequence of 2D slices obtained from the scanner is used to reconstruct the volumetric model in 3D space. The 3D reconstructed models can be used for better visualization of internal structure, for diagnosis or to plan the recovery process.

8.2.2 Volume Data

Volume data is a set of samples (x, y, z, v). The values of v represent the quantitative property of scanning material at a 3D location (x, y, z). The values of v can be scalar or vector. More generally, scalar values are single-valued, and represent some measurable property of the data such as color, intensity, heat, or pressure. However, the vector value is multi-valued, and represents the velocity or direction in addition to the scalar value at that location [6]. Vector values are useful in the fields where direction plays a vital role during volume visualization and interpretation process (computational fluid dynamics for instance). A further discussion of vector volume data is out of the scope of our study, because this paper aims to provide extensive information about volume visualization techniques thatare useful in the biomedical field. In this field, the volume data generated with the help of 3D scanning devices by scanning material of interest is scalar data. The value of v represents the gray intensity values at that location.

In the biomedical field, volume is a 3D array of volume-elements (voxels). This 3D array can be visualized as a stack of 2D slices as shown in Figure 8.1. Slice-oriented representation is another way to visualize 3D data. This is the traditional representation, nothing but how the physicians look at volume dataset. It is denoted by a matrix $V = \Gamma^{X \times Y \times Z}$ where X, Y, and Z represent rows, columns, and slices respectively. The V is a discrete grid of voxels v. Each voxel v is represented by $I(v)$: $N^3 \rightarrow \Gamma$ [9]. Concerning CT imaging, it is

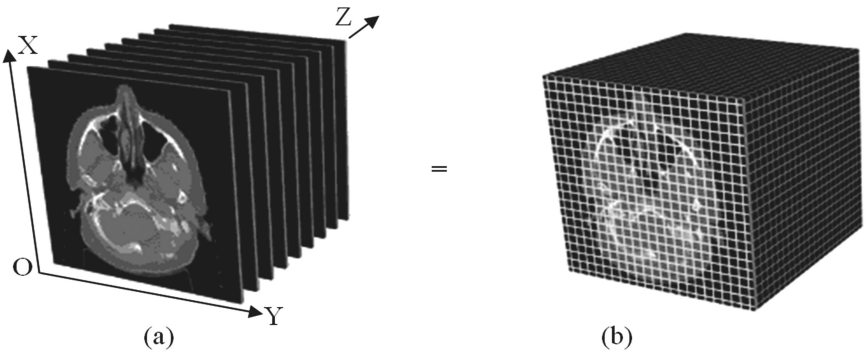


FIGURE 8.1
3D volume data representation (a) n × 2D slices and (b) 3D stack.

gray values which are X-ray attenuation coefficient of the material of interest at that point. The volume data obtained from CTscanners are anisotropic with an equal sampling density in the x- and y-directions (more generally, 512 × 512 voxels in each direction), whereas it has a coarser density along the z-direction (i.e. the number of slices may vary per stack). The number of slices per stack ranges from 100 to 600, and typically depends upon area under supervision and severity of disease/trauma. The datasets V generated by the CT-scanning process arethe basis for the development and analysis of volume visualization (rendering) algorithms.

8.2.3 Grid Structures

Grid (mesh) structures determine the volume visualization technique. They structures depend on the source of volumetric data. A grid structure can be structured or unstructured [10]. A structured grid shows regular connectivity between the grid points, whereas an unstructured grid shows intermittent connectivity. Figure 8.2 shows the different types of grid structures. Uniform, rectilinear, and curvilinear are examples of structured grids as shown in Figure 8.2 (a), (b), and (c) respectively. Unstructured grids (Figure 8.2 (d)) are usually generated by physical simulation, whereas scanning devices produce volumetric data in structured grids. The grids generated by CT or MRI scanners are a rectilinear structured grid. Further discussion on rest of the grid structure is beyond the scope of this paper.

8.2.4 Volume Visualization

Volume visualization techniques are used to create 2D graphical representations from volume datasets defined over 3D grids. Based on the types of volume datasets, different techniques can be adopted for volume visualization (rendering) [11]. Techniques such as isosurfaces, slice planes, and contour

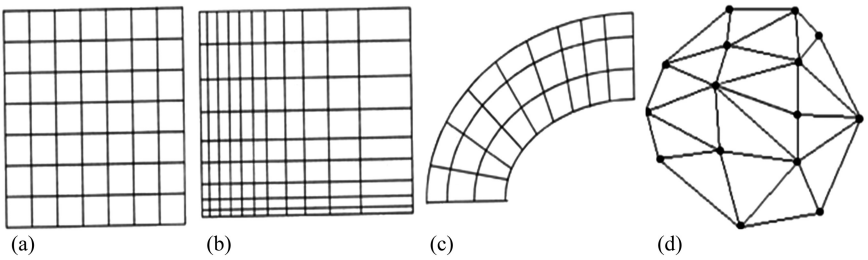


FIGURE 8.2
Types of grid (a) uniform, (b) rectilinear, (c) curvilinear, and (d) unstructured grid.

slices are best suited for the visualization of scalar data, whereas streamlines, cone plots, and arrow plots are suitable visualization techniques for vector data. More explanation on visualization techniques of vector data is out of the scope of this paper, as medical 3D scanning devices generate scalar data. Further, scalar data volume visualization techniques can be categorized into parts: direct and indirect volume visualization. Sections 8.3 and 8.4 provide a detailed explanation of these two techniques.

8.2.5 Steps in Volume Visualization

Most of the volume visualization methods use common steps to render volumetric data. Data acquisition, preprocessing, and define view are some common steps included in many state of the art volume visualization methods [7].

8.2.5.1 Data Acquisition and Dimension Reconstruction

The initial step of every visualization method is data acquisition. The material of interest is scanned via scanning devices to generate the data. In the biomedical field, CT or MRI scanners are commonly used to generate the volume data. The acquiesced data need to be reconstructed to match the dimensions of the scanned material of interest to the dimensions of the display coordinate system.

In addition to this, reconstruction is required to make equidistant (regularly spaced) scanned data. This is required when aradiographer intentionally constructs the irregular spaced data to capture a complicated part in detail. The radiographer may increase the distance between the adjacent slices where the scanned portion is less interesting and may decrease the distance at the area of interest. For example, during CT stack construction, the radiographer may increase the distance while scanning healthy bone and decrease the distance during the scanning of the fracture-prone area to obtainmore details [23]. In such a situation, new slices need to be reconstructed or to replicate existing slices to make the volume equidistant. Most of the methods adopted the interpolation method to predict the values in new slices.

8.2.5.2 Data Preprocessing and Extraction

The acquired slices may contain unwanted artifacts (CT bed, cables, and flesh, for instance), or noise may get added during the image acquisition process. At this point, suitable preprocessing and segmentation techniques need to be applied. The preprocessing technique is responsible for removing unwanted artifacts and forenhancing the desired portion with precision [73]. After that, the application of a segmentation technique necessary to extract the desired portion out of the image. Some authors named this as "data classification step."

8.2.5.3 View Definition

The main aim of the view definition step is the selection of an appropriate TF. The TF is responsible for mapping volume data onto geometric or display primitives. In addition to this, it specifies the coloring and lighting effects. Lighting effects are used to enhance the visibility of the surface shape and to provide the 3D perspective of the volume data. Next, the view definition includes adjusting camera position, specifying aspect ratio, and selecting-projection type. This step may vary in every algorithm. During rendering, these primitives can be stored, manipulated, and intermixed with each other to display the view on the screen.

8.3 Indirect Volume Rendering (Surface Fitting)

Indirect volume (i.e. surface) rendering techniques render the only surface of the given volume data. The rendered surface is opaque, and it is easily manageable. The coherent structures like skin and bone are represented by point sets with the same sampling rate. Polygons are used to approximate the surface from the point set. Then the triangular mesh is created to better represent the given volume, and at last the created 2D triangular mesh is rendered on screen with proper shading and lighting effects [6].

The surface rendering approach commonly extracts geometric primitives (the boundary of an object, for instance) to approximate the surface of objects being rendered [9]. To extract such geometric primitives, a suitable segmentation technique is adopted. Most surface rendering methods use a thresholding-based segmentation technique to extract the boundaries of the desired objects. This is true even for commercial medical imaging tools such as DICOM* viewer and 3D slicer, for instance. They also use a user-defined

* DICOM: digital imaging and communications in medicine.

threshold to extract boundaries of objects before generation of polygon mesh for surface rendering.

In short, the surface rendering approach segregates each voxel in one of two classes: one part of the desired object and another part of the background. To do this, it needs the help of a user-defined threshold value. Then, object boundaries are extracted using an edge detector operator (Canny or Sobel operator, for instance). At last, a polygonal mesh is generated and the surface is displayed on screen by applying proper shading effects. Shading effects are determined by the isosurfaces value at that location [9]. Opaque cubes (cuberilles), contour tracing, marching cubes/tetrahedral, and dividing cubes are afew examples of surface rending algorithms. Amongst them, marching cube is a widely used method.

8.3.1 Opaque Cubes (Cuberilles)

This method initially performs the binarization of a given volume by considering the isovalue. Then all boundary front faces are identified such that the normal points of the faces are pointing towards the viewpoint. Lastly, these faces are rendered as shaded polygons. This method does not use any interpolation method to determine points, thus the object boundary is not identified precisely; moreover, it results in a block surface.

8.3.2 Contour Tracing

This is one of the most flexible methods in the biomedical field for volume visualization before the invention of the marching cube algorithm. Like cuberille, this method also affects binarization of the given volume. Then polyline is obtained by traversing the boundary pixels in a clockwise direction. Finally, the polylines of adjacent slices thatrepresent the same objects are connected to form triangles, and the generated triangles are displayed on the screen as an isosurface of the object. This method becomes stuck if there is considerable variation between two adjacent slices and there are some objects in the slice.

8.3.3 Marching Cube

Lorensen et al. [31] devised the marching cubes algorithm to determine the isovalued surface with a triangle mesh. The algorithm defines a voxel in terms of a cube having pixels values at eight corners of the cube. These cubes are marched through the entire volume. During marching, each vertex of the cube is classified as being inside or outside the isosurface. The edges where one vertex is classified as outside and another one is classified as inside are used to form a triangular patch. Then these triangular patches from adjacent lines are connected to each other to form the isosurface. To form a triangular patch between two adjacent slices, the cube (voxel) can be represented by 256 different cases. These are reduced to only 15 cases observing symmetrical

structures. These 15 generic triangles are stored into a lookup table for future reference. The actual vertices to form a triangle are determined by the linear interpolation function.

Lastly, the normal value is determined for each vertex, and the triangular mesh is projected on standard graphics hardware for rendering. Due to simplicity and efficiency, the marching cube algorithm is adopted by several researchers to reconstruct 3D models from acquired medical data [12]. The divide cubes are used to render the isosurfaces in large datasets, whereas the variation of marching cubes, the marching tetrahedra [32], is used to render the surface in unstructured grids.

The surface rendering technique best approximates the surface of the volume data, andis efficient and straightforward concerning space and time. However, in addition to surface most of the applications require internal details also. For instance, in VR-based orthopedic simulators developed to compute drilling parameters such as depth, an entire volume is required to perform erosion. That is to remove the part of volume drilled by the drill bit to show the depth. In such applications' surfaces, rendering techniques are not efficient. Hence, many such applications havenow adopted the direct volume rendering technique.

8.4 Direct Volume Rendering

Surface rendering techniques extract geometric primitives to display the surface of the volume data. If the data size is too large, then the time required to extract geometric features and render a surface may be very long. In such cases, surface rendering techniques are not efficient. In addition to this, the volume is represented by the surface, thus most of the internal information is lost during the rendering process. That information is valuable and required for a precise understanding of the data. To avoid information loss andto gain high accuracy, direct volume rendering (DVR) techniques are being used frequently in biomedical applications [6].

DVR techniques directly map voxels to pixels in a 2D plane. That is, they directly render the segmented volume with the help of a suitable TF without extracting any kind of geometric features or structures. The DVR approach reconstructs the 2D image directly from 3D volume data. Raycasting, splatting, shear-wrap, maximum intensity projection, and 3D texture mapping are the most commonly used DVR techniques. These techniques operate on actual volume data.

8.4.1 Raycasting

Raycasting is one of the most widely used DVR techniques to display volume data in the 2D viewing plane. It gained a lot of importance in the literature

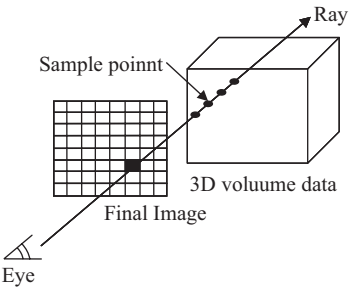


FIGURE 8.3
Raycasting pipeline.

as it has the most substantial body of publications. This technique fires a ray from each pixel in the viewing plane into the volume to determine the color and opacity values from each scalar value. The tri-linear interpolation function is used to compute these values [8, 9]. This technique does not generate any shadows or reflection effects [7]. The raycasting pipeline is shown in Figure 8.3. A few researchers worked to improve the performance of the raycasting method. Levoy [13] developed an adaptive refinement model. The method skips the empty spaces during raycasting and sample points computation to improve the performance.

Further, Levoy [14] discussed two improvements. In the first technique, spatial coherence in the volume data is encoded through a pyramid of binary volume; the second method uses the opacity threshold to terminate the ray tracing. A thresholding level parallelism and a bricked volume layout are used by Grimm et al. [15] to improve the performance of the raycasting method.

8.4.2 Splatting

The splatting DVR technique represents each voxel by 3D reconstruction (Gaussian) kernel. Each projected kernel leaves a splat (footprint) on the viewing plane [6, 7]. The voxel contribution to each pixel in the image plane is calculated using a look table. In addition to this, the color and opacity values in the pixels are computed by TF [8]. Figure 8.4 shows the splatting pipeline. This technique is named splatting because the rendering output is similar to a scenario generated on a glass plate when throwing a snowball on it. The concentration of snow is more at the center, whereas it fades while moving away from the center. The basic splatting algorithm suffers from a color blending problem [16]. An aligned sheet buffer technique is adopted by Westor to solve this problem [17]. A self-sufficient data structure, "FreeVoxels," was developed [18] to remove the same problem. Zwicker et al. [19] used an elliptical weighted average filter to overcome the color blending problem. Xeu et al. [20] adopted texture mapping to improve the performance of the splatting process.

8.4.3 Shear-Warp

A hybrid DVR algorithm shear-warp is one of the fastest volume rendering algorithms. It attempts to combine the advantages of both the image order-based and object order-based volume rendering methods. It uses run-length encoding (RLE) compression to compress volume the data, which allows fast streaming through it. In addition to this, it factorizes the viewing transformation into 3D shear and warp transformation in the 2D plane [8, 9]. Figure 8.5 shows the transformation of shared object space from the parallel projection tothe original object space. The shear-warp algorithm achieves greater rendering speed at the cost of image quality. To overcome this drawback, i.e. to maintain image quality while rendering the pre-integrated volume, rendering in the shear-warp algorithm for parallel projection is implemented in Schulze et al. [21]. Further, Kye et al. [22] presented two methods to improve image quality. In the first method, super-sampling is performed in an intermediate image space, whereas the second method uses a pre-integrated rendering technique with the help of a new data structure named "overlapped min-max block."

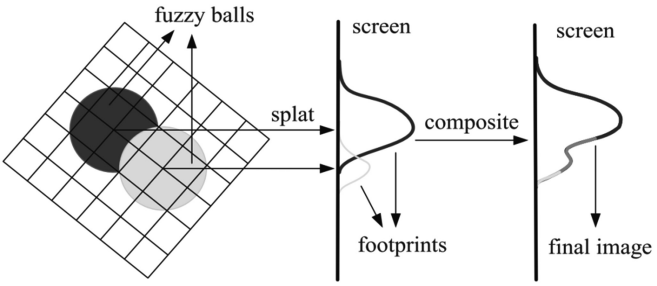


FIGURE 8.4
Splatting pipeline. (Image courtesy of Zhang et al. [8].)

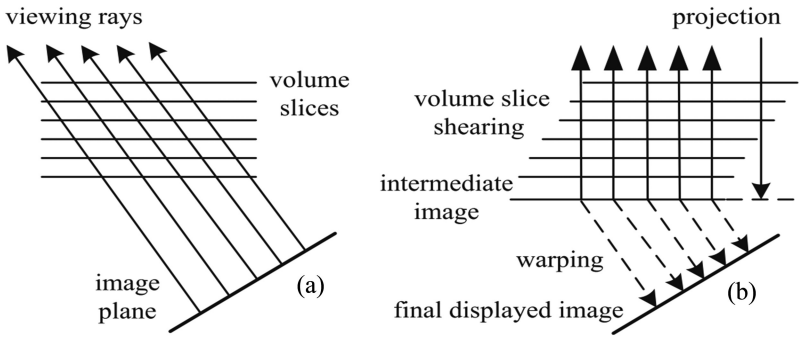


FIGURE 8.5
(a) standard transformation and (b) shear-warp factorized transformation for a parallel projection. (Image courtesy of Zhang et al. [8].)

8.4.4 Maximum Intensity Projection

The basic idea of maximum intensity projection (MIP) is to assign a maximum intensity to the pixels. The maximum intensity is determined by evaluating each voxel of volume thatlies on the path of ray coming from the viewer's eye [22]. The idea of MIP is illustrated in Figure 8.6. It is mainly used to visualize high-intensity structures such as blood vessels from the volume data. The limitations of MIP techniques are as follows: no shading information is given, and the depth and occlusion information is lost [9]. To overcome one of the limitations (depth information loss) of MIP, tri-linear interpolation—a function-based, interactive, high-quality MIP method—is implemented in Mroz et al. [24].

8.4.5 3D Texture Mapping Volume

The texture mapping volume rendering technique is widely supported by traditional display devices to render synthetic images. Thanks to Cullip et al. [25] and Cabral et al. [26], the texture mapping technique has become popular. The core idea of the texture mapping method is to decompose the volume into three stacks and interpret each voxel as 3D texture defined over $[0, 1]^3$. During rasterization, the texture information at an arbitrary point is extracted by using tri-linear interpolation within the volume data. Figure 8.7 shows the pipeline of the texture mapping volume rendering technique. Van Gelder et al. [27] introduced shading in texture mapping to improve the image quality. Further diffuse and specular shading models are integrated into Rezk-Salama et al. [28] to improve image quality. Abellán et al. [29] proposed three types of shading for the multiple-model dataset to accelerate the performance of the texture mapping technique.

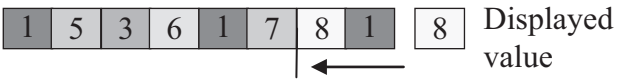


FIGURE 8.6
Concept of maximum intensity projection.

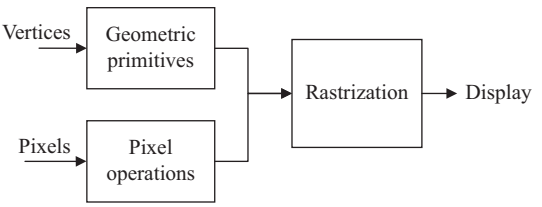


FIGURE 8.7
3D texture mapping pipeline.

8.5 Recent Advances in Volume Visualization

Visualization data arestored in multi-dimensional volumes. These volumes may grow large in spatial terms as well as in a number of dimensions. The data architecture used to store volumetric data supports various file formats (such as DICOM and STL*, for instance) and software (several DICOM viewers or 3D slicers, for instance). Devising and developing an efficient volume rendering algorithm is still a challenging task. Visualization hardware, TFs, and rendering styles are a few of the primary challenges to develop such algorithms. In addition to this, researchers need to focus on some additional challenges, such as view-dependent algorithms, image-based rendering, multi-resolution techniques, importance-based methods, adaptive resource-aware algorithms, remote and collaborative visualization algorithms, and networking [11].

8.5.1 Advances in Hardware (GPU)-Based Volume Rendering

In traditional volume rendering algorithms, the software calculates the frame, whereas the CPU is used only to render the frames. There are certain disadvantages associated with software-based volume rendering. It assumes the hardware device (CPU and memory, for instance) is fast enough. In addition to this, algorithms forcefully ask the hardware to handle data which is high in volume. However, CPU and memory by themselves cannot handle such voluminous data (dedicated hardware is required to handle such data). The reason behind this is the gap present between the speed of memory and CPU. Software-based rendering algorithms ask for precalculated data, because precalculated data are always faster than the data calculated on the fly (i.e. during the rendering process). Though the processing of precalculated data makes things faster, the process still does not meet to the desired efficiency, because memory is not fast enough to feed the data that is required by CPU. For example, the clock speed of memory (DDR3 RAM for instance) is ~1333 MHz, whereas the clock speed of the processor is ~2.6 GHz. There is a massive gap between processing speeds. Additionally, cache memories are too small.

Due to the limitations mentioned above relating to software-based volume rendering algorithms, several recent studies introduced innovative hardware-based approaches that are better suited for parallel processing. In addition to this, various advances are being made in parallel processing hardware, i.e. graphical processing unit (GPU). Hence nowadays the GPUs are at the core of volume visualization.

* STL: surface template library.

8.5.1.1 The Need for GPU

Generating volumetric data and rendering them on large size displays is always computationally expensive. To render the data in inner slices, the algorithms have to create pixels for each of the slices based on direct or indirect volume rendering techniques. Processing such large data is always challenging for the developers, and it is also a hardware-bound task. In order to explore the insights of the data, the user has to view the data from different point of views. This, in turn, creates a demand for view-dependent processing. All such processing is computationally expensive in terms of speed and memory. This leads to the adaptation of GPU for volume visualization [67].

The data processing in visualization is required at several stages such as data acquisition, data preprocessing, volume generation, volume rendering, and visualization. This has to be done in real time, or sometimes on the fly. To process such high-dimension data, a parallel processing system becomes an essential requirement. In addition to this, the pixels/voxels need to be created using volume generation algorithms for direct volume rendering. The direct volume rendering method requires the abstract and underlying data together. That is, it requires the volume to be transparent and should display the propagation of light and shadows through the volume. Such rendering makes it necessary for itto be processed by parallel systems. The required parallel processing can be achieved by GPUs [71]. In addition to this, some applications require for the volume data to be explored with insights. To explore such insights with minute details, the display screen has to be large, and the resolution of the screen must be high enough.

Other than this, the advancements in GPU are far better than CPU as far as processing speed and memory capacity are concerned. These advancements help GPUs to render frames at a very high rate. This is possible because the large size data can also be fitted easily into the GPU [71]. This property of GPU is beneficial for hardware-based volume rendering. The visualization in the medical field is increasingly dependent on high-performance computing (HPC) to calculate underlying pixel data, volumes properties, and rendering phenomena [73]. For this kind of tasks, multi-GPU workstations, i.e. distributed GPU clusters and cloud-based GPU, are a more feasible option. GPU units are multicore, and each GPU unit has multi-threaded blocks. These are the core of parallel processing. The introduction of high-bandwidth networking is the backbone for distributed image processing and is being actively applied in the field of medical visualization. Simulation-based hardware and software areincreasingly applied in visualization to generate volumes. In addition to GPU, some recent applications are making use of tensor processing units (TPU) for medical visualization tasks.

8.5.1.2 Accelerators on the GPU

To accelerate parallel processing with GPU, there is a need for software accelerators to leverage the parallel processing power of GPUs. One such

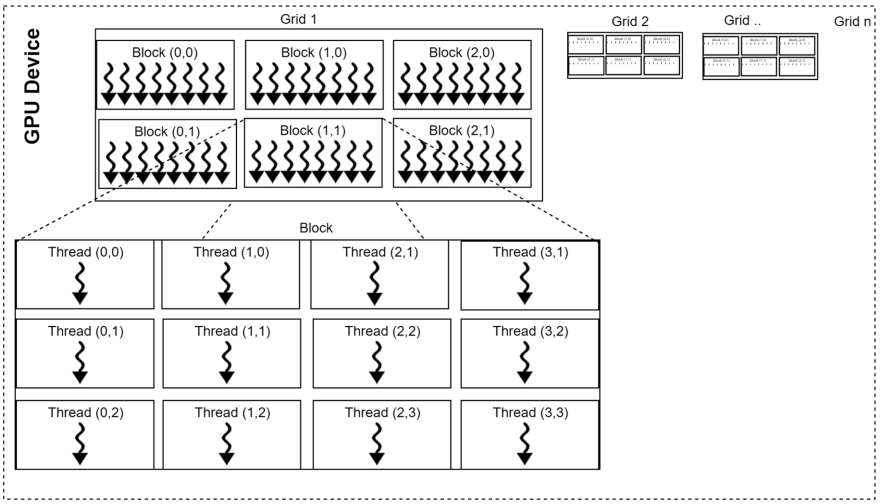


FIGURE 8.8
CUDA architecture. (Image courtesy of NVIDIA [71].)

accelerator is the compute unified data architecture (CUDA). Other than CUDA, OpenCL* is another accelerator used to improve the visualization. The GPU manufacturer AMD† supports OpenCL acceleration, and this is being used in many visualization devices for processing multi-dimensional visualization data. In addition to this, these devices are relatively inexpensive. Thus nowadays many medical visualization applications use GPU-accelerated devices [73]. However, visualization quality and execution time are dependent on factors such as GPU models, number of GPUs, number of threads, and acceleration libraries used [71].

NVIDIA introduced the CUDA model in 2006. This is used to process instructions in parallel. Most of the programming features are an extended form of C programming language, which is compiled using an NVCC compiler. The CUDA programming model uses the acceleration in GPU by executing instructions in parallel threads. Figure 8.8 shows the CUDA architecture. CUDA has several threads in a single block, whereas each grid contains several such blocks [71, 72]. By assigning image data and by using the thread IDs, the data is processed in a parallel manner.

Recent studies show that researchers prefer hardware accelerators along with software libraries for their visualization experiments. Weinrich et al. [33, 65] used CUDA and OpenGL‡, along with two GPUs, namely, GeForce 8800 GTX and Quadro fx 5600, for their visualization experiments. The

* OpenCL: open computing language.
† AMD: advanced micro devices.
‡ OpenGL: open graphics language.

speed was increased up to 148 times higher than CPU. These were earlier versions of CUDA and OpenGL. At the time of writing this paper, NVIDIA has released a new CUDA version: CUDA 10.0. Recent improvements in CUDA include support for various libraries such as deep learning for medical images [34]. Libraries such as cuDNN, GIE, cuBLAS, cuSPARSE, and NCCL support deep learning in CUDA. The 4D processing of medical images is supported by some of the CUDA libraries (CUFFT and NPP). The present study demonstrated latencies in the visualization pipeline using GPU [35]. In this study, the visualization algorithm is designed using-3Dtexture mapping (3DTM), software-based raycasting (SOFTRC), and hardware-accelerated raycasting (HWRC). They used three GPUs to compare the performance gain. Recently, NVIDIA introduced an RTX* feature in its flagship GPU RTX2080, which is beneficial for direct as well as indirect volume rendering [70].

8.5.2 Advances in TFs

Designing TFs is a complex task. Developers need to consider several parameters (color, opacity, and texture, for instance) while designing an efficient TF. A TF is used to create volumetric data and to extract optical properties (color and opacity, for instance) from volume [60]. The TF is a mandatory part of the visualization pipeline. Medical volume data is a scalar entity, and it is in the 3D spatial domain. 3D image generation involves mapping data from voxel to pixel through a TF. A TF can be categorized into two types: image-centric and data-centric. More generally, the data-centric TF can be designed in four different ways: manual, semiautomatic, automatic, and machine learning–based. In manual TFs, handcrafted parameters are used, so these are not considered for further discussion. In semiautomatic partial user, intervention is required (for view selection), whereas in automatic TF no user intervention is required and the machine learning–based TF learns the data and extracts all the parameters on its own. Researchers have designed several TFs, including image-centric and data-centric TFs which consider all the parameters and design challenges.

8.5.2.1 Image-Centric TFs

In image-centric TFs, the parameters are calculated from the resultant images. These are used to generate a final result from the initial results. Usually, image-centric TFs extract the external shape of objects as a parameter. However, some additional parameters (texture, for instance) are not extracted with precision, which means they miss the minute details. Such minute details are the primary requirement of almost all visualization applications in the medical field, because they provide better insights and are

* RTX: real-time raytracing

useful to identify exact scenarios such as cancerous cells, damages to blood veins and arteries, and minute fractures in bones, for instance. Thus, image-centric TFs are not preferable for medical visualizations [60].

8.5.2.2 Data-Centric TFs

Unlike in image-centric TFs, in data-centric TFs parameters are derived from original data instead of resultant images. To do that, the information from voxels consulting the volume is taken into consideration [60].

Recent research has revealed that volume rendering is a crucial stage in volume visualization. The result of rendering is highly based on the selection of a suitable TF. However, the practical design of the TF is a more complex and time-consuming task. Recently, the research effort has paved the way for semiautomatic and automatic TFs. These TFs are being used to overcome several design challenges in volume visualization. The primary aim of TFs is to identify the objects in underlying data and extract the desired parameters. In addition to this, these parameters are used based on material and projection.

We studied recent advances in TFs that are useful for medical visualization. The following sub-section investigates some of the recent designs of semiautomatic, automatic, and machine learning–based TFs. Earlier research shows that TFs used histogram values of voxels being rendered. Such TFs with scalar values (shown in Equation 8.1) are usually called 1D-TFs.

$$q(d) = C\big(M(d)\big) \tag{8.1}$$

However, by only using histogram values, 1D-TFs can not classify objects properly. Along with histogram values, gradient magnitudes are also required. 2D-TFs based on intensity values and gradient magnitudes (as shown in Equation 8.2) were proposed. They were more effective in detecting multiple materials as well as their boundaries [36].

$$q_{\text{separable}} = (d_1, d_2) = V(M(d_1), d_2) \tag{8.2}$$

Some TFs are considered derivatives of scalar value, along with curvature value [37], feature dimensions [38], and occlusion for ambience [39]. To highlight vital patterns of volume data, a few visibility-based methods were used[40]. Equation 8.3 shows the parameters used in the baseline TF.

$$I = \int_a^b q(s) e^{-\int_a^s K(u)du} ds, \tag{8.3}$$

where I = light intensity between *a* and *b.* The *a* and *b* are traversing points in the volume. *q*(*s*)= light distribution. *K* is the attenuation of the light. The material properties and light transportation are redefined by functions *q*(*s*)

and $K(u)$ respectively. The TFs are used to estimate $q(s)$ and $K(u)$. Since baseline TFs are independent of physical properties, baseline TFs are unable to represent objects as ray points and their color values, which represent optical properties of the object. Max et al. [41] proposed Equation 8.4 to include the colors of ray points and their opacity in TF.

$$I = \sum_{i=1}^{n} C_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha j), \tag{8.4}$$

where C_i is the color of the ray points and α_i is the fraction of light of points in the ray, C_i can be solved recursively in reverse order by Equations 8.5 and 8.6.

$$C'_{i+1} = C'_i + (1 + \alpha'_i) C_i \alpha_i \tag{8.5}$$

$$\alpha'_{i+1} = \alpha'_i + (1 + \alpha'_i) \alpha_i, \tag{8.6}$$

where C_i is accumulated color and α_i is the opacity of the ray points which gives weighted color and opacity distribution. To make TFs design effective and usable across the rendering techniques the semiautomatic, automatic, and machine-learning-based approaches have been proposed by researches in recent studies. We will discuss these studies in the next sub-sections.

8.5.2.2.1 SemiAutomatic TFs

Initially, He et al. [42] proposed an approach to semiautomatic TFs. For the generation of TFs, they used a genetic algorithm. This algorithm could identify the best fit TF, which could be as either user-centric or system-centric for each iteration. The component-based approach proposed by Castro et al. [43] uses a different TF for each of the components such as bones, tissues, and soft flesh. A weighted mixture of this component was used to design a TF. In an attempt to design a new semiautomatic TF, Fang et al. [44] used image enhancement and boundary detection for the transformation of the dataset, which was followed by a linear color operation. This two-step process was useful in designing a semiautomatic algorithm.

Durkin [45] proposed a multi-dimensional volume of attribute values and directional derivatives to construct a histogram. Along with the histogram, they considered the object boundaries in the images. To construct a semiautomatic TF, a model with object boundaries and histogram information has been used to build an opacity function. Further, Prauchner et al. [46] extended their approach and evaluated the various settings. To classify the dataset, a histogram of values or gradients along with 3D coordinations have been used by Roettger et al. [47]. The visibility of hidden patterns was enhanced by optimizing the ranges of absolute values and modulating the opacity using the proposed algorithm inCorrea and Ma [48]. This approach used the histogram for visibility of raypoints. In later work, they added iterative mode views to the visibility data.

The use of clustering techniques to segregate the object features has been implemented to enhance the performance of TFs. Several researchers used clustering techniques to detect the boundaries of objects and materials using similarity amongst the boundary information. Two kinds of clusters, one for identifying boundaries and another for identifying their connectivity in the volumes, were proposed by Šereda et al. [49]. The hierarchy of clusters was used to design a semiautomatic TF. A TF for abdominal visualization withcomponents generated by a cluster of features was proposed by Maciejewski et al. [50]. They used both value and gradient as features for building a TF. The semiautomatic transfer of properties was studied by some authors instead of the semiautomatic design of TF. Such a transfer of properties is designed in Praßni et al. [51]. This worked as an aid in the rendering of objects in various volumes. The properties were identified during the preprocessing of data.

A hierarchical structure of segments was proposed by Ip et al. [52]. The user could select the appropriate segment of a histogram to generate a TF for each selection. Recently, Liu et al. [53] proposed a method with data containing multiple voxel features. These multiple features can be used to present dynamic projections so that the user can have multiple views of features space that can be used to select aTF.

As thumbnails, semiautomatic TFs require user intervention. Hence, they are not suitable for real-time volume rendering. The volume rendering process gets slower due to user interaction.

8.5.2.2.2 Automatic TF

Semiautomatic TFs require some user interaction, whereas the automatic TF does not require user intervention. An approximate deviation of spatial values of the surface to be rendered was calculated from the mean surface, and it was colored by using the automatic TF in Pfaffelmoser et al. [54]. Wang et al. [55] proposed a simple method to assign color and opacity values to the set of cells. The set of cells isformed by decomposing the feature space extracted from the volume. The cells are separated, and only cells with vital features are kept for further processing. In addition to this, the cells that are derived from noisy data are rejected. Before assigning color and opacity to the cells, a hierarchical structure was created by merging the cells successively. Similarly, 3D field topology and graph representation–based automatic color assignment of TF was proposed in Fujishiro et al. [56]. An importance-based function was used to select the TF automatically as proposed in Wang and Kaufman [57]. In the proposed method, the important color features were used to calculate the importance of objects.

Multi-functional automatic TFs were proposed by Bramon et al. [58]. The information and divergence factors were used to define the object properties while designing the TF. The intensity and gradient magnitude–based automatic method was proposed by Tianjin [60]. Along with value and gradient, they also considered using voxel information and itsspatial arrangement. Finally, they used clustering for generating the TFs.

In summary, automatic TFs are superior to semiautomatic TFs. They fasten the rendering process, and the performance is also better. Most of them use a clustering technique to design TFs.

8.5.2.2.3 Machine Learning–Based TFs

Machine learning helps developers to come up with learned parameters of TFs that are used to project the objects along with their properties. In this type of TFs, the supervised machine learning algorithm [74] uses the training data and the transformation of training data to learn the TF. During this process, the underlying objects are classified, and then TFs are determined. For this purpose, researchers have used various machine learning techniques such as artificial neural networks, hidden Markov models, support vector machines, and clustering techniques.

Various methods based on machine learning have been proposed, and their effects on automatic TFs have been studied by Sundararajan et al. [62]. Their study suggests that the random forest approach is quite appropriate for the design of TFs. An unsupervised learning technique to reduce the dimensionality of volumes and TF space is proposed in De Moura Pinto and Freitas [59]. The proposed technique also used unsupervised machine learning to assign object properties such as color and opacity on the reduced dimensions. The artificial neural network with back propagation has been used by Wang et al. to find the similarities in the input volumes [64]. The TF is generated by classifying the information from a set of parameters. Selver et al. [61] proposed a method in which both high- and low-frequency structures were represented in different quadrants of transformation for enhancement of volume data. They used brushlet expansion for volume reconstruction with tiling of selected quadrants.

In recently studied techniques, it has been observed that Gaussian-based naive Bayesian classifiers have limited offerings for the complex dataset and that they cannot handle the effect of outliers, although they are quite fast for classification. Significant accuracy is achieved by using k-nearest neighbors (k-NN) classifiers. However, they are quite expensive in terms of computational cost, since they cannot obtain high-dimensional data from the input training data [62].

Similarly, in the case of the support vector machine (SVM), the training time is high, but it is faster than k-NN; for the classification provided the data should be normalized. Single-layer perceptrons also provedto be good, but they are not comparable with other classifiers. Machine learning–based TFs have benefited from simple Bayesian networks, but random forests have been significant in all kinds of challenges related to TFs.

In summary, there are various other categories of TFs based on the design aspect. Many of them require acomplicated method to design and estimate the parameters of the TF. Such methods cannot be considered as an advancement. More information on these methods is available in Ljung et al. [63].

8.5.3 Generative Adversarial Networks(GANs)

GANs are proving to be milestones, particularly in image processing domains. A GAN generator network generates images based on supervised training data, whereas an adversarial network assesses generated images. The learning function adjusts the generator network's parameters based on the advice received from the adversarial network. Figure 8.9 shows the application of GAN to generate visualization data.

A view variant TF space is learned using a generative network, which quantifies the desired changes to be made by analyzing the output image in the training data [67]. The generated space is used directly for rendering, which enables the user to explore the entire space of generated images. Since the model is independent of the rendering process, the algorithm also demonstrates resultant images generated by global illumination lighting on various datasets. A new interactive visualization tool (GAN Lab) can be used to experiment with GAN using selective and popular deep learning models. It allows users to train various generative models and also to visualize intermediate results during the training process [68].

Recently, a GAN model with various stages was proposed to generate 3D volumes for a variety of objects [69]. The generator network captures the object structure and produces 3D objects with high quality. It also creates a mapping between probabilistic space and3D space for objects. This can be done without using a reference image or external models (CAD, for instance). The discriminator network in the model creates a 3D descriptor for the shape, which is learned without any supervision; such a model can be used to recognize 3D objects in the generated volumes.

In summary, looking at the advancements shown by GAN models, it is esteemed that they can play a vital role in medical imaging and visualization. More and more research is expected tomake use of GANs to generate, render, and recognize 3D data.

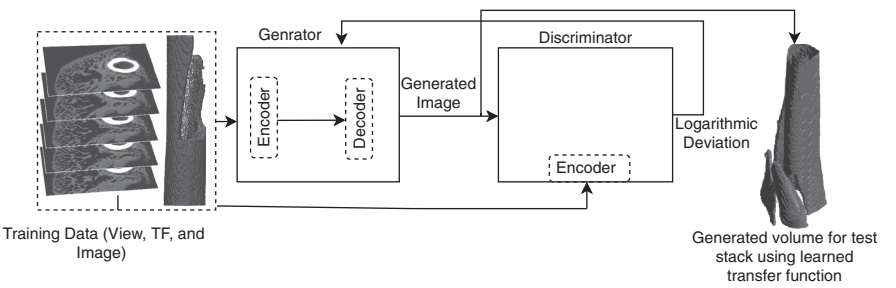


FIGURE 8.9
How GAN works with labeled images to learn the visualization transformation.

8.6 Tools and Libraries for Volume Visualization

Several software tools and graphics libraries are available for volume visualization. Most of them are open source. DICOM viewers are commonly used by the radiographer to visualize the volume data acquired by 3D scanning devices. They have the provisions of 3D model generation from that data. Other than this, 3D Slicer and 3D-DOCTOR are the tools generally used by researchers to reconstruct 3D models. The models generated by these tools can be stored in the STL file for future use. A blender tool is used in Johns [30] to develop an augmented reality–based driving simulator. In addition to this, CUDA, OpenCL, OpenGL, VolPack, and Visualization Toolkit (VTK) are the libraries available to do effective volume visualization.

8.7 Conclusion and Future Directions

In this paper, we presented the systematic review on several volume rendering techniques including both direct and indirect volume rendering. The typical steps involved in the volume visualization process are discussed in detail. Recent advances in TF, which include automatic and semiautomatic designs, are discussed in detail. Along with this, we discussed machine learning–based TFs and their performance. Hardware-based advancements such as CUDA programming using various GPUs and their effect on visualization processing are also specified. In addition to this, a GAN-based architecture for realistic visualization is discussed in brief.

In the data classification (segmentation) step, most of the previous attempts used thresholding-based approach. Thresholding is a traditional segmentation technique that neither analyses the image contents nor assigns unique labels (required in case there are multiple objects present in the volume). In the future, more research attempts should be made to devise image content analysis–based preprocessing and segmentation techniques. The preprocessing technique should be able to remove unwanted artifacts and to enhance the required region of interest by analyzing image contents. The segmentation technique should not only extract the desired portion, but also assign unique labels to each object in the volume.

In the future, researchers can think about using automatic (deep and/or machine learning–based) TFs to develop volume rendering algorithms. The adaptation of automatic TFs will definitely reduce the development time; moreover, researchers can focus on the development of the desired system. Other than this, more research must be conducted to design hybrid volume visualization techniques which initially render volume by surface rendering approach, and after detecting a collision, the technique should generate internal details on the fly through DVR to perform erosion logic.

References

1. P. Reddy, M. Balaram, C. Bones, and Y. B. Reddy, *Visualization in Scientific Computing*, Grambling State University, Grambling, LA, 1996.
2. A. Kaufman, "Volume visualization," *The Visual Computer*, vol. 6, no. 1, pp. 1–1, 1990.
3. B. H. McCormick, T. A. DeFanti, and M. D. Brown (eds), "Visualization in scientific computing," special issue of *Computer Graphics*, ACM, vol. 21, no. 6, 1987.
4. T. A. Defanti and M. D. Brown, "Visualization in scientific computing," *Advances in Computers*, vol. 33, pp. 247–307, 1991.
5. D. D. Ruikar, R. S. Hegadi, and K. C. Santosh, "A systematic review on orthopedic simulators for psycho-motor skill and surgical procedure training," *Journal of Medical Systems*, vol. 42, no. 9, p. 168, 2018.
6. A. Kaufman, "Volume visualization: principles and advances," Course notes, 24, 1997.
7. T. T. Elvins, "A survey of algorithms for volume visualization," *ACM Siggraph Computer Graphics*, vol. 26, no. 3, pp. 194–201, 1992.
8. Q. Zhang, R. Eagleson, and T. M. Peters, "Volume visualization: a technical overview with a focus on medical applications," *Journal of Digital Imaging*, vol. 24, no. 4, pp. 640–664, 2011.
9. J. Zhou and K. D. Tonnies, "State of the art for volume rendering," *Simulation*, pp. 1–29, 2003.
10. H. Pfister and C. M. Wittenbrink, Volume visualization and volume rendering techniques, Tutorial presented at Eurographics, 2000.
11. C. D. Hansen, and C. R. Johnson, *Visualization Handbook*, Elsevier, 2011.
12. W. E. Lorensen and H. E. Cline, "Marching cubes: a high-resolution 3D surface construction algorithm," *Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.
13. M. Levoy, "Volume rendering by adaptive refinement," *The Visual Computer*, vol. 6, no. 1, pp. 2–7, 1990.
14. M. Levoy, "Efficient ray tracing of volume data," *ACM Transactions on Graphics (TOG)*, vol. 9, no. 3, pp. 245–261, 1990.
15. S. Grimm, S. Bruckner, A. Kanitsar, and E. Gröller, "A refined data addressing and processing scheme to accelerate volume raycasting," *Computers & Graphics*, vol. 28, no. 5, pp. 719–729, 2004.
16. L. Westover, "Interactive volume rendering," in *Proceedings of the 1989 Chapel Hill Workshop on Volume Visualization*, ACM, 1989, pp. 9–16.
17. L. Westover, "Footprint evaluation for volume rendering," *ACM Siggraph Computer Graphics*, vol. 24, no. 4, pp. 367–376, 1990.
18. K. Subr, P. Diaz-Gutierrez, R. Pajarola, and M. Gopi, Order independent, attenuation-leakage free splatting using freevoxels, Tech. Rep. IFI-2007.01, Department of Informatics, University of Zürich, 2007.
19. M. Zwicker, H. Pfister, J. Van Baar, and M. Gross, "EWA splatting," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 3, pp. 223–238, 2002.
20. D. Xue and R. Crawfis, "Efficient splatting using modern graphics hardware," *Journal of Graphics Tools*, vol. 8, no. 3, pp. 1–21, 2003.
21. J. P. Schulze, M. Kraus, U. Lang, and T. Ertl, "Integrating pre-integration into the shear-warp algorithm," in *Proceedings of the* 2003 Eurographics/IEEE TVCG Workshop on Volume Graphics, ACM, 2003, pp. 109–118.

22. H. Kye and K. Oh, "High-quality shear-warp volume rendering using efficient supersampling and pre-integration technique," in *Advances in Artificial Reality and Tele-Existence*, 2006, pp. 624–632, Springer, Berlin, Heidelberg, Germany.
23. P. S. Calhoun, B. S. Kuszyk, D. G. Heath, J. C. Carley, and E. K. Fishman, "Three-dimensional volume rendering of spiral CT data: theory and method," *Radiographics*, vol. 19, no. 3, pp. 745–764, 1999.
24. L. Mroz, H. Hauser, and E. Gröller, "Interactive high-quality maximum intensity projection," in *Computer Graphics Forum* (Vol. 19, No. 3), 2000, pp. 341–350, Blackwell Publishers Ltd, Oxford, UK and Boston, MA.
25. T. J. Cullip and U. Neumann, *Accelerating Volume Reconstruction with 3D Texture Hardware*, 1993, University of North Carolina at Chapel Hill. Department of Computer Science.
26. B. Cabral, N. Cam, and J. Foran, "Accelerated volume rendering and tomographic reconstruction using texture mapping hardware," in *Proceedings of the 1994 Symposium on Volume Visualization*, ACM, 1994, pp. 91–98.
27. A. Van Gelder and K. Kim, "Direct volume rendering with shading via three-dimensional textures," in *Proceedings of the 1996 Symposiumon Volume Visualization*, IEEE, 1996, pp. 23–30.
28. C. Rezk-Salama, K. Engel, M. Bauer, G. Greiner, and T. Ertl, "Interactive volume on standard PC graphics hardware using multi-textures and multi-stage rasterization," in *Proceedings of the* ACM SIGGRAPH/EUROGRAPHICS Workshop on Graphics Hardware, ACM, 2000, pp. 109–118.
29. P. Abellán and D. Tost, "Multimodal volume rendering with 3D textures," *Computers & Graphics*, vol. 32, no. 4, pp. 412–419, 2008.
30. B. D. Johns, The creation and validation of an augmented reality orthopaedic drilling simulator for surgical training. Thesis, University of Iowa, 2014.
31. W. E. Lorensen and H. E. Cline, "Marching cubes: a high-resolution 3D surface construction algorithm," in *ACM Siggraph Computer Graphics* (Vol. 21, No. 4), ACM, 1987, pp. 163–169.
32. G. M. Treece, R. W. Prager, and A. H. Gee, "Regularised marching tetrahedra: improved iso-surface extraction," *Computers & Graphics*, vol. 23, no. 4, pp. 583–598, 1999.
33. A. Weinlich, B. Keck, H. Scherl, M. Kowarschik, and J. Hornegger, "Comparison of high-speed ray casting on GPU using CUDA and OpenGL," in *Proceedings of the First International Workshop on New Frontiers in High-performance and Hardware-aware Computing* (Vol. 1), Proceedings of HipHaC'08, 2008, pp. 25–30.
34. NVIDIA. Deep learning software. Available at: https://developer.nvidia.com/deep-learning-software. Accessed October 29, 2018.
35. Q. Zhang, R. Eagleson, T. M. Peters, "Dynamic real-time 4D cardiac MDCT image display using GPU-accelerated volume rendering," *Computerized Medical Imaging and Graphics*, vol. 33, no. 6, pp. 461–476, 2009.
36. G. Kindlmann and J. W. Durkin, "Semi-automatic generation of transfer functions for direct volume rendering," in 1998 IEEE *Symposium on Volume Visualization*, IEEE, 1998, pp. 79–86.
37. G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Möller, "Curvature-based transfer functions for direct volume rendering: methods and applications," in *Proceedings of the 14th* IEEE Visualization (*VIS '03*), IEEE, 2003, pp. 513–520.

38. C. D. Correa and K.-L. Ma, "Size-based transfer functions: a new volume exploration technique," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1380–1387, 2008.
39. C. D. Correa and K.-L. Ma, "The occlusion spectrum for volume classification and visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1465–1472, 2009.
40. C. D. Correa and K.-L. Ma, "Visibility histograms and visibility-driven transfer functions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 2, pp. 192–204, 2011.
41. N. Max, "Optical models for direct volume rendering," *IEEE Transactions on Visualization & Computer Graphics*, vol. 1, no. 2, pp. 99–108, 1995.
42. T. He, L. Hong, A. Kaufman, and H. Pfister, "Generation of transfer functions with stochastic search techniques," in *Visualization'96. Proceedings*, IEEE, 1996, pp. 227–234.
43. S. Castro, A. König, H. Löffelmann, and E. Gröller, Transfer function specification for the visualization of medical data, Vienne University of Technology, 1998.
44. S. Fang, T. Biddlecome, and M. Tuceryan, "Image-based transfer function design for data exploration in volume visualization," in *IEEE Visualization*, 1998, pp. 319–326.
45. G. Kindlmann and J. W. Durkin, "Semi-automatic generation of transfer functions for direct volume rendering," in *1998 IEEE Symposium on Volume Visualization*, IEEE, 1998, pp. 79–86.
46. J. L. Prauchner, C. M. Freitas, and Comba, J. L. D., "Two-level interaction approach for transfer function specification," in *Computer Graphics and Image Processing*, 2005, SIBGRAPI, 18th Brazilian Symposium on, IEEE, 2005, pp. 265–272.
47. S. Roettger, M. Bauer, and M. Stamminger, "Spatialized transfer functions," in *Euro Vis*, 2005, pp. 271–278.
48. C. D. Correa and K. L. Ma, "Visibility-driven transfer functions," in *Visualization Symposium*, 2009. *Pacific Vis'09*. IEEE Pacific, IEEE, 2009, pp. 177–184.
49. P. Sereda, A. Vilanova, and F. A. Gerritsen, "Automating transfer function design for volume rendering using hierarchical clustering of material boundaries," in *Euro Vis*, 2006, pp. 243–250.
50. R. Maciejewski, I. Woo, W. Chen, and D. Ebert, "Structuring feature space: a non-parametric method for volumetric transfer function generation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1473–1480, 2009.
51. J. S. Praßni, T. Ropinski, J. Mensmann, and K. Hinrichs, "Shape-based transfer functions for volume visualization," in *Visualization Symposium (Pacific Vis)*, 2010 IEEE Pacific, IEEE, 2010, pp. 9–16.
52. C. Y. Ip, A. Varshney, and J. JaJa, "Hierarchical exploration of volumes using multilevel segmentation of the intensity-gradient histograms," *IEEE Transactions on Visualization & Computer Graphics*, vol. 18, no. 12, pp. 2355–2363, 2012.
53. S. Liu, B. Wang, J. J. Thiagarajan, P. T. Bremer, and V. Pascucci, "Multivariate volume visualization through dynamic projections," in *2014 IEEE 4th Symposium on Large Data Analysis and Visualization (LDAV)*, IEEE, 2014, pp. 35–42.
54. T. Pfaffelmoser, M. Reitinger, and R. Westermann, "Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields," in Computer Graphics *Forum* (Vol. 30, No. 3), Blackwell Publishing Ltd, Oxford, UK, 2011, pp. 951–960.

55. Y. Wang, J. Zhang, D. J. Lehmann, H. Theisel, and X. Chi, "Automating transfer function design with valley cell-based clustering of 2D density plots," in Computer Graphics *Forum* (Vol. 31, No. 3 pt 4), Blackwell Publishing Ltd, Oxford, UK, 2012, pp. 1295–1304.
56. I. Fujishiro, Y. Takeshima, T. Azuma, and S. Takahashi, "Volume data mining using 3D field topology analysis," *IEEE Computer Graphics and Applications*, vol. 20, no. 5, pp. 46–51, 2000.
57. L. Wang and A. Kaufman, "Importance driven automatic color design for direct volume rendering," in Computer Graphics *Forum* (Vol. 31, No. 3 pt 4), Blackwell Publishing Ltd, Oxford, UK, 2012, pp. 1305–1314.
58. R. Bramon, M. Ruiz, A. Bardera, I. Boada, M. Feixas, and M. Sbert, "Information theory-based automatic multimodal transfer function design," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 870–880, 2013.
59. F. De Moura Pinto and C. M. Freitas, "Design of multi-dimensional transfer functions using dimensional reduction," in *Proceedings of the 9th Joint Eurographics/IEEE VGTC Conference on Visualization*, Eurographics Association, 2007, pp. 131–138.
60. T. Zhang, Z. Yi, J. Zheng, D. C. Liu, W. M. Pang, Q. Wang, and J. Qin, "A clustering-based automatic transfer function design for volume visualization," *Mathematical Problems in Engineering*, 2016.
61. M. A. Selver, "Exploring brushlet based 3D textures in transfer function specification for direct volume rendering of abdominal organs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 2, pp. 174–187, 2015.
62. K. P. Soundararajan and T. Schultz, "Learning probabilistic transfer functions: a comparative study of classifiers," in *Computer Graphics Forum* (Vol. 34, No. 3), 2015, pp. 111–120.
63. P. Ljung, J. Krüger, E. Groller, M. Hadwiger, C. D. Hansen, and A. Ynnerman, "State of the art in transfer functions for direct volume rendering," in *Computer Graphics Forum* (Vol. 35, No. 3), 2016, pp. 669–691.
64. L. Wang, X. Chen, S. Li, and X. Cai, "General adaptive transfer functions design for volume rendering by using neural networks," in *International Conference on Neural Information Processing*, Springer, Berlin, Heidelberg, 2006, pp. 661–670.
65. A. Weinlich, B. Keck, H. Scherl, M. Kowarschik, and J. Hornegger, "Comparison of high-speed ray casting on GPU using CUDA and OpenGL," in *Proceedings of the First International Workshop on New Frontiers in High-performance and Hardware-aware Computing* (Vol. 1), Proceedings of HipHaC'08, 2008, pp. 25–30.
66. Q. Zhang, R. Eagleson, and T. M. Peters, "Dynamic real-time 4D cardiac MDCT image display using GPU-accelerated volume rendering," *Computerized Medical Imaging and Graphics*, vol. 33, no. 6, pp. 461–476, 2009.
67. M. Berger, J. Li, and J. A. Levine, "A generative model for volume rendering," *IEEE Transactions on Visualization & Computer Graphics*, no. 1, pp. 1–1.
68. M. Kahng, N. Thorat, D. H. P. Chau, F. B. Viégas, and M. Wattenberg, "GAN lab: understanding complex deep generative models using interactive visual experimentation," *IEEE Transactions on Visualization and Computer Graphics*, 2018.
69. J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum, "Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling," In *Advances in Neural Information Processing Systems*, 2016, pp. 82–90.
70. NVIDIA. GE Forxe RTX. Available at: https://www.nvidia.com/en-us/geforce/20-series/rtx/.

71. Nvidia CUDA, *Nvidia Cuda c Programming Guide*, Nvidia Corporation, 120(18), 8, 2011.
72. T. Kalaiselvi, P. Sriramakrishnan, and K. Somasundaram, "Survey of using GPU CUDA programming model in medical image analysis," *Informatics in Medicine Unlocked*, vol. 9, pp. 133–144, 2017.
73. D. D. Ruikar, R. S. Hegadi, and K. C. Santosh, Contrast stretching-based unwanted artifacts removal from CT images in recent trends in image processing and pattern recognition (accepted), Springer, 2019.
74. S. Vajda and K. C. Santosh, "A fast k-nearest neighbor classifier using unsupervised clustering," in *International Conference on Recent Trends in Image Processing and Pattern Recognition*, Springer, Singapore, 2016, pp. 185–193.

Applied Machine Learning for Healthcare

Prashant Natarajan and Bob Rogers

Computers are useless. They can only give you answers.

—Pablo Picasso

7.1 Introduction

Collecting, managing, and storing big data is a costly exercise if we can't convert such data into high-value, actionable insights or influence workflows in a timely fashion. Generating knowledge from big data increasingly requires the use of machine learning for various reasons—cognitive, organizational, technical, and operational. Any discussion on big data must include a corresponding discussion on machine learning; frankly, they can seldom be separated anymore. "To be useful, data must be analyzed, interpreted, and acted on . . . [and] attention has to shift to new statistical tools from the field of machine learning that will be critical for anyone practicing medicine in the 21st century."[1]

In order to obtain the most value of out of large, diverse, and fast data, we need to consider options beyond rules-based deductive reasoning, "traditional" systems engineering, and descriptive analytics. Artificial intelligence, specifically the sub-fields of machine and deep learning, provides optimal and cost-effective options to expand the universe of knowledge and solutions in healthcare.

[1] Obermeyer, Z. and Emanuel, E.J. (2016, September 29). "Predicting the Future—Big Data, Machine Learning, and Clinical Medicine." *New England Journal of Medicine*, Vol. 375, p. 1216.

Machine learning enables new use cases by:

- Ameliorating the effects of certain human limitations—cognitive (repetitive accuracy, human limitations and information overload), physical (fatigue), emotional (mood, human biases, etc.)
- Enabling new knowledge creation or data reduction via learning and prediction
- Learning to generate computational biomarkers—finding hidden patterns/insights that are not visible to the eye
- Processing repetitive data management tasks more efficiently, consistently, and with greater performance
- Serving as the foundation for clinical workflows and comprehensive secondary use that includes predictive and prescriptive analytics, intelligent search, speech-to-text conversion, real-time image processing, among other uses

7.2 Chapter Overview

While there is a plethora of books, videos, websites, and other resources on machine learning, most content is either too rudimentary or, on the other end of the spectrum, requires an advanced understanding of linear algebra, probability, statistics, and/or computer science. In addition, there appears to be a paucity of resources on applied machine learning—in which learning algorithms, data sets, and best practices are optimized for and applied in a specific domain/industry such as healthcare. As with any emerging technology, ensuring the successful design/deployment and "production" use of machine learning requires knowledge that enables you to connect theory to practice and convert general principles into domain-specific applications.

If you are interested in learning more about this exciting field, or just want a better understanding of the truth behind the hype of "how <<Brand X>> <<machine/deep learning>> can cure <<disease 1>>," then this chapter is for you.

7.3 A Brief History

AI and machine learning are not new topics. They have been researched, argued over, and used by computer scientists, applied linguists, engineers, etc. for more than 60 years. The mathematical foundations of machine learning are rooted in algebra, statistics, and probability developed over the last 2000 years. However, modern development of AI and machine learning in the 1950s and '60s began with the works of Alan Turing, John McCarthy, Arthur Samuels, Alan Newell, and Frank Rosenblatt, among others. Samuel's self-learning and optimizing Checkers program is recognized as the first working instance of a machine-learning system. Rosenblatt was instrumental in creating the Perceptron, a learning algorithm inspired by biological neurons that became the basis for the field of artificial neural networks, which we will touch upon later in this chapter. "Feigenbaum and others advocated the case for building expert systems—knowledge repositories tailored for specialized domains such as chemistry and medical diagnosis."[2]

[2] "One Hundred Year Study on Artificial Intelligence, Appendix I, A Short History of AI." Stanford (2016). Available at https://ai100.stanford.edu/2016-report/appendix-i-short-history-ai

In the 1990s, research on machine learning moved from knowledge-engineering–based expert systems to statistical and data-driven approaches. The subsequent time period saw the refinement of backpropagation ("the workhorse algorithm of learning in neural networks"[3]) as also the development of the precursors of what we call *deep learning* today by Hinton and others.[4,5] "Something that can be considered a breakthrough happened in 2006: Hinton et al. [. . .] introduced Deep Belief Networks (DBNs), with a learning algorithm that greedily trains one layer at a time, exploiting an unsupervised learning algorithm for each layer, a Restricted Boltzmann Machine (RBM)."[6]

A more in-depth history of machine learning is beyond the scope of this chapter.

7.4 What's Different About Machine Learning Today?

After many fits and starts over the past decades, machine learning has come out of the hibernation that happened during the "AI winter" that followed the last hype cycle in the 1980s and '90s.[7] Machine learning is also no longer a knowledge-engineering effort as it once was. It's been redefined and optimized to be data intensive instead, hence its appropriateness to handle big data. Today, machine learning (and for that matter, deep learning) is maturing to a point where targeted applications are practical and real. There is definitely increasing market demand, and machine learning is here to stay.

Machine learning is ready for prime time for the following reasons:

1. **Moore's Law.** Continuing advances in computing and storage are allowing us to store and process very large data sets in a cost effective and scalable manner.
2. **Availability of more data.** Machine learning is primarily a data-driven endeavor. As a result, the creation/availability of large data sets coupled with the ability to share/transport such data are allowing us to get further than ever before in predicting or determining new knowledge.
3. **New sources in native unstructured data formats.** Several big-data sources such as the ones discussed in Table 2.1 in Chapter 2 ("Table 2.1 Sources for Big Data in Healthcare" on page 23) are unstructured. Machine learning is ideally suited and is rapidly evolving to better support the generation of insights and analytics directly off native formats such as videos, images, voice, and large un- or semi-structured text.

[3] Nielson, M. (2016, January). "How the Backpropagation Algorithm Works." Retrieved from http://neuralnetworksanddeeplearning.com/chap2.html

[4] Hinton, G.E. and Salakhutdinov, R.R. (2006, July 28). "Reducing the Dimensionality of Data with Neural Networks." *Science,* Vol. 313. Available at https://www.cs.toronto.edu/~hinton/science.pdf

[5] Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986, October 9). "Learning Representations by Back-Propagating Errors." *Nature,* Vol. 323, pp. 533–536. Retrieved from http://www.nature.com/nature/journal/v323/n6088/pdf/323533a0.pdf. DOI: 10.1038/323533a0

[6] Bengio, Y. (2009). *Foundations and Trends in Machine Learning,* Vol. 2, No. 1, p. 6. DOI: http://dx.doi.org/10.1561/2200000006

[7] Katz, Y. (2012) "Noam Chomsky on Where Artificial Intelligence Went Wrong." *The Atlantic* [online]. Available at http://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/?single_page=true

We interact with machine learning (and learning algorithms) on a daily basis; examples include self-driving cars, email spam filters, Netflix movie suggestions, Amazon shopping recommendations, and postal-code–based mail sorting using handwriting recognition. Machine learning applications are rapidly being deployed and used in the commercial space across diverse verticals—retail and e-commerce, government, finance, healthcare (providers, payers, and pharma, and personal/public health), cyber security, transportation, agriculture, space exploration, and manufacturing, among many others.

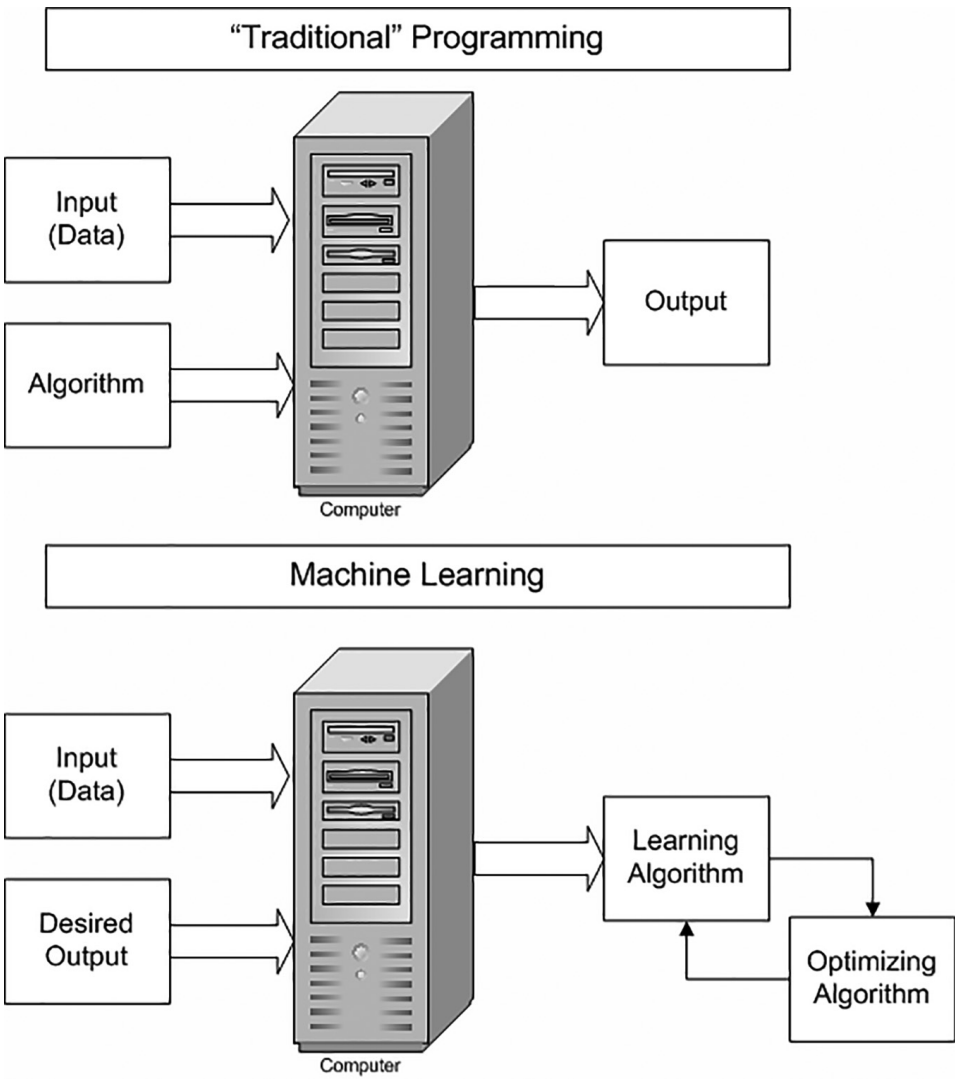


Figure 7.1 Traditional programming and machine learning: a comparison.

7.4.1 What Is Machine Learning?

Arthur Samuel is credited with defining machine learning as the the field of study that gives computers the ability to learn without being explicitly programmed. While its simplicity ensures that this definition is oft-quoted, others provide added perspectives and useful clarifications:

> Machine Learning is a paradigm that enables systems to automatically improve their performance at a task by observing relevant data. Indeed, machine learning has been the key contributor to the AI surge in the past few decades, ranging from search and product recommendation engines, to systems for speech recognition, fraud detection, image understanding, and countless other tasks that once relied on human skill and judgment.[8]
>
> One useful perspective on machine learning is that it involves searching a very large space of possible hypotheses to determine one that best fits the observed data and any prior knowledge held by the learner.[9]

As we will review in the next section, machine learning is different from traditional software programming due to its emphasis on:

- Learning algorithms versus "traditional" algorithms
- Reasoning that is primarily induction and abduction, with a selective emphasis on deduction
- Dealing with uncertainty ("the unknown unknowns") via the use of mathematical models that are driven by probability and statistics as compared to deterministic rules
- Prediction: using data you have to extrapolate data you don't have in order to infer probability of outcomes

7.5 How Do Machines Reason and Learn: A Crash Course in Learning Algorithms

"A learning algorithm is an algorithm that is able to learn from data."[10] Machine learning is different from traditional programming (see Figure 7.1). "In machine learning, we provide the input (data), the desired result and out comes the [learning] algorithm."[11] Learning algorithms—also known as *learners*—are algorithms that create new knowledge or demonstrate new skills by learning from old (training) data and new (generalized) data. A learning algorithm uses data and experience to self-learn and also to perform better over time. During the process, a learner also optimizes itself to progressively come up with better predictions (see Figure 7.1).

[8] "One Hundred Year Study on Artificial Intelligence, Appendix I, A Short History of AI." *op. cit.*

[9] Mitchell, T.M. (1997). *Machine Learning.* McGraw Hill. p. 27.

[10] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning.* Book in preparation for MIT Press. Information available at http://www.deeplearningbook.org. p. 98.

[11] Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.* Basic Books. p. 5.

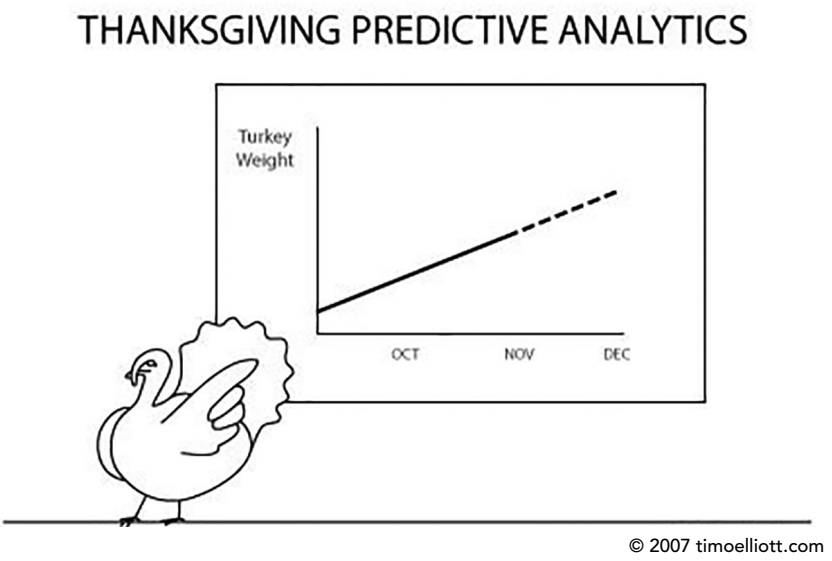


Figure 7.2 Prediction is not certainty. (*Source:* Timo Elliot; used with permission.)

Learners are the foundation of any machine-learning system, and they help us achieve generalization via induction or abduction. *Generalization* is a core concept in machine learning; to be useful, machine-learning algorithms can't just *memorize* the past, they must *learn from* the past. Generalization is the ability to respond properly to new situations based on experience from past situations.

We will introduce two basic concepts here: *Training Dataset* (the data you have that is used as input to the learner to train the model), and *Test Dataset* (dataset is used by the learner for validation and optimization).

Training model refers to the ML artifact that is created coming out of the training process. Training is supplying the learning algorithm with training data to learn from.

A *cost function* is something (usually, a function) that you want to minimize in the ML system. For example, your cost function might be the sum of squared errors over your training data set.[12]

In summary, a machine-learning system consists of the following basic components:

- Learning and optimizing algorithms
- Training and test datasets
- Training model
- Cost function[13]

[12] "Can someone explain to me the difference between a cost function and the gradient descent equation in logistic regression?" (2012). StackOverflow, StackExchange. http://stackoverflow.com/questions/13623113/can-someone-explain-to-me-the-difference-between-a-cost-function-and-the-gradien

[13] Goodfellow, I., et al. *op. cit.* p. 99.

Table 7.1 Some Example Learning Problems[a]

Example Learning Problem	Task T	Performance Measure P	Training Experience E
Learning Checkers	Playing checkers	% of games won against opponents	Learner playing practice games against itself
Handwriting recognition	Recognizing and classifying handwritten words within images	% of words correctly classified	Training dataset of handwritten words with given classifications
Self driving car	Driving from Cupertino to Livermore on public roads	Average distance travelled before an error (as judged by a human overseer)	Sequence of videos, still images, and steering commands recorded while observing a human driver

[a] Table adapted from Mitchell, T.M. (1997). *Machine Learning.* McGraw Hill. pp. 3–4.

In summary, "a well-defined learning problem requires a well-specified task, T; performance metric, P; and source of training experience, E."[14] A formal (and personal favorite) definition of learning states, "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."[15]

7.6 Mastering the Basics of Machine Learning

Now that we've reviewed Mitchell's definition, let's take a look at Task T, Performance P, and Experience E using examples in Table 7.1.

7.6.1 Task, T

A *Task* is something that we want the machine learning system to do: "the process of learning itself is not the task. Learning is our means of attaining the ability to perform the task."[16] To illustrate further, a task for a self-driven car would be to make the journey autonomously from Peoria, Illinois, to Livermore, California. The learning algorithms and the rest of the machine learning system that the car uses to learn and recognize street signs, sidewalks, other vehicles, people, etc. constitute the means by which this task is completed successfully. Examples of tasks that can be done by a machine-learning system are classification, classification with missing inputs, regression, transcription, machine translation, structured output, anomaly

[14] Mitchell, T.M. *op. cit.*
[15] *Ibid.* p. 2.
[16] Goodfellow, *op. cit.* p. 99.

detection, synthesis, density estimation, denoising, imputation of missing values, named entity recognition, etc.[17]

In machine learning, "a *Dataset* is a collection of Examples."[18] "A *Task* is defined in terms of how the learning system should process an Example. An *Example* is defined as a collection of features that have been quantitatively measured from some object or event that we want the system to process. A *Feature* is the combination of an attribute and its value." For instance, "Color is blue" is a feature, where color is the attribute and blue is the value. Another example: "The features of an image are usually the values of the pixels in the image."[19]

Note that some authors prefer to use the terms *Feature* and *Example* as synonyms.

In machine learning, we are very interested in understanding and addressing *dimensionality,* which is defined as the number of features that contain most useful or actionable information, as well as *Parameters,* which are attribute values (of high-value features) that control the behavior of the learning system. Parameters are important, as they can be modified by the learning algorithm to determine better performance (accuracy) or contextualization of prediction. So for example, the predicted sale price of a house may be impacted more by parameters such as nearness to school or number of bedrooms, rather than the type of roof shingles or color of exterior paint.

As we saw in earlier chapters, with big data, it has become easier to collect/store/manage data than to obsess over or minimize what is collected. "We are in the era of massive automatic data collection, systematically obtaining many measurements, not knowing which ones will be relevant to the phenomenon of interest. Our task is to find a needle in a haystack, teasing the relevant information out of a vast pile of glut. This is a big break from the past, when it was assumed that one was dealing with a few well-chosen variables—for example, using scientific knowledge to measure just the right variables in advance."[20]

Understanding Dimensionality allows us to discuss *dimensional reduction,* which is reducing the number of features required in an example. Dimensional reduction refers to the algorithmic processes by which a dataset having high dimensions is converted into a dataset with fewer dimensions so that as much information as possible about the original data is preserved. This is done in machine learning to address what is known as Bellman's *Curse of Dimensionality,* in which "many learning algorithms that work fine when the dimensions are low become intractable when the input is high-dimensional. But in machine learning [the curse] refers to much more. Generalizing correctly becomes exponentially harder as the dimensionality (number of features) of the examples grows."[21]

> Dimension reduction can be accomplished via *Feature Extraction* and *Feature Selection.* Feature extraction creates new features resulting from the combination of the original

[17] *Ibid.* pp. 100–103.
[18] *Ibid.* pp. 104.
[19] *Ibid.* pp. 99.
[20] Donoho, D.L. (2000). "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality." Stanford University. Retrieved from http://statweb.stanford.edu/~donoho/Lectures/CBMS/Curses.pdf. p. 17.
[21] Domingos, P. (2012, October). "A Few Useful Things to Know about Machine Learning." *Communications of the ACM,* Vol. 55, No. 10, p. 81.

> features; and feature selection produces a subset of the original features. Both attempt to reduce the dimensionality of a dataset in order to facilitate efficient data processing tasks.[22]
>
> *Weights* determine how each feature affects the prediction. If a feature receives a positive weight, then increasing the value of that feature increases the value of our prediction. If a feature receives a negative weight, increasing the value of that feature reduces the value of our prediction. If the weight is 0, there is no effect on prediction.[23]

One of the primary purposes of machine learning is for the system to perform on "unknown unknowns," or new, previously unseen data—not just the training dataset with which the model was trained. A good machine-learning system generalizes well from the training dataset to any data from the problem domain. This allows the system to extrapolate and predict on "never seen before" data. Wilson explains further:

> Learning in backprop seems to operate by first of all getting a rough set of weights which fit the training patterns in a general sort of way, and then working progressively towards a set of weights that fit the training patterns exactly. If learning goes too far down this path, one may reach a set of weights that fits the idiosyncrasies of the particular set of patterns very well, but does not interpolate (i.e., generalize) well.[24]
>
> Generalization is the ability to perform well on previously unobserved inputs. Generalization Error (also called Test Error) is defined as the expected value of the error on a new input.[25]

Overfitting has significant impacts on the performance of the machine learning system. Overfitting happens when a learner mimics random fluctuations, anomalies, and noise in the training dataset, thus adversely impacting the performance of the system on new data. "[W]ith large complex sets of training patterns, it is likely that some errors may occur, either in the inputs or in the outputs. In that case, it is likely that [the learner] will be contorting the weights so as to fit precisely around training patterns that are actually erroneous."[26]

7.6.2 Performance, P

In order to evaluate the abilities of a machine-learning algorithm, we must design a quantitative measure of its *Performance, P.* Performance is usually measured on the task being carried out by the machine-learning system and is typically measured in terms of *Accuracy,* which is the "proportion of examples for which the model produces the correct output," or *Error Rate,* which is the "proportion of examples for which the model produces the incorrect output."[27]

---

[22] Dash, M. and Liu, H. (n.d.). "Dimensional Reduction." Retrieved from: http://www.public.asu.edu/~huanliu/papers/dm07.pdf

[23] Goodfellow, I., et al. *op. cit.* pp. 107–108.

[24] Wilson, H.B. (1998, updated June 24, 2012). *The Machine Learning Dictionary.* http://www.cse.unsw.edu.au/~billw/mldict.html#generalizebp

[25] Goodfellow, I., et al. *op. cit.* p. 110.

[26] Wilson, B. *op. cit.*

[27] Goodfellow, I., et al. *op. cit.* pp. 103–104.

We conclude our brief discussion on performance by reviewing *Noise,* which in machine learning refers to "errors in the training data for machine learning algorithms. If a problem is difficult enough and complicated enough to be worth doing with machine learning techniques, then any reasonable training set is going to be large enough that there are likely to be errors in it. This will of course cause problems for the learning algorithm."[28]

7.6.3 Experience, E

Experience in machine learning is primarily determined by the amount of supervision (during the learning process) and the availability of labeled data in the dataset.

In *supervised learning,* "algorithms experience a dataset containing features but each example is associated with a *Label* (AKA *Target*)."[29] Wilson defines supervised learning as "a kind of machine learning where the learning algorithm is provided with a set of inputs for the algorithm along with the corresponding correct outputs, and learning involves the algorithm comparing its current actual output with the correct or target outputs, so that it knows what its error is, and modify [*sic*] things accordingly."[30] Input data is labeled based on existing knowledge (for example, is the email in the training dataset spam or not-spam?) The model continues to train until it achieves a desired level of performance on the training dataset, and the training model is then fed new and unknown data, as described earlier.

In *unsupervised learning,* input data is not labeled, and furthermore, "the system is not told the 'right answer'—for example, it is not trained on pairs consisting of an input and the desired output. Instead the system is given the input patterns and is left to find interesting patterns, regularities, or clusterings among them."[31]

In *semi-supervised learning,* as the experience suggests, input data may be only partially labeled, and the expected results may or may not be known. The machine learning system will include both supervised and unsupervised learners.

Active learning is a semi-supervised learning experience in which "the model chooses by itself what unlabelled data would be most informative for it, and asks an external 'oracle' (for example, a human annotator) for a label for the new data points."[32] The learner aims to "achieve high accuracy using as few labeled instances as possible, thereby minimizing the cost of obtaining labeled data (something that remains challenging in healthcare)."[33]

Deep learning is a type of machine-learning experience that uses learning algorithms called *artificial neural networks* that attempt to simulate or replicate the functioning of the human

[28] Wilson, B. *op. cit.* http://www.cse.unsw.edu.au/~billw/mldict.html#firstN

[29] Goodfellow, I., et al. *op. cit.* pp. 105.

[30] Wilson, B. *op. cit.* http://www.cse.unsw.edu.au/~billw/mldict.html#firstS

[31] *Ibid.*

[32] Gal, Y. (2016). *Uncertainty in Deep Learning,* PhD Thesis, University of Cambridge. Retrieved from http://mlg.eng.cam.ac.uk/yarin/blog_2248.html. p. 11.

[33] Settles, B. (updated 2010, January 6). "Active Learning Literature Survey." *Computer Sciences Technical Report 1648,* University of Wisconsin–Madison. Retrieved from http://burrsettles.com/pub/settles.activelearning.pdf

brain. Think of deep neural networks as "ANNs with lotsa depth."[34] Before we review deep learning, let's take a quick look at artificial neural networks and understand why they serve as a basis for understanding what deep learning does.

7.7 Artificial Neural Networks: An Overview

- **Biological neuron.** "From the artificial neural network point of view, a biological neuron operates as follows: electrical pulses from other neurons cause the transfer of substances called neurotransmitters (of which there are several varieties) from the synaptic terminals of a neuron's axon (think "output") across a structure called a synapse to the dendrites of other neurons (call them downstream neurons). The arrival of the neurotransmitter in the dendrite of the downstream neuron increases the tendency of the downstream neuron to send an electrical pulse itself ("fire"). If enough dendrites of a neuron receive neurotransmitters in a short enough period of time, the neuron will fire."[35]
- **Artificial neuron.** "A simple model of a biological neuron used in neural networks to perform a small part of some overall computational problem. It has inputs from other neurons, with each of which is associated a weight—that is, a number which indicates the degree of importance which this neuron attaches to that input."[36]
- **Artificial neural network.** "An artificial neural network is a collection of simple artificial neurons connected by directed weighted connections. When the system is set running, the activation levels of the input units is clamped to desired values. After this the activation is propagated, at each time step, along the directed weighted connections to other units. The activations of non-input neurons are computing using each neuron's activation function. The system might either settle into a stable state after a number of time steps, or in the case of a feed forward network, the activation might flow through to output units. Learning might or might not occur, depending on the type of neural network and the mode of operation of the network."[37]

7.8 Deep Learning

"*Deep learning* is a specific kind of machine learning. In order to understand deep learning well, one must have a solid understanding of the basic principles of machine learning."[38] It is a "kind of learning where the representations you form have several levels of abstraction, rather than a direct input to output."[39] Think of "deep" in deep learning as having many more layers

[34] @natarpr (author) on Twitter. (2016, December 8).
[35] Wilson, B. *op.cit.* http://www.cse.unsw.edu.au/~billw/mldict.html#bioneuron
[36] Wilson, B. *op.cit.* http://www.cse.unsw.edu.au/~billw/mldict.html#neuron
[37] *Ibid.*
[38] Goodfellow, I., et al. *op. cit.* p. 98.
[39] Norvig, P. (2016, March 18). "Deep Learning and Understandability versus Software Engineering and Verification." Available at http://youtu.be/X769cyzBNVw

(or *Depth*) than were possible with ANNs and as the ability to deal with very large datasets due to Moore's law and data availability. The principle driving deep learning is "guiding the training of intermediate levels of representation using unsupervised learning, which can be performed locally at each level."[40]

Deep learning particularly does well on sequential, unstructured, or analog data such as images, audio, and video and is becoming very popular today due to its high performance. "Deep Learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer."[41]

Deep learning currently excels at supervised learning. However, we see as much or greater potential for using deep learning in unsupervised learning, "primarily because large datasets contain greater amounts of unlabeled data that require labeling, which is time- and effort-intensive."[42]

Let's discuss some types of deep neural nets, including Feed Forward Neural Networks; Recurrent Neural Networks; Convolutional Neural Networks; and Reinforcement Neural Networks.

- **Feed Forward Neural Network.** A "kind of neural network in which the nodes can be numbered, in such a way that each node has weighted connections only to nodes with higher numbers. [. . .] In practice, the nodes of most feedforward nets are partitioned into layers—that is, sets of nodes, and the layers may be numbered in such a way that the nodes in each layer are connected only to nodes in the next layer. The first layer has no input connections and is termed the input layer. The last layer has no output connections and is termed the output layer. The layers in between the input and output layers are termed hidden layers, and consist of hidden units."[43]
- **Recurrent Neural Network.** Sequence-based neural networks that play a key role in natural language processing, machine translation, video processing, and many other tasks.[44] "The idea behind RNNs is to make use of sequential information. In a traditional neural network, we assume that all inputs (and outputs) are independent of each other. But for many tasks that's a very bad idea. If you want to predict the next word in a sentence you better know which words came before it. RNNs are called *recurrent* because they perform the same task for every element of a sequence, with the output being depende[nt] on the previous computations. RNNs [. . .] have a 'memory' which captures information about what has been calculated so far."[45]

[40] Bengio, Y. (2009). "Learning Deep Architectures for AI." *Foundations and Trends® in Machine Learning,* Vol. 2, No. 1, p. 7. Information available at http://dx.doi.org/10.1561/2200000006

[41] LeCun, Y., Bengio, Y., and Hinton, G. (2015 May 28). "Deep Learning." *Nature,* Vol. 521, pp. 436–444. Retrieved from http://www.nature.com/nature/journal/v521/n7553/abs/nature14539.html

[42] Ng, A. (2014, August 26). "Deep Learning: Machine Learning via Large-scale Brain Simulations." Invited Talk: Deep Learning, Stanford University. http://youtu.be/W15K9PegQt0

[43] Wilson. *op. cit.* http://www.cse.unsw.edu.au/~billw/mldict.html#firstF

[44] Yarin, G. *op. cit.* p. 5.

[45] "Recurrent Neural Networks Tutorial, Part 1: Introduction to RNNs." (2015, September 7). Retrieved from http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

- **Convolutional Neural Network.** A "type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex."[46] CNNs excel at dealing with sequential, analog, or unstructured data and are showing great promise in healthcare—particularly in image/audio/video recognition, recommender systems, and natural language processing. The model is made of a "recursive application of convolution and pooling layers, followed by simple NNs. A convolution layer is a linear transformation that preserves spatial information in the input image. Pooling layers simply take the output of a convolution layer and reduce its dimensionality."[47] An excellent example of how deep learning and CNNs are being used in healthcare is the work being done by Pratik Mukherjee MD and his team at UCSF (see UCSF Case Study on page 149).
- **Reinforcement Neural Network**. Neural networks in which the learner learns and performs tasks via trial and error, much like a child learning to ride her bicycle. Reinforcement learning is inspired by behaviorist psychology and focuses on how software agents ought to take *actions* in an *environment* so as to maximize some notion of cumulative *reward.*

  "Reinforcement learning differs from supervised learning in that correct input/output pairs are never presented, nor sub-optimal actions explicitly corrected."[48]

7.9 A Guided Tour of Machine-Learning Algorithms in Healthcare

Every machine-learning algorithm is good at answering a specific kind of question. Let's take a look at some of the most important algorithms in healthcare and the questions they are being used to answer.

Don't be intimidated by the sheer number of machine-learning algorithms that are out there. While there are currently 54 Wikipedia pages dedicated to specific machine-learning algorithms, all of these are really variations on a few major themes.

The list of algorithms described below is not exhaustive. Our goal is to give examples of the most commonly used methods to answer different types of questions in healthcare analytics. In addition to the type of question being answered, there is another useful way to characterize machine-learning algorithms: whether or not they need input data that is labeled with known answers to create them. Methods that require input data with known labels are called *supervised training* algorithms, and those that do not require any prior knowledge of what answers are expected are called *unsupervised.* The majority of the algorithms below are supervised learning algorithms. We indicate the unsupervised learning algorithms with an asterisk (*).

7.9.1 Classifier

Does data belong to class A?
Example: Is this really a heart failure patient?

[46] "Convolutional Neural Network." https://en.m.wikipedia.org/wiki/Convolutional_neural_network
[47] Yarin, G. *op. cit.*
[48] "Reinforcement Learning." (n.d.) Wikipedia. https://en.m.wikipedia.org/wiki/Reinforcement_learning

- **Logistic regression.** Logistic regression is the workhorse of classifiers. It is a linear classifier, which means it uses a single, straight cut to divide the world of possible features into two groups. If a patient's characteristics fall on one side of this cut, they are in class A (i.e., they are judged to have heart failure), otherwise they are not in class A. In problems with many features (some problems can require millions of features), logistic regression is the preferred method because it works well and is straightforward to compute.
- **Support Vector Machine (SVM).** SVM is a linear classifier with a twist: the world of possible features is split by a single line as in logistic regression, but this line can be curved. This additional flexibility makes SVM highly adaptable, but because of the way the curvature is introduced (though something called a kernel), they are still simple to compute and to interpret.
- **Decision tree, random forest, boosted trees.** Trees and forests are an entire family of algorithms, all based on the idea of creating a tree of decisions about features that lead to a specific classification. For example, to identify heart failure, the algorithm may start with ejection fraction. Is ejection fraction above or below 50? For each of these paths, a new question would be considered, such as: does the echocardiogram show Left Ventricle Hypertrophy? At the end of each series of questions, the patient falls into either a "heart failure" or a "not heart failure" bucket. Random forests improve upon decision trees by dividing the input data into many different subsets and creating a different decision tree for each of these subsets. All of the different resulting decision trees then vote to determine the final classification of the input. This process reduces the risk of making the final buckets too small and subsequently being fooled by random variations in the original labeled training data. Boosting is a trick for creating decision trees and random forests that can significantly improve their ability to generalize from example data. We call them out specifically here because boosted tree classifiers tend to be among the best-performing algorithms in public classification competitions such as Kaggle (www.kaggle.com).
- **Deep Learning.** Deep learning, and indeed neural networks in general, can take raw data as input and produce a class (or a vector of probabilities for many classes) as output. All neural network models consist of multiple layers of "neurons": each neuron in a layer receives the outputs of neurons in previous layers, combines these inputs, and uses a threshold to determine whether to output a value closer to 0 or closer to 1 for processing by the next layer. Deep-learning algorithms are unique in their ability to automatically generate features of interest in input data, as long as they are provided with a sufficient number of training examples (usually in the millions). Deep learning is already extensively used in image, video, and audio understanding in healthcare, and it will eventually become more common in other classification problems in healthcare as larger sets of labeled training data become available for healthcare applications.

Common uses: Classifiers are the most commonly used machine learning algorithms in all analytics applications, including healthcare. In healthcare, classifiers are used to:

- Suggest possible patient diagnoses
- Identify patients with high readmission risk

- Automatically alert care providers early in the development of sepsis
- Define the thresholds for "abnormal" lab results
- Automatically differentiate between clinical and administrative documents
- Recommend the most effective wellness or disease management intervention for a patient
- Many, many more

7.9.2 Memory-Based Learning*

How does this new piece of data compare to past data?
Example: Who are the patients most like this patient?

- **Associative memory.** An associative memory system compares incoming data with past data to identify what the new data is most like. The comparison can be based on any subset of the attributes of the data, so no assumptions need to be made about what is "important" in the data, and very large numbers of features can be included. This makes these algorithms especially useful in healthcare applications, because the number of conditions and measurement results that could be applicable to a patient is very large.

Common uses: Memory-based learning is commonly used in healthcare to:

- Create cohorts of patients with which to compare a specific patient. This is the "patients like mine" question that plays a role in treatment planning, pharmaceutical research, and risk adjustment modeling.
- Identify insurance fraud.
- Calculate risk of readmission and other costly future events.

7.9.3 Topic Modeling

What is this document about?
Example: What conditions are being addressed for this patient in this SOAP note?

- **Latent Dirichlet Allocation (LDA).** LDA assumes that content is made up of a combination of underlying topics. A single doctor's note may be 80% about a patient's diabetes and 20% about pain management. LDA can identify the combinations of terms and phrases that make up the underlying topics. LDA can be applied to many different sources of information, from single documents to groups of documents, to even a patient's entire clinical history.
- **Probabilistic Latent Semantic Analysis (pLSA), Latent Semantic Analysis (LSA).** These algorithms are similar to LDA but make stricter simplifying assumptions about how topics are distributed in documents and how words are distributed in topics. With modern computing and large datasets available for analysis, these assumptions are no longer necessary, so LDA is the dominant methodology.

Common uses in healthcare:

- Reliably identify the conditions that a patient has based on clinical text combined with structured data for use in acute disease detection, such as:
  - Sepsis detection
  - Heart failure detection for prevention of hospital readmission
  - Drug-seeking and drug fraud

7.9.4 Forecasting

How much will this time series change in the next time period?
Example: How likely is this CKD patient to progress in the next six months?

- **Linear regression.*** Draws a straight line through the time series of past data, assuming that the current linear trend will continue. This approach requires that the predicted output is a continuous variable.
- **Neural networks.*** A neural network, which can be as simple as a Multi-Layer Perceptron (MLP) or as complex as a recurrent deep-learning model (e.g., Long Short-Term Memory, LSTM), takes past values as inputs and produces the predicted next value as output. All neural network models consist of multiple layers of "neurons": each neuron in a layer receives the outputs of neurons in previous layers, combines these inputs, and uses a threshold to determine whether to output a value closer to 0 or closer to 1 for processing by the next layer.
- **Exponential smoothing.*** This is a simple but surprisingly useful method for predicting the next value in a time series based on a weighted average of the most recent past values. It gives the most weight to the most recent measured value, then reduces the weight by multiplying by a number between 0 and 1 for each subsequent previous value, resulting in an exponential decrease in the impacts of older previous time-series values.
- **Auto-Regressive Integrated Moving Average (ARIMA) modeling.*** This is a general group of forecasting methods (of which exponential smoothing is actually a member) that uses weighted averages of past time-series values, past differences between time-series values, past differences between rates of change of past values, and so on, to calculate future values of the time series.

Common uses in healthcare: The expected next value of a time series is often used as part of a larger predictive modeling or clinical decision-support application. For example:

- Chase lists for disease management: the predicted future values of key diagnostic measurements such as Hemoglobin A1C or creatinine are used to determine who should be included on a chase list for chronic disease management.
- Risk prediction for individuals: a number of healthcare companies, payors, providers, and third-party analytics vendors use predictive models to compute the likelihood that a patient will convert to a new diagnosis within future time periods ranging between six months and two years.

- Very time-sensitive detection of disease: acute applications, such as sepsis detection in the hospital, commonly include time-series forecasts of key measurements such as temperature, white blood cell count, or respiratory rate.

7.9.5 Probability Estimation

What is the most likely interpretation of the data?
Example: What is the most likely diagnosis, given the patient's signs, symptoms, and measurements?

- **Probabilistic Graph Model (PGM).** PGM algorithms, such as Bayes networks, identify key observations, measurements, and outcomes and link them together to identify causal relationships. Each of these factors would be represented as a node in a graph, with connections between nodes indicating causal relationships. These graphs can be learned directly from data or constructed by human experts. The PGM algorithm then uses actual data to determine the amount of influence each combination of variables (nodes) has on the others.
- **Logistic regression.** Logistic regression models assume that the log of the odds of an event occurring (or an interpretation being applicable) can be calculated from a simple weighted average of a set of observations. In practice, this means that they can be used to predict the likelihoods of categorical values such as diagnoses or specific outcomes.

Common uses in healthcare: Sepsis detection, readmission prevention.

- CDS and diagnosis tools: For example, a model for diagnosing COPD might include historical and demographic information such as age, sex, smoking history, exposure to chemicals, and signs and symptoms such as coughing, dyspnea, and blood oxygen saturation, along with comorbidities such as bronchitis, diabetes, and lung cancer. Each of these factors would be represented as a node in the graph, with connections between nodes indicating a causal relationship. The presence or absence of a combination of these factors will influence the probability that a COPD diagnosis is applicable to the patient.
- Disease risk forecasting: PGM and logistic regression can both be used to compute the probabilities of different diagnoses or interpretations of data.

7.9.6 Image and Video Understanding

What is in this image? What is happening in this video?
Example: Is there a nodule in this chest x-ray?

- **Deep learning.** Deep-learning systems, especially convolutional neural networks (CNNs), are very powerful methods for recognizing objects or patterns in complex images. The power of deep-learning algorithms is that, given enough data, they can learn what is important for understanding an image without being explicitly told. In practice, to recognize a cat in a photo, or a nodule in a chest x-ray, a deep-learning system may need to

be shown millions of images, each labeled with the desired answer for that image. This training process can be very computationally intensive and require long times to complete (hours, days, and even weeks), but once the algorithm is trained, it can easily be used to quickly recognize the objects it has been taught—a process called "inference" or "scoring."

Common uses in healthcare: This area is exploding right now. At the time of writing, there are compelling deep-learning results being developed for:

- Automated detection of "findings" in radiology images: for example, features such as nodules, pneumonia, or pneumothorax can be automatically detected in chest x-rays using deep-learning systems. These results can be used to route time-sensitive cases to radiologists or to enhance the productivity of radiologists without sacrificing accuracy. Development is underway to develop commercially viable radiology detection systems for all modalities, including MRI, CT, and ultrasound.
- Workflow monitoring and procedural compliance: prototype systems have been developed that can use video and other data streams (such as RFID-based location tracking) to track compliance with standard workflows and procedures. For example, nosocomial infection prevention, in which deep-learning systems have been developed to recognize activities in hospitals that increase the risk of nosocomial infection. These systems, for instance, can flag when a wound is handled but hands are not washed before an IV is placed.
- Patient safety monitoring: Patients can be monitored via video to predict their fall risk in general and to identify when they are getting out of the bed or performing a risky activity, so that personnel can be alerted to assist.

7.9.7 Speech to Text

What is the transcribed text for this audio stream?
Example: What did the clinician dictate?

- **Hidden Markov Models (HMMs).** HMMs assume that there are underlying processes that we can't see—but which are nonetheless consistent and predictable—that create outputs that we can see. For example, in a sentence, if the word "mellitus" is detected, the previous word is far more likely to be "diabetes" than it is to be "disabilities." This is very valuable in speech-to-text processing, because it is not possible to clearly hear or identify each word in an audio stream. The HMM can help choose the right interpretation of the sounds to result in the correct overall transcription. HMM has been used historically in a number of commercial dictation transcription systems.
- **Deep learning, especially Long Short-Term Memory (LSTM).** LSTM models, like the Convolutional Neural Networks described above, can automatically learn what attributes of an audio stream are important for predicting what words it represents. Given sufficient data, which is readily available to online service providers such as Google and Baidu, it is possible to train LSTM models to accurately convert spoken language into

text in almost any language. This technology has become the state of the art for spoken language understanding applications and will likely play an increasing role in clinical transcription applications.

Common uses in healthcare:

- Dictation and clinical note transcription.
- Interpretation and automated documentation of clinical encounters, including speech from clinicians, patients, and support staff.
- Voice controls for computer systems in the clinic and in the surgical theater.
- Call-center resources and agent coaching to help call-center operators provide appropriate information and resources to patients or members.
- Patient coaching: applications are being developed in which patients are given context-dependent coaching for disease management, wellness, and particularly behavioral health applications.

7.9.8 Recommender Systems

What was the behavior of other people like you?
Example: What chronic disease management intervention is most likely to be effective for this patient?

- **Collaborative filtering.** Collaborative filtering includes several different methods for predicting a user's rating for a specific item given the user's history of ratings for other items, combined with the history of all users' ratings for all items. Intuitively, if a user rates an item highly, then that user is likely to give a high rating to a very similar item. For healthcare, users could be patients, items could be treatments or interventions, and ratings could be outcome or level of compliance.
- **Memory-based learning.** Memory-based learning systems, such as Saffron, compute the difference between a new data point and previously seen data, for a number of different contexts. When the new data is near previous data, it is possible to predict the outcome based on what happened in the past. These systems tend to learn continuously on an ongoing basis as they are exposed to more data, and they can be used to reason on very complex data.
- **Association rules.** Association rules are a data-mining method in which algorithms use historical data to identify items or events that commonly occur together. For example, a patient needing health education and weight management is also highly likely to need nutrition management.

Common uses in healthcare:

- Matching patients with interventions and coaching resources
- Detecting fraudulent claims
- Call-center optimization and customer experience

7.9.9 Clustering

Can the data be grouped into natural categories or buckets?
Example: Are there natural groupings that can help me understand my patients?

- **Unsupervised Clustering.*** Unsupervised clustering algorithms, such as K-Means, can automatically identify naturally occurring groups of similar items. Typically, the algorithm is given a set of attributes for each item (for example, diagnoses and lab measurements for each patient) and a number of clusters to create. The algorithm will then work out which combinations of attributes most accurately divide the items into that number of groups. The resulting "clusters" can usually be interpreted by humans by looking at which attributes are most important in the cluster.
- **Hierarchical clustering.*** This family of methods creates a tree or dendogram of clustering scenarios for data, creating a single cluster containing all the items, which then splits into two clusters, each of which further splits into two clusters, and so on until each "cluster" contains only a single item. Based on the problem under consideration, this process can be stopped at any point to create meaningful clusters.

Common uses in healthcare:

- **Risk adjustment.** Risk adjustment is a crucial analytical tool for many applications in healthcare, from clinical-outcomes studies to determination of reimbursement for patients in capitated care delivery programs (such as Medicare Advantage). The problem is that, when calculating the impacts of different activities on outcomes, the baseline level of illness for each patient needs to be computed to create a consistent baseline for comparison of methods across all patients. Clustering can be used to group patients into meaningful groups of similar comorbidities or risks.
- **Patients like mine.** In the care of complex or rare disease, it is valuable to understand how different treatments have worked on other patients in similar situations, but this information is only useful if the past patients are similar enough to the current patient to have predictive value. Clustering can be very useful to identify the most similar patients.
- **Population health management and chronic disease management.** Current population health management methods rely on identifying broad groups of patients for whom interventions can be implemented to improve outcomes in general. Clustering is very powerful for finding these groups of patients.

7.9.10 Text Understanding

What does this text mean?
Example: Is this a properly documented diagnosis of diabetes with peripheral neuropathy?

- **Natural Language Processing (NLP).** Natural language processing includes an extensive toolkit of different text processing tools and steps, combined with the goal of understanding the meaning or practical implications of a piece of text. In clinical text analysis,

text understanding depends on being able to recognize distinctions among diagnoses attributed to a patient, those mentioned in a differential diagnosis, family history, and negations ("patient does not have diabetes mellitus"). There are a number of open-source tools for general NLP and for clinical NLP that can be incorporated into application development. Full NLP analysis of text can be very computationally expensive, both in terms of CPU cycles and memory required.

- **Text mining.** Text mining is the application of extensive dictionaries of terms to identify occurrences of key terms in text such as clinical notes, consult letters, and discharge summaries. Text mining has the advantage of being able to recognize vast variations in terminology, including abbreviations, misspellings, regional variations in usage, and transcription errors from scanned documents (often using optical character recognition, or OCR). Text-mining methods are often augmented with specific NLP tools to help understand the context of the terms that are identified in the text. For example, negation detection can be combined with search for diagnoses to help understand the difference between a positive statement of a diagnosis and a negative statement that a diagnosis does not apply.
- **Deep learning.** As described above, deep learning has the ability to learn key features of data without explicit programming. In the case of text understanding, deep learning has begun to show value for identifying complex ideas in text and interpreting the implications of their context.

Common uses in healthcare: Studies show that structured data in healthcare can be deeply flawed. For example, structured problem lists consistently suffer from extensive false negatives and false positives, even for impactful conditions such as heart failure. As a result, clinical text is one of the most reliable sources of usable information in healthcare, and the number of healthcare systems using text analytics as part of their reporting, decision support, and care optimization efforts is growing rapidly. Examples of applications include:

- Identify conditions that have been addressed in face-to-face encounters but not submitted to Medicare Advantage for capitated payment.
- Automated coding of typed or dictated encounter notes.
- Extraction of key findings in radiology reports to correlate with information in the EHR.
- Identification of medications and other key findings for inclusion in risk-prediction models.
- Automated chart review to identify care, such as annual diabetic foot exam, which is routinely performed without being separately coded. These reviews can directly impact reported performance measures.
- Mapping of patient care history, across multiple provider organizations, without requiring access to data sets from all providers.

Figure 7.4 shows how machine can be combined with other types of analytics to solve a large swath of business problems.

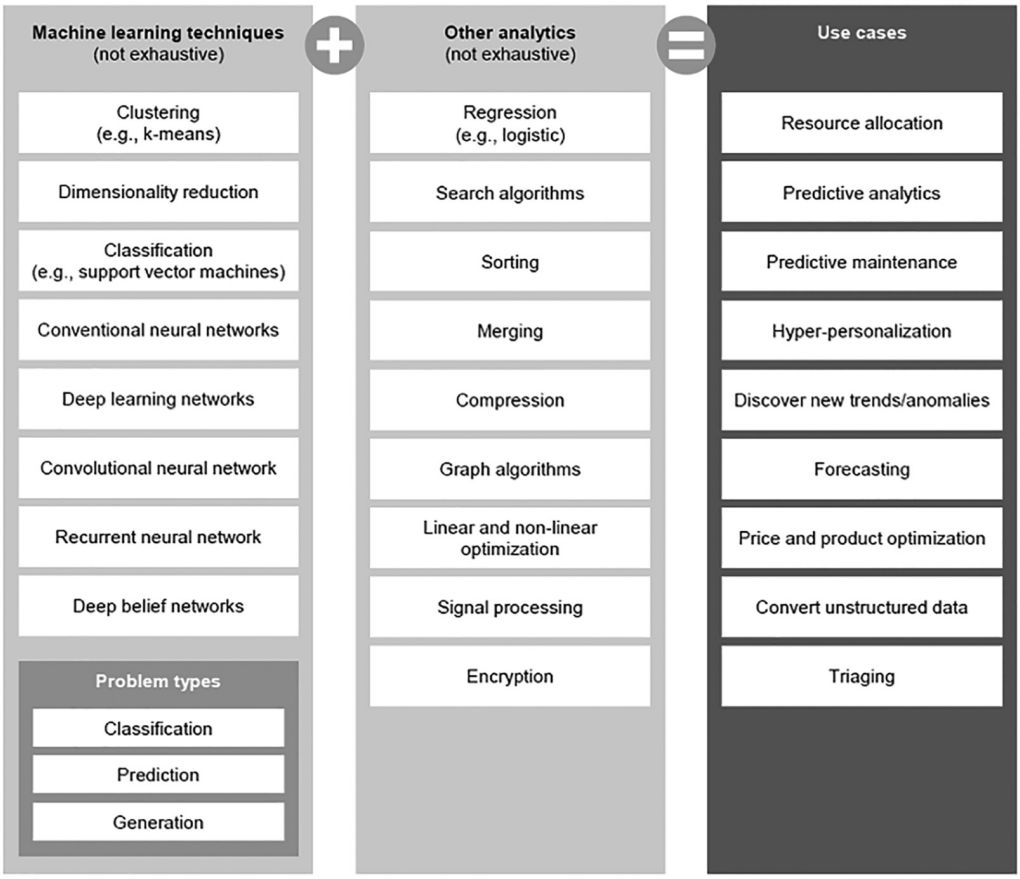


Figure 7.4 Machine learning can be combined with other types of analytics to solve a large swath of business problems. (*Source:* Exhibit from "The Age of Analytics: Competing in a Data-Driven World," December 2016, McKinsey Global Institute, www.mckinsey.com. Copyright © 2016 McKinsey&Company. All rights reserved. Reprinted by permission.)

7.10 Machine Learning and the Contextually Intelligent Agent

Machine learning will begin to realize its full potential in healthcare when contextually intelligent agents (CIAs) are put into widespread use. A CIA is a system that can interact directly with a human, via spoken or written communication, and that can understand context to identify what's important in a given situation. CIAs are a first, and crucial, stop on the journey to artificial intelligence. (See Figure 7.5).

Why is context so important? As we just described, most machine-learning algorithms have been created to answer a specific question. How can I compare two patients? Is this a heart failure patient? When will this patient convert from pre-diabetes to full-blown diabetes? But for any given situation, there are many, many such questions that could be reasonably asked, and then answered with a machine-learning algorithm. Intelligence depends, in part, on the ability to know which question to ask at any given time, based on the context of the situation.

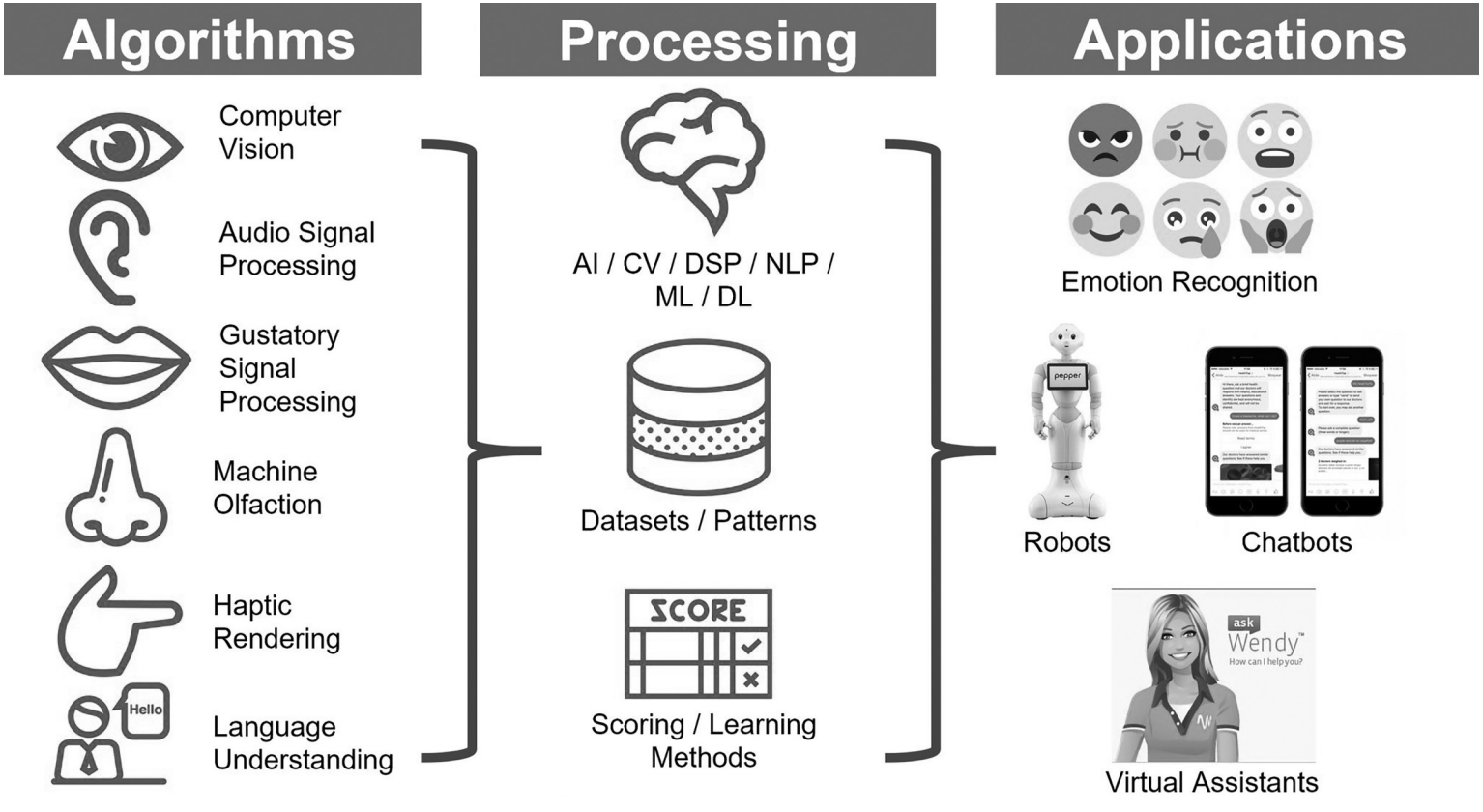


Figure 7.5 Learning algorithms, processing, and CIAs in healthcare. (Used with permission of Catalaize, Nardo Manaloto, and Wen Dombrowski MD.)

Who is asking the question? What has recently changed? Where is the question being asked? We refer to systems that are sufficiently contextually aware to be able to ask (and then answer) the right questions as CIAs.

CIAs are beginning to have an impact in healthcare. Chatbots are effectively communicating with patients to help them find an effective disease management intervention or to coach them through physical therapy following a surgical procedure. Or consider the portable ultrasound, a revolutionary technology that is being actively developed right now. To have non-invasive, highly actionable ultrasound information available in the field, where it can be most valuable, is a huge advance, but the challenge is that it has traditionally required an enormous amount of training for an ultrasound technician to be able to collect clinically meaningful images. How can a relatively untrained user in the field collect clinically useful ultrasound images? The answer is that the portable ultrasound comes with a CIA that can guide the user to collect good data given the circumstances that are driving the need for the imaging. Such contextually aware applications will become increasingly important in healthcare and will drive the value of machine-learning applications.

We would argue that, just as machine learning was a necessary technology to extract value from big data, the widespread availability of CIAs will be required for machine learning to realize its full value in healthcare. Not only do algorithms need to give the right answers, but they need to ask the right questions: questions that are sensitive to the current clinical situation, and which fit neatly into a clinical workflow rather than disrupting it. CIAs, which are appearing in many areas of healthcare, from payor disease management programs to patient wellness and diagnostic imaging in the field, are the key enabling technology to allow this transformation to happen.

7.11 Some Best Practices for Successful Machine Learning

7.11.1 Ask a Specific Question

Your machine-learning algorithm should answer a very specific question that tells you something you need to know and that can be answered appropriately by the data you have access to. The best first question is something you already know the answer to, so that you have a reference and some intuition to compare your results with. Remember: you are solving a business problem, not a math problem. Ask yourself, "What valuable action will be taken as the result of my analytics?"

Analytics and artificial intelligence systems come in two flavors: (1) knowledge management systems that interpret questions and provide information to answer these questions, and (2) very targeted quantitative systems designed to provide information for a specific use case. Don't try to build both types of system in a single effort.

7.11.2 Start Simple

This is true for model selection and the data you consider using for your analysis. You want your results to be robust, so less model complexity and fewer parameters are always beneficial. Regarding data, don't start by building a huge data lake with every kind of data you could possibly get your hands on. Instead, start with the minimal set of data that could get you to a good result.

7.11.3 Try Many Algorithms

Most machine-learning toolkits support multiple algorithms. Try a few to see how they work. This allows you to find the best tool for the job. Also, if one classifier works incredibly well and another doesn't seem to work well at all, be cautious. You may have an overfitting situation, which means you won't really have much predictive power. You may also want to combine methods: use deep learning to extract features from unstructured data and then use these features, along with others, in a classical machine-learning algorithm to get interesting results.

Remember that data is more important than the exact algorithm you use. More training data is always desirable. In addition, for classical machine-learning applications, the better your features, the better your performance will be.

7.11.4 Treat Your Data with Suspicion

Look at your data, dig into its details, look for correlations, suspicious gaps, systematic biases, errors, and flaws. Use statistics and visualizations here. Text has transcription errors, misspellings, and abbreviations. These challenges often exist for structured data as well: you will find that data is recorded inconsistently both across your data set and even within a single field.

7.11.5 Normalize Your Inputs

Machine-learning algorithms can perform poorly if there are large differences in scale between different features.

7.11.6 Validate Your Model

Separate your data into training, test, and validation sets, or if you are using K-fold cross validation, at least hold out a validation set. You need to keep some powder dry for most applications. Also, be aware of biases in your split. Remember: there is no such thing as a random set of data, only a random process to generate data. If you randomly flip six coins and they all come up heads, that's not going to be a very good validation set.

7.11.7 Focus on Data Fidelity But Ensure Quality of Training Data

Data fidelity is more appropriate for machine-learning systems than data quality for the reasons discussed in Section 2.4.2. For supervised learning algorithms, you will want to look closely at your training data. Does it cover all the use cases? Is it biased in some way? For example, did multiple humans create it? Can you see biases or differences among different folks?

This is particularly challenging in healthcare, where unstructured data is critical and source data comes from multiple silos. Extra effort in developing a high-quality training set will pay major dividends and will improve fidelity. Because of the variation in how information is represented in different healthcare settings, the more diverse the sources of data you use in your training set, the more transferable your results will be.

7.11.8 Set Up a Feedback Loop

Think through how you will use the output errors of your machine-learning system to improve it. Downstream users can provide feedback on when your algorithm got it wrong. How are you capturing this feedback so you can bring it back into training?

Note: This is great for false positives, but can miss false negatives, so you will want to pay special attention to false negatives as you train and use this experience to help you find missed results in production data review so you can include them in your next round of training.

7.11.9 Healthcare Doesn't Trust Black Boxes

Some machine-learning methods are more transparent than others. Clustering, topic modeling, and recommender systems tend to be easy for humans to interpret, because they create groupings of concepts that humans can associate with known influences. Linear regression can tell you how important each feature is to the final output. This is true to a lesser extent with linear classifiers. Random forests are difficult to interpret. Deep learning is truly a black box, with very little transparency to what is important in the decision-making process.

Note that there is a lot of research in this area of machine learning, so better tools for helping us understand the decision process are coming. In fact, some third-party healthcare analytics providers have instrumented interesting explanatory tools for their machine-learning algorithms.

7.11.10 Correlation Is Not Causation

It's easy convince yourself that two factors that move together imply that one causes the other. Just remember that in many cases there is a hidden factor that could be causing both factors to move together.

7.11.11 Monitor Ongoing Performance

How will you monitor the performance of your algorithm on an ongoing basis? Data drifts and systems evolve. You can do this manually by spot checking your results against the incoming data, and you can monitor data and algorithm statistics with a dashboard. Simple moving averages can tell you a lot.

7.11.12 Keep Track Of Your Model Changes

Always track the revision of your model and report it with your results. As you improve different parts of your data analytics pipeline, you will want to go back and re-analyze data. Recording which model was used at which time helps you understand what to recalculate.

7.11.13 Don't be Fooled by "Accuracy"

If you're looking for a rare event that only happens 1% of the time, and you never actually find it, you can report your accuracy as 99%. Obviously, that's meaningless. Instead, figure out before you start your project what precision and recall your application requires to be useful. Build your application to these metrics.

7.12 Conclusion

In conclusion, let us look at some next steps for machine learning and AI in healthcare—and what it means to professional practice, personal skill sets/knowledge, and jobs.

1. The current state of the industry, as evidenced by rapid recent progress and increasing investments, is promising. However, we haven't reached the Promised Land yet. We will get closer to it when the predictions and data-driven insights from machine learning are connected to contextually intelligent agents (CIAs), as described in the preceding section. CIAs and applications (see Figure 7.5) are necessary to complete the "data to action" or "data to behavior modification" workflow loops. They will be critical to the "mainstreaming" of machine learning within each HC organization and by individuals—patients, providers, and consumers included.
2. As you can expect, machine learning and AI in healthcare must deal with or account for the foundational five V's of big data in healthcare, as described in Chapter 2. Successful advanced analytics (predictive or prescriptive analytics) and CIAs will "need to be able to easily integrate more data sources, harness machine learning and advanced technology for faster, more sophisticated analyses, and extract insights that will improve business performance."[49]
3. Labeled data in healthcare remains a challenge and a time-consuming effort. While making more data available to machine-learning systems is always helpful, in healthcare setting that data must be prepared and labeled in the context of the source and intended use. Context-based labeling is essential to ensure the relevance and validity of learning, prediction, and action.
4. Feedback loops are essential to any healthcare machine-learning system. Period. Feedback loops can be of two types:
   - Explicit annotation by human experts
   - Implicit inferencing from downstream use of the data as evidenced by the behavior of both
     - contextually intelligent agents, and
     - individuals who are the recipients of the recommendations/actions suggested by the agent/application

[49] "How Analytics and Machine Learning Help Organizations Reap Competitive Advantage." (2016, December.) *MIT Technology Review.* Available at https://s3.amazonaws.com/files.technologyreview.com/whitepapers/Google-Analytics-Machine-Learning.pdf. p. 3.

5. The topic of data fidelity (as opposed to just data quality, as discussed in Sections 2.2.2 and 2.4.2 in Chapter 2) remains an important one in deep learning today. Some practitioners believe that deep-learning systems require the highest data quality as inputs for the predictions/results to be relevant and useful in healthcare. Other practitioners believe that imposing more stringent data quality requires more up-front investments and longer time to market. According to this school of thought, the advantages of deep-learning black boxes are proportionately reduced with increased attempts at enforcing transparency and management of uncertainty in hidden layers of a deep-learning system.

   The reason for this debate is that current deep-learning experiences don't provide needed transparency and do a sub-optimal job of determining and managing uncertainty in hidden layers or can't always identify blind spots between input and output layers. We agree that more transparent management of uncertainty and weights will allow us to do more deep learning on healthcare data of lower-data quality than is currently required.

   Given the current state of deep-learning architectures, we also think this debate is yet to be fully settled in healthcare. As a result, we don't yet have validated and settled best practices for this topic. However, given the coming importance of contextual usage-driven intelligent/applications for deep learning, we strongly recommend you use data fidelity (as defined in Section 2.2.2 of Chapter 2) instead of data quality in the design and deployment of deep-learning systems, both in the interim and for the future.

6. Questions for readers:
   - How are you approaching data fidelity in deep learning?
   - Do you think we're close to settling this debate or will that happen only when layers are "unhidden"?
   - What are the estimated added costs (or conversely, savings) of going from black-box to white-box deep learning or improving data fidelity?

7. In an Analytics 3.0 environment, the roles of data integration and MDM remain as critical as they do in the Analytics 1.0 and 2.0 environments. Analytics and workflows based on machine learning and CIAs must be integrated with all relevant data—both little and big—across silos in order to benefit the healthcare enterprise and users. Your EDW will continue to play a key role in supporting newer processes and technologies in an Analytics 3.0 world.

8. Questions on ethics and privacy as they relate to machine learning (and AI) are as relevant today (if not more so) as in the past. Current AI systems are not designed to account for the morality of learning algorithms and machine learning. It is useful to point out an important distinction between human and machine morality: "The moral constraints to which we are subject in our dealings with contemporary AI systems are all grounded in our responsibilities to other beings, such as our fellow humans, not in any duties to the systems themselves."[50]

[50] Bostrom, N. and Yudkowsky, E. "The Ethics of Artificial Intelligence." (n.d.). Machine Intelligence Research Institute. Available at https://intelligence.org/files/EthicsofAI.pdf. p. 7.

While it's beyond the scope of this chapter to provide answers to the various questions already being raised, we would like to suggest the following topics for more discussion and research.

a. How do we forward the knowledge coming out of machine learning and AI to the patient in a transparent and ethical way? How do we establish provenannce of insights being shared with the patient and provider?

b. Ethics of the human participant as related to providing inputs or acting on outputs.

c. Robo-ethics: ethics built into the machine-learning system by design or during learning.

d. Transparency around access to and inspection of the machine-learning system.

e. Reliability of predictions and validated performance.

f. Clear demarcation or sharing of human and machine-learning/CIA responsibilities when failure happens.

g. Legal, privacy, and innovation (patents, copyrights, etc.) considerations.

h. As you can guess, points a–g above will also create new discussions on the ethics of AI as related to public policy, reimbursements, population health management, SDoH, and access/cyber security at the global and individual levels.

9. Jobs: While applied machine learning and AI are here to stay/thrive, the impact on healthcare jobs will mostly be positive (helping humans do their tasks or creating more jobs) in the immediate- to mid-term. As a result, we will see increased augmentation of human tasks in healthcare—and not the overblown wholescale replacement of physicians, surgeons, radiologists, nurses, CXOs, or IT geeks as often portrayed in the popular press.

   However, we do not recommend complacency. In the mid- to long-term, we fully expect to see more administrative roles and even some specialties to be eclipsed or replaced by machine learning and AI (and in certain geographies and organizations at an accelerated pace). So, how do readers prepare for coming changes in healthcare jobs?

   a. Understand what's available and coming; we hope this book has helped you get started.

   b. Do projects to investigate and be prepared for emerging technologies in this space.

   c. Leverage your existing skills to drive question determination, feature definition, labeling, data integration, feedback loops, and promotion of data exchange and use.

   d. Leverage intelligent bots, agents, etc. and try things out via smartphones, personal wellness applications, CIAs.

27. S. G. Armato III, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, and A. P. Reeves, "Lung image database consortium: developing a resource for the medical imaging research community," *Radiology*, vol. 232, no. 3, pp. 739–748, 2004.
28. D. Testi, P. Quadrani, and M. Viceconti, "PhysiomeSpace: digital library service for biomedical data," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1921, pp. 2853–2861, 2010.
29. D. D. Ruikar, R. S. Hegadi, and K. C. Santosh, "Contrast stretching-based unwanted artifacts removal from CT images in recent trends in image processing and pattern recognition (accepted)," Springer, 2019.
30. K. C. Santosh, S. Candemir, S. Jäger, L. Folio, A. Karargyris, S. Antani, and G. Thoma, "Rotation detection in chest radiographs based on generalized line histogram of rib-orientations,"in *Computer-Based Medical Systems* (*CBMS*), *2014 IEEE 27th International Symposium on IEEE*, 2014, pp. 138–142.
31. R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (2nd ed.) Publishing House of Electronics Industry, Beijing, China, 2002, p. 455.
32. S. Vajda and K. C. Santosh, "A fast k-nearest neighbor classifier using unsupervised clustering," in *International Conference on Recent Trends in Image Processing and Pattern Recognition*, Springer, Singapore, 2016, pp. 185–193.
33. A. Karargyris, J. Siegelman, D. Tzortzis, S. Jaeger, S. Candemir, Z. Xue, K. C. Santosh, S. Vajda, S. Antani, L. Folio, and G. R. Thoma, "Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 1, 2016, pp. 99–106.
34. R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.
35. J. Y. Lai, T. Essomba, and P. Y. Lee, "Algorithm for segmentation and reduction of fractured bones in computer-aided preoperative surgery," in *Proceedings of the 3rd International Conference on Biomedical and Bioinformatics Engineering*, ACM, 2016, pp. 12–18.

Intelligent Light Therapy for Older Adults: Ambient Assisted Living

Joost van Hoof, Eveline J. M. Wouters, Björn Schräder, Harold T. G. Weffers, Mariëlle P. J. Aarts, Myriam B. C. Aries, and Adriana C. Westerlaken

21.1 Introduction

Light therapy is increasingly administered and studied as a nonpharmacologic treatment for a variety of health-related problems, including treatment of people with dementia. It is applied in a variety of ways, ranging from being exposed to daylight (in sanatoria) to being exposed to light emitted from electrical sources. These include light boxes, light showers, and ambient bright light. Light therapy covers an area in medicine where medical sciences meet the realms of physics, engineering, and technology (van Hoof et al. 2012).

One of the areas within medicine in which light therapy is administered is geriatric psychiatry, which includes the care of older adults with dementia. Dementia can be caused by a number of progressive disorders that affect memory, thinking, behavior, and the ability to perform everyday activities. Alzheimer's disease is the most common cause of dementia. Other types include vascular dementia, dementia with Lewy bodies, and frontotemporal dementia. Dementia mainly affects older people, although there is a growing awareness of a substantial amount of cases that start before the age of 65. After age 65, the likelihood of developing dementia roughly doubles every 5 years.

In the 2009 World Alzheimer Report, Alzheimer's Disease International estimated that there would be 35.6 million people living with dementia worldwide in 2010, increasing to 65.7 million by 2030 and 115.4 million by 2050. Within Europe, nearly two-thirds live in low- and middle-income countries, where the sharpest increases in numbers will occur.

The societal cost of dementia is already enormous. Dementia is significantly affecting every health and social care system in the world. The economic impact on families is insufficiently

appreciated. New technological services and government policies may help to address this problem.

Low-income countries accounted for just under 1% of the total worldwide costs (but 14% of the prevalence), middle-income countries for 10% of the costs (but 40% of the prevalence), and high-income countries for 89% of the costs (but 46% of the prevalence). About 70% of the global costs occurred in just two regions: Western Europe and North America.

Recent research indicates that dementia could be slowed down significantly by treatments that reset the body's biological clock. This kind of research started with the work of van Someren et al. (1997), who conducted a study on the effects of ambient bright light emitted from ceiling-mounted luminaires. In a randomized controlled study by Riemersma-van der Lek et al. (2008), brighter daytime lighting was applied to improve the sleep of persons with dementia (PwDs) and slow down cognitive decline. Applying bright light techniques is expected to lengthen the period PwDs can continue to live in their own home. It can also reduce the speed of health decline once the PwD has been moved to a care facility. In these situations, the quality of life will be positively impacted, and the workload of the care professionals and assisting relatives is expected to decrease. At the same time, Forbes et al. (2009) concluded that due to the lack of randomized controlled trials, there are no clear beneficial outcomes of light therapy for older persons. Similar lighting systems have also been tested in school environments; however, these systems did not have any effect on school children in a controlled setting (Sleegers et al. 2013).

PwDs do not venture outdoors as much as healthy younger adults, due to mobility impairments, and inside their homes, they are exposed to light levels that are not sufficient for proper vision, let alone yielding positive outcomes to circadian rhythmicity and mood (Aarts and Westerlaken 2005; Sinoo et al. 2011). Using light as a care instrument does not only apply to people with dementia, it also applies to ageing in general. Ageing impacts the circadian rhythm of people and increases the gap between level of sleep required and level of sleep achieved (older people do not get enough deep sleep). Moreover, it impacts the vision, since the eyes become affected due to biological ageing. These ageing effects include, among others, the yellowing of the lens and vitreous.

Although the evidence regarding the positive impact of light on well-being, especially of older people and PwDs or persons with other neurological diseases, is hopeful but not convincingly and scientifically affirmed (Forbes et al. 2009), these insights are already being converted to implementable solutions. Applying light as an instrument for care has tremendous benefits. It is noninvasive, it is cheap in implementation and maintenance, and it has a high level of intuitive use, creating a low threshold for acceptance.

Applying bright light techniques can slow down cognitive decline and reduce the speed of health decline. There is, however, no conclusive evidence on which lighting conditions are most favorable for yielding positive health outcomes. We also lack a clear definition of what technicians and product developers call healthy lighting. It is also not known how to design such healthy lighting systems, for instance, in relation to the emergence of new energy-friendly light sources such as LED (van Hoof et al. 2012). There is no validated set of algorithms as used in current lighting systems, including the effects of static versus dynamic lighting protocols (Barroso and den Brinker 2013), the contribution of the dynamic component of daylight, vertical and horizontal illuminance levels, and color temperature. The available knowledge has not yet been converted into widespread implementable lighting solutions, and the solutions available are often technologically unsophisticated, uneducated guesses and poorly evaluated from the perspective of end users. New validated approaches in terms of ceiling-mounted luminaires, the inclusion of low-energy light sources, and integration of computerized controls are needed.

This chapter will focus first on the effects of biological ageing and dementia on our lighting needs and second on the application of intelligent light therapies for older adults with dementia.

21.2 The Effects of Biological Ageing and Dementia

The age-related sensory changes, involving sensory receptors in the eyes, ears, nose, buccal cavity, and peripheral afferent nerves, frequently affect the way we perceive the environment. Apart from the sensory changes, incorrect or malfunctioning visual aids and hearing aids may have negative effects, too. Sensory losses or impairments, together with cognitive deficits, make it difficult for the individual to interpret and understand the environment (perception and comprehension phase) (van Hoof et al. 2010).

21.2.1 Ageing-Related Changes in Vision

Ageing negatively affects vision. In general, the performance of the human eye deteriorates already at a relatively early age. Many people aged 45 and over wear glasses to compensate for impaired vision due to presbyopia, caused by reduced elasticity of the lens of the eye resulting in significant loss of focusing power. Older people are known to have vision impairments stemming from the normal ageing process, which include an impaired ability to quickly adapt to changes in light levels, extreme sensitivity to glare, reduced visual acuity, restricted field of vision and depth perception, reduced contrast sensitivity, and restricted color recognition. Changes in vision do not happen overnight and depend on the progress of age. After the age of 50, glare and low levels of light become increasingly problematic. People require more contrast for proper vision and have difficulty perceiving patterns. After the age of 70, fine details become even harder to see, and color and depth perception may be affected. Apart from the influence of ageing, there are pathological changes leading to low vision and eventual blindness, such as cataracts, macular degeneration, glaucoma, and diabetic retinopathy (van Hoof et al. 2010; Sinoo et al. 2011). In Table 21.1 an overview of age-related eye pathology is given.

21.2.2 Ageing and Nonvisual Effects of Light

Apart from being indispensable for proper vision, light plays a role in regulating important biochemical processes, immunologic mechanisms, and neuroendocrine control (for instance, melatonin and cortisol pathways), via the skin and via the eye (Hughes and Neer 1981). Light exposure ($\lambda \sim 460$–480 nm) is the most important stimulus for synchronizing the biological clock, suppressing pineal melatonin production, elevating core body temperature, and enhancing alertness (van Hoof et al. 2010, 2012). The circadian system, which is orchestrated by the hypothalamic suprachiasmatic nuclei (SCN), influences virtually all tissues in the human body. In the eye, light activates intrinsically photosensitive retinal ganglion cells (Brainard et al. 2001; Thapan et al. 2001), which discharge nerve impulses that are transmitted directly to the SCN and, together with the photoreceptors for scotopic and photopic vision, participate in mammalian circadian phototransduction.

In older adults, the orchestration by the SCN requires ocular light levels that are significantly higher than those required for proper vision, but the exact thresholds are unknown

TABLE 21.1

Age-Related Sensory Changes to Vision

- Lid elasticity diminished, leading to pouches under the eyes.
- Loss of orbital fat, leading to excessive dryness of eyes.
- (1) Decreased tears; (2) arcus senilis becomes visible; (3) sclera yellows and becomes less elastic; (4) yellowing and increased opacity of cornea, which may lead to a lack of corneal luster.
- (1) Increased sclerosis and rigidity of the iris and (2) a decrease in elasticity and convergence ability of the lens, leading to presbyopia.
- Decline in light accommodation response leads to lessened acuity.
- Diminished pupillary size leads to a decline in depth perception.
- Atrophy of the ciliary muscles (holding the lens) leads to a diminished recovery from glare.
- Night vision diminishes leading to night blindness.
- Yellowing of the lens may lead to a diminished color perception (blues and greens).
- Lens opacity may develop, leading to cataract.
- Increased ocular pressure may lead to seeing rainbows around lights.
- Shrinkage of gelatinous substance in the vitreous, which may lead to altered peripheral vision.
- Vitreous floaters appear.
- Ability to gaze upward decreases.
- Thinning and sclerosis of retinal blood vessels.
- Atrophy of photoreceptor cells.
- Degeneration of neurons in visual cortex.

Sources: Hughes, P.C., and Neer, R.M., *Hum. Factors*, 23(1), 65–85, 1981. Ebersole, P. et al., editors, *Toward Healthy Aging*, sixth edition, Mosby, St. Louis, MO, 2004.

to date. Research by Aarts and Westerlaken (2005) in the Netherlands has shown that light levels, even during daytime, are too low both to allow for proper vision and, consequently, also for non–image-forming effects, even though the semi-independently living older persons were satisfied with their lighting conditions. A similar study was carried out among 40 community-dwelling older people in New York City by Bakker et al. (2004). Even though nearly all of them had inadequate light levels for both image-forming and non-image-forming effects, subjects rated their lighting conditions as adequate.

An additional problem is formed by the ageing of the eye, which leads to opacification and yellowing of the vitreous and the lens, limiting the amount of bluish light reaching the retinal ganglion cells. This can be as much as a 50% reduction in 60-year olds compared to 20-year olds.

Many older adults are not exposed to high-enough illuminance levels, due to decreased lens transmittance, poorly lit homes (up to 400 lx), and the short periods of time spent outdoors. The indoor illuminance levels are too low for any non-image-forming effects to take place.

21.2.3 Dementia-Related Changes in Vision

Dementia has a severe impact on the human visual system (Guo et al. 2010), and the effects of biological ageing often aggravate the visual dysfunctions stemming from dementia. Persons with Alzheimer's disease frequently show a number of visual dysfunctions, even in the early stages of the disease (Kergoat et al. 2001; Redel et al. 2012). These dysfunctions include impaired spatial contrast sensitivity, motion discrimination, and color vision, as well as blurred vision. Altered visual function may even be present if people with dementia have normal visual acuity and have no ocular diseases (Kergoat et al. 2001). Another dysfunction is diminished contrast sensitivity, which may exacerbate the effects of other

cognitive losses and increase confusion and social isolation (Boyce 2003). Impaired visual acuity may be associated with visual hallucinations (Desai and Grossberg 2001). According to Mendez et al. (1996), persons with Alzheimer's disease have disturbed interpretation of monocular as well as binocular depth cues, which contributes to visuospatial deficits. The impairment is largely attributed to disturbances in local stereopsis and in the interpretation of depth from perspective, independent of other visuospatial functions.

21.2.4 Dementia and Nonvisual Effects of Light

In people with Alzheimer's disease, the SCN is affected by the general atrophy of the brain, leading to nocturnal restlessness due to a disturbed sleep–wake rhythm and wandering (van Someren 2000; Waterhouse et al. 2002). The timing of the sleep–wake cycle can show a far wider variation; times of sleep and activity can vary substantially from day to day or can be temporarily inverted (Waterhouse et al. 2002), which has great implications for both the PwD and his/her family carer. Restlessness and wandering form a high burden for carers and are among the main reasons for institutionalization (Health Council of the Netherlands 2002; Abbott 2003; Harper et al. 2005). Marshall (1995) stated that lighting technology deserves more attention as a means to help with managing problem behavior. Hopkins et al. (1992) have suggested a relation between illuminance levels and this type of behavior before, and today, light therapy is used as a treatment to improve sleep in people experiencing sundowning behavior (Brawley 2006). Sundowning is associated with increased confusion and restlessness in PwDs in the evening.

It is hypothesized that high-intensity lighting, with vertical (instead of horizontal, as our eye is located in a vertical plain) illuminance levels of well over 1000 lx (eye height), may play a role in the management of dementia. Bright light treatment with the use of light boxes is applied to entrain the biological clock, to modify behavioral symptoms, and to improve cognitive functions, by exposing people with dementia to high levels of ocular light (see, for instance, Lovell et al. 1995; Thorpe et al. 2000; Yamadera et al. 2000; Graf et al. 2001; Dowling et al. 2005). This intervention requires supervision to make PwDs follow the total protocol and may cause a bias in the outcomes of the therapy, for instance, as the level of personal attention is higher. The results of bright light therapy on managing sleep and behavioral, mood, and cognitive disturbances show preliminary positive signs, but there is a lack of adequate evidence obtained via randomized controlled trials to allow for widespread implementation in the field (Kim et al. 2003; Terman 2007; Forbes et al. 2009).

Another approach that is gaining popularity, from a research, ethical, and practical point of view, is to increase the general illuminance level in rooms where people with dementia spend their days in order for non-image-forming effects of light to take place (Boyce 2003). Studies by Rheaume et al. (1998), van Someren et al. (1997), Riemersma-van der Lek et al. (2008), and van Hoof et al. (2009a,b) that exposed institutionalized PwDs to ambient bright light through ceiling-mounted luminaires showed short-term and long-term effects, such as lessened nocturnal restlessness, a more stable sleep–wake cycle, possible improvement to restless and agitated behavior as well as better sleep quality, increased amplitude of the circadian body temperature cycle, and a lessening of cognitive decline.

The occurrence of nonvisual effects of light does not only depend on light intensity. As stated before, certain parts of the light spectrum (specific short wavelengths) are more effective than others. The human circadian photoreception sensitivity peaks at approximately 480 nm, which is associated with the neuroendocrine and neurobiological systems. This sensitivity is graphically represented in the so-called C(λ) curve, which is used

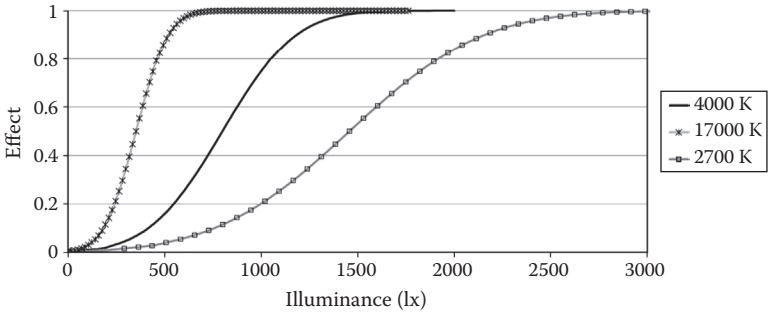


FIGURE 21.1
Hypothesized size of nonvisual effects of light during daytime for different illuminance levels at the eye and different color temperatures. (Adapted from Górnicka, G.B., Lighting at Work. Environmental Study of Direct Effects of Lighting Level and Spectrum on Psychophysiological Variables. Dissertation, Eindhoven University of Technology, Eindhoven, 2008. Reprinted from *Building and Environment*, 44(9), van Hoof, J. et al., High colour temperature lighting for institutionalised older people with dementia, 1959–1969, Copyright 2009, with permission from Elsevier. License number: 3087171431049.)

in lighting research and practice (Pechacek et al. 2008). Generally, required illuminance levels are higher than average, and so is the (correlated) color temperature of the light (Górnicka 2008) (Figure 21.1). The (correlated) color temperature is one of the measures for the amount of short-wavelength light present in the spectrum. As stated, there are several short-term and long-term effects (van Hoof et al. 2012). A cluster-unit crossover intervention trial by Sloane et al. (2007) on the effects of high-intensity light found that nighttime sleep of older adults with dementia improved when exposed to morning and all-day light, with the increase most prominent in participants with severe or very severe dementia. Hickman et al. (2007) studied the effects on depressive symptoms in the same setting as Sloane et al. (2007). Their findings did not support the use of ambient bright light therapy as a treatment for depressive symptoms. To date, it is unknown if the light therapy is effective, how long effects of bright light last, and how to predict which persons (may) respond favorably to light treatment. These points were already made by the Health Council of the Netherlands in 2002.

21.3 Technological Solutions: Design and Practice

The increasing numbers of older people with dementia in combination with the lack of available care professionals go together with a need for technological solutions and services to support activities of daily living and reduce the burden on carers. Light therapy is hypothesized to play a role in improving the well-being and quality of life of PwDs. To date, the administration of light therapy via ceiling-mounted luminaires is a relatively new area of study and innovation. As the current state of science permits us to design and model healthy lighting solutions, it is time to improve the quality of life of PwDs. In order to do so, we need to investigate the recent innovations in the field of lighting technology in relation to the underlying algorithms of the lighting equipment's steering mechanism.

New dynamic lighting protocols are being implemented in the lighting solutions offered to older adults (Figures 21.2 and 21.3). The underlying assumption of such systems is that

FIGURE 21.2
(See color insert.) Examples of dynamic lighting installed in Dutch nursing homes: Amadea by Derungs (left), Biosun by Van Doorn (middle), and Strato by Philips (right).

FIGURE 21.3
Examples of dynamic lighting systems in Dutch nursing homes. Figure upper right. (Reprinted from *Building and Environment*, 44(9), van Hoof, J. et al., High colour temperature lighting for institutionalised older people with dementia, 1959–1969, Copyright 2009, with permission from Elsevier. License number: 3087171431049.)

human beings evolved in daylight conditions and that the dynamic component further contributes to the positive effects of the lighting systems. As there is no validated set of algorithms as used in current lighting systems, including the relative effects of static versus dynamic lighting protocols (Rea et al. 2002; Figueiro 2008; Barroso and den Brinker 2012), there is still plenty of room for innovation and research. Figure 21.4 shows the rationale behind a dynamic lighting protocol and the way it has been shaped in practice. As can be seen in the figure, both illuminance and color temperature are controlled through dedicated software. In most projects, only the main luminaire in the living room is steered via a dynamic protocol.

In all the studies concerning light therapy, the exposure to daylight is often an ill-described aspect. We therefore do not fully understand the effects of these interactions. Daylight has a dynamic character, which is mimicked by dynamic lighting systems. With new technologies, lighting can be supplemented to the available daylight, which also has positive effects on energy consumption. This requires that new lighting solutions are to

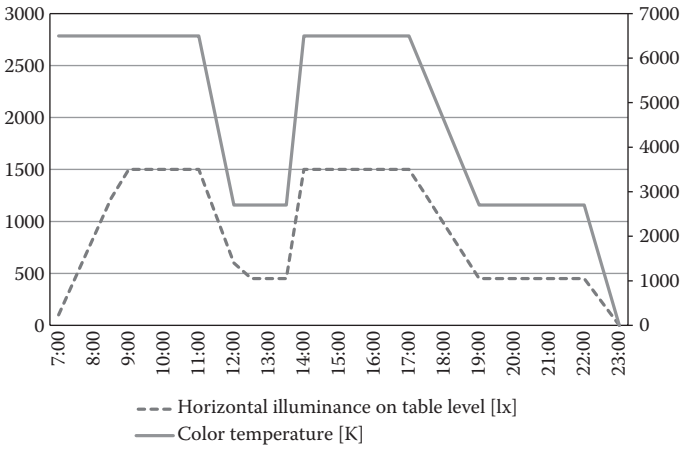


FIGURE 21.4
Example of a protocol for dynamic lighting. The horizontal illuminance on the table level varies between 0 and 1500 lx. The color temperature of the light varies between 0 and 6500 K. The so-called post-lunch dip can be found in the first hour of noon.

be integrated within a framework of the digitalization of the home (including sensor-based networks, ambient assisted living, and e-health) (van Hoof et al. 2011a,b) and smart façade systems (measuring incoming daylight and its spectral composition), which can go together with an optimization of exposure to certain light levels and, hence, energy use. Moreover, the exposure to light can be part of a digital patient or care file and form an additional source of information to medication protocols. Ambient intelligence in the home environment can thus lead to new improvements in the administration of ambient bright light, starting with institutional settings.

New lighting solutions should combine all sources of light (including daylight); novel lighting technologies such as LED technology (with energy-efficient options for a high output in the blue region of the spectrum) or ceiling- or wall-mounted lighting applications (also referred to as ambient bright light); smart sensors that measure the quality and quantity of light needed at a specific moment for a specific individual; and the platform that both controls the supply of light and, in parallel, measures the effectiveness of the light. The intricate balance between daylight and electrical light sources calls for smart sensors in the dwelling and dedicated software to steer the lighting exposure of persons. These sensors should also measure the presence of occupants and the type of activity they are engaged in. The dip in lighting protocols (Figure 21.4) is related to the length of the lunch break. If this break takes longer, perhaps the dip in the protocol should be longer too. The protocol should ideally follow the care regime. Apart from manual controls, a sensor-based network can help achieve accounting for the regime.

In addition, light therapy also calls for an attractive design of the lighting equipment. Indoor lighting conditions may be perceived as unfavorable from the perspective of personal preferences and taste, particularly with higher color temperature lighting or light with a dedicated, spectrum which accounts for the human circadian photoreception sensitivity that peaks at approximately 480 nm [following from the C(λ) curve (Pechacek et al. 2008)]. This can be solved by adding an additional decorative luminaire underneath the luminaires used for light therapy.

One other factor that makes it hard to implement innovations in the field is that to date, there are no extensive documents on lighting for older people, which can serve as an underlying basis when conducting research and for product and service innovation. There is, however, a growing interest within the building community in the nonvisual aspects of light (Webb 2006; McNair et al. 2010). CEN EN 12464-1 (2011) summarizes recommendations of lighting of indoor work places. The standard specifies horizontal illuminances for health care facilities, such as waiting rooms, corridors, examination rooms, and spaces for diagnostics in hospitals. Nursing homes are not included in this standard. The standard does not specifically include color temperature for health care facilities in general, too. There need to be more efforts to improve current standards and guidelines to account for light therapy and ambient bright light. We need to look at how science can find its way into practice in order to improve the quality of life of PwDs and the work of their carers and the installers.

At the same time, we need to be critical. There is still a lot we do not know. To date, we do not know if dynamic systems have better outcomes than static lighting systems and how such systems (should) interact with available daylight. We do know that vision can be improved by raising general illuminance levels and glare control; the nonvisual benefits cannot be quantified yet. Therefore, the economic and well-being benefits of accounting for these parameters are not yet clear. As long as there are many uncertainties, maybe we should suggest exposing our older citizens to plenty of daylight, for instance, by taking them out for a stroll. The innovations in the realm of sensors and ambient intelligence in the home environment hold a promise that, in the future, the nursing home can administer the right amount of lighting to its residents, which allows them to enjoy the highest degree of quality of life.

References

Aarts MPJ, Westerlaken AC. Field study of visual and biological light conditions of independently-living elderly people. *Gerontechnology* 2005;(3):41–152.

Abbott A. Restless nights, listless days. *Nature* 2003;245(6961):896–898.

Alzheimer's Disease International (2009). World Alzheimer report. Alzheimer's Disease International, London.

Bakker R, Iofel Y, Mark S, Lachs MS. Lighting levels in the dwellings of homebound older adults. *Journal of Housing for the Elderly* 2004;18(2):17–27.

Barroso A, den Brinker B. Boosting circadian rhythms with lighting: a model driven approach. *Lighting Research and Technology* 2013;45(2):197–216 doi: 10.1177/1477153512453667.

Boyce PR (2003). *Human Factors in Lighting*. *Lighting Research and Technology,* 2013;45(2):197–216.

Brainard GC, Hanifin JR, Greeson JM, Byrne B, Glickman G, Gerner E, Rollag MD. Action spectrum for melatonin regulation in humans: Evidence for a novel circadian photoreceptor. *Journal of Neuroscience* 2001;21(16):6405–6412.

Brawley EC (2006). *Design Innovations for Aging and Alzheimer's. Creating Caring Environments*. John Wiley & Sons, Hoboken, NJ.

CEN. EN 12464-1 (2011). *Light and Lighting—Lighting of Work Places—Part 1: Indoor Work Places*. Comité Européen de Normalisation CEN, Brussels, Belgium.

Desai AK, Grossberg GT. Recognition and management of behavioral disturbances in dementia. *Primary Care Companion to the Journal of Clinical Psychiatry* 2001;3(3):93–109.

Dowling GA, Hubbard EM, Mastick J, Luxenberg JS, Burr RL, van Someren EJW. Effect of morning bright light treatment for rest-activity disruption in institutionalized patients with severe Alzheimer's disease. *International Psychogeriatrics* 2005;17(2):221–236.

Ebersole P, Hess P, Schmidt-Luggen A, editors (2004). *Toward Healthy Aging*. Sixth edition. Mosby, St. Louis, MO.

Figueiro MG. A proposed 24 h lighting scheme for older adults. *Lighting Research and Technology* 2008;40(2):153–160.

Forbes D, Culum I, Lischka AR, Morgan DG, Peacock S, Forbes J, Forbes S. Light therapy for managing cognitive, sleep, behavioural, or psychiatric disturbances in dementia. *Cochrane Database of Systematic Reviews* 2009;(4):CD003946.

Górnicka GB (2008). Lighting at work. Environmental study of direct effects of lighting level and spectrum on psychophysiological variables. Dissertation. Eindhoven: Eindhoven University of Technology.

Graf A, Wallner C, Schubert V, Willeit M, Wlk W, Fischer P, Kasper S, Neumeister A. The effects of light therapy on Mini-Mental State Examination in demented patients. *Biological Psychiatry* 2001;50(9):725–727.

Guo L, Duggan J, Cordeiro MF. Alzheimer's disease and retinal neurodegeneration. C*urrent Alzheimer Research* 2010;7(1):3–14.

Harper DG, Volicer L, Stopa EG, McKee AC, Nitta M, Satlin A. Disturbance of endogenous circadian rhythm in aging and Alzheimer disease. *The American Journal of Geriatric Psychiatry* 2005;13(5):359–368.

Health Council of The Netherlands (2002). Dementia. Publication no. 2002/04. Health Council of The Netherlands, The Hague, The Netherlands [in Dutch].

Hickman SE, Barrick AL, Williams CS, Zimmerman S, Connell BR, Preisser JS, Mitchell CM, Sloane PD. The effect of ambient bright light therapy on depressive symptoms in persons with dementia. *Journal of the American Geriatrics Society* 2007;55(11):1817–1824.

Hopkins RW, Rindlisbacher P, Grant NT. An investigation of the sundowning syndrome and ambient light. *The American Journal of Alzheime*r's *Care and Related Disorders and Research* 1992;7(2): 22–27.

Hughes PC, Neer RM. Lighting for the elderly: a psychobiological approach to lighting. *Human Factors* 1981;23(1):65–85.

Kergoat H, Kergoat M-J, Justino L, Robillard A, Bergman H, Chertkow H. Normal optic nerve head topography in the early stages of dementia of the Alzheimer type. *Dementia and Geriatric Cognitive Disorders* 2001;12(6):359–363.

Kim S, Song HH, Yoo SJ. The effect of bright light on sleep and behavior in dementia: an analytic review. *Geriatric Nursing* 2003;24(4):239–243.

Lovell BB, Ancoli-Israel S, Gevirtz R. Effect of bright light treatment on agitated behavior in institutionalized elderly subjects. *Psychiatric Research* 1995;57(1):7–12.

Marshall M. Technology is the shape of the future. *Journal of Dementia Care* 1995;3(3):12–14.

McNair D, Cunningham C, Pollock R, McGuire B (2010). *Light and Lighting Design for People With Dementia*. Stirling, Dementia Services Development Centre, University of Stirling, UK.

Mendez MF, Cherrier MM, Meadows RS. Depth perception in Alzheimer's disease. *Perceptual and Motor Skills* 1996;83(3 Pt 1):987–995.

Pechacek CS, Andersen M, Lockley SW. Preliminary method for prospective analysis of the circadian efficacy of (day)light with applications to healthcare architecture. *LEUKOS* 2008;5(1):1–26.

Rea MS, Figueiro MG, Bullough JD. Circadian photobiology: an emerging framework for lighting practice and research. *Lighting Research and Technology* 2002;34(3):177–187.

Redel P, Bublak P, Sorg C, Kurz A, Förstl H, Müller HJ, Schneider WX, Perneczky R, Finke K. Deficits of spatial and task-related attentional selection in mild cognitive impairment and Alzheimer's disease. *Neurobiology of Aging* 2012;33(1):195.e27–195.e42.

Rheaume YL, Manning BC, Harper DG, Volicer L. Effect of light therapy upon disturbed behaviors in Alzheimer patients. *American Journal of Alzheimer's Disease* 1998;13(6):291–295.

Riemersma-van der Lek RF, Swaab DF, Twisk J, Hol EM, Hoogendijk WJG, van Someren EJW. Effect of bright light and melatonin on cognitive and noncognitive function in elderly residents of group care facilities. A randomized controlled trial. *The Journal of the American Medical Association* 2008;299(22):2642–2655.

Sinoo MM, van Hoof J, Kort HSM. Lighting conditions for older adults in the nursing home: assessment of environmental illuminances and colour temperature. *Building and Environment* 2011;46(10):1917–1927.

Sleegers PJC, Moolenaar NM, Galetzka M, Pruyn A, Sarroukh BE, van der Zande B. Lighting affects students' concentration positively: findings from three Dutch studies. *Lighting Research and Technology* 2013;45(2):159–175, doi:10.1177/1477153512446099

Sloane PD, Williams CS, Mitchell CM, Preisser JS, Wood W, Barrick AL, Hickman SE, Gill KS, Connell BR, Edinger J, Zimmerman S. High-intensity environmental light in dementia: effect on sleep and activity. *Journal of the American Geriatrics Society* 2007;55(10):1524–1533.

Terman M. Evolving applications of light therapy. *Sleep Medicine Reviews* 2007;11(6):497–507.

Thapan K, Arendt J, Skene DJ. An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. *Journal of Physiology* 2001;535(1):261–267.

Thorpe L, Middleton J, Russell G, Stewart N. Bright light therapy for demented nursing home patients with behavioral disturbance. *American Journal of Alzheimer's Disease* 2000;15(1):18–26.

van Hoof J, Aarts MPJ, Rense CG, Schoutens AMC. Ambient bright light in dementia: effects on behavior and circadian rhythmicity. *Building and Environment* 2009a;44(1):146–155.

van Hoof J, Schoutens AMC, Aarts MPJ. High colour temperature lighting for institutionalised older people with dementia. *Building and Environment* 2009b;44(9):1959–1969.

van Hoof J, Kort HSM, Duijnstee MSH, Rutten PGS, Hensen JLM. The indoor environment and the integrated building design of homes for older people with dementia. *Building and Environment* 2010;45(5):1244–1261.

van Hoof J, Kort HSM, Rutten PGS, Duijnstee MSH. Ageing-in-place with the use of ambient intelligence technology: perspectives of older users. *International Journal of Medical Informatics* 2011a;80(5):310–331.

van Hoof J, Wouters EJM, Marston HR, Vanrumste B, Overdiep RA. Ambient assisted living and care in The Netherlands: the voice of the user. *International Journal of Ambient Computing and Intelligence* 2011b;3(4):25–40.

van Hoof J, Westerlaken AC, Aarts MPJ, Wouters EJM, Schoutens AMC, Sinoo MM, Aries MBC. Light therapy: methodological issues from an engineering perspective. *Health Care and Technology* 2012;20(1):11–23.

van Someren EJW. Circadian and sleep disturbances in the elderly. *Experimental Gerontology* 2000;35(9–10):1229–1237.

van Someren EJW, Kessler A, Mirmiran M, Swaab DF. Indirect bright light improves circadian rest-activity rhythm disturbances in demented patients. *Biological Psychiatry* 1997;41(9):955–963.

Waterhouse JM, Minors DS, Waterhouse ME, Reilly T, Atkinson G (2002). *Keeping in Time With Your Body Clock*. Oxford University Press, Oxford.

Webb AR. Considerations for lighting in the built environment: non-visual effects of light. *Energy and Buildings* 2006;38(7):721–727.

Yamadera H, Ito T, Suzuki H, Asayama K, Ito R, Endo S. Effects of bright light on cognitive and sleep–wake (circadian) rhythm disturbances in Alzheimer-type dementia. *Psychiatry and Clinical Neurosciences* 2000;54(3):352–353.

Artificial Intelligence Approaches for Drug Safety Surveillance and Analysis

Mei Liu, Yong Hu, Michael E. Matheny, Lian Duan, and Hua Xu

27.1 Introduction

From the earliest moments in the modern history of computers, scientists have conceptualized the potential of artificial intelligence (AI) in medicine with the hope to create AI systems that can assist clinicians in the complex medical diagnosis processes [1,2]. The earliest work in medical AI dates back to the early 1970s with applications primarily focused on constructing AI programs that perform diagnoses and make therapy recommendations [3]. Much has changed since then. A wide array of AI-inspired methods have been developed to solve a broad range of important clinical and biological problems such as computer-based knowledge generation, decision support systems, and clinical data mining [4]. This chapter focuses on one specific application of AI methodologies to medical research—medication safety surveillance.

Every year, the US public spends billions of dollars on prescription medications. However, an awareness of the potential side effects that can occur with the use of medications is important for patients, providers, health care systems, payers, and regulatory agencies. Some side effects are minor, but others can turn out to be severe adverse drug reactions (ADRs) leading to patient morbidity. ADRs are defined as "any unintended and undesirable effects of a drug beyond its anticipated therapeutic effects occurring during clinical use at normal dose" [5]. According to the national surveillance study conducted by Budnitz et al. [6] on outpatient emergency visits, a total of 21,298 adverse drug events occurred from January 1, 2004, through December 31, 2005, which would yield a weighted annual estimate of 701,547 individuals or 2.4 individuals per 1000 population treated in the emergency departments experiencing an adverse event. Moreover, Lazarou et al. [7] estimated that 6%–7% of hospitalized patients would experience severe ADRs, causing 100,000 deaths annually in the United States. Additionally, both reported ADRs and related deaths have increased ~2.6 times over the past decade, and numerous drugs withdrew from the US market after presenting unexpected severe ADRs [8,9]. Consequently, ADRs presents a huge financial burden on the national economy, with an estimated $136 billion annual cost in the United States [10,11].

Drug discovery and development is a long and expensive process (Figure 27.1). To bring a new drug from discovery to market, it can take 10–17 years and millions of dollars [12]. And yet the majority of drug discovery endeavors are not successful due to the high failure rate of drug candidates in clinical trials, and approximately 30% of the failures are linked to unacceptable toxicities [13]. Hence, predicting potential ADRs at early stages is essential to reduce the risks of costly failures. Furthermore, even after a drug is approved to market, undiscovered ADRs may occur and lead to withdrawals, which can be financially detrimental for the manufacturers. Therefore, it is critical to predict and monitor a drug's ADRs throughout its life cycle, from preclinical screening phases to postmarket surveillance. Pharmacovigilance (PhV) is the science to address this problem.

In this chapter, we will cover a broad spectrum of the current AI approaches for PhV at both premarketing and postmarketing stages. The methodologies are presented along

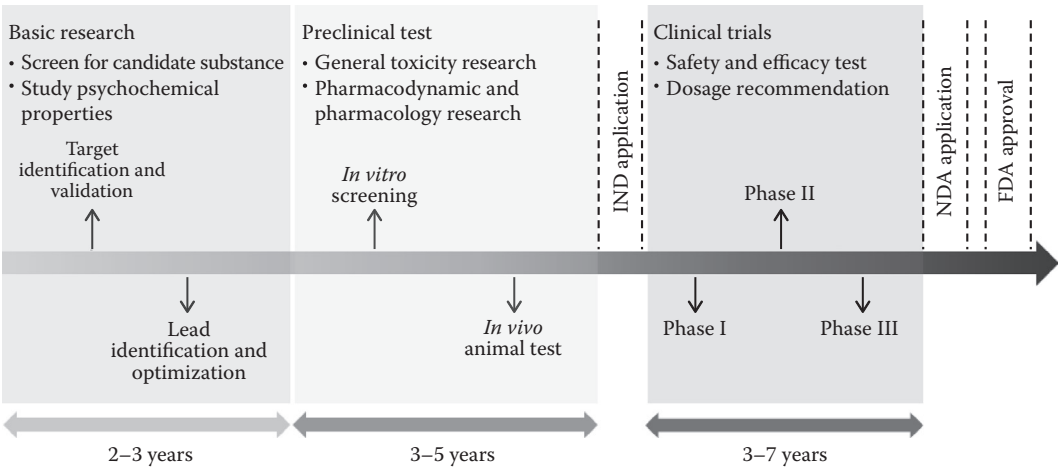


FIGURE 27.1
(See color insert.) Overview of the drug development process. Once a new chemical compound passes preclinical test, pharmaceutical company files an IND with the US Food and Drug Administration (FDA) to obtain approval for testing the new drug in humans. After the drug passes all phases of clinical trials, a pharmaceutical company formally submits a proposal to the FDA to approve the new drug for sale and marketing. IND, investigational new drug; NDA = new drug application.

different axes according to the data sources utilized with respect to each PhV stage. We will present a general overview of PhV in Section 27.2, followed by applications of AI methods in the premarketing surveillance stage (Section 27.3) and postmarketing surveillance stage (Section 27.4), and conclude with useful resources for future research needs (Section 27.5) and future perspectives and challenges (Section 27.6).

27.2 PhV—Drug Safety Surveillance: An Overview

PhV, also known as drug safety surveillance, is to enhance patient care and patient safety with regard to the use of medicines through the collection, monitoring, assessment, and evaluation of information from health care providers and patients. Generally speaking, PhV is conducted throughout the drug development process and market life and can be divided into two major stages: (1) premarketing surveillance—analyzing information collected from preclinical screening and phase I to III clinical trials; and (2) postmarketing surveillance—evaluating data accumulated in the postapproval stage and throughout a drug's market life (Figure 27.2).

Historical PhV efforts have relied on biological experiments or manual review of case reports; however, due to the vast quantities and complexity of the data to be analyzed, computational methods that can accurately and efficiently identify ADR signals have become a critical component in PhV. A variety of enabling resources are available for the computerized ADR detection methods, which include large-scale compound databases containing structure, bioassay, and genomic information, such as the National Institutes of Health's (NIH's) Molecular Libraries Initiative (MLI) [14], as well as comprehensive clinical observational data like electronic medical record (EMR) databases.

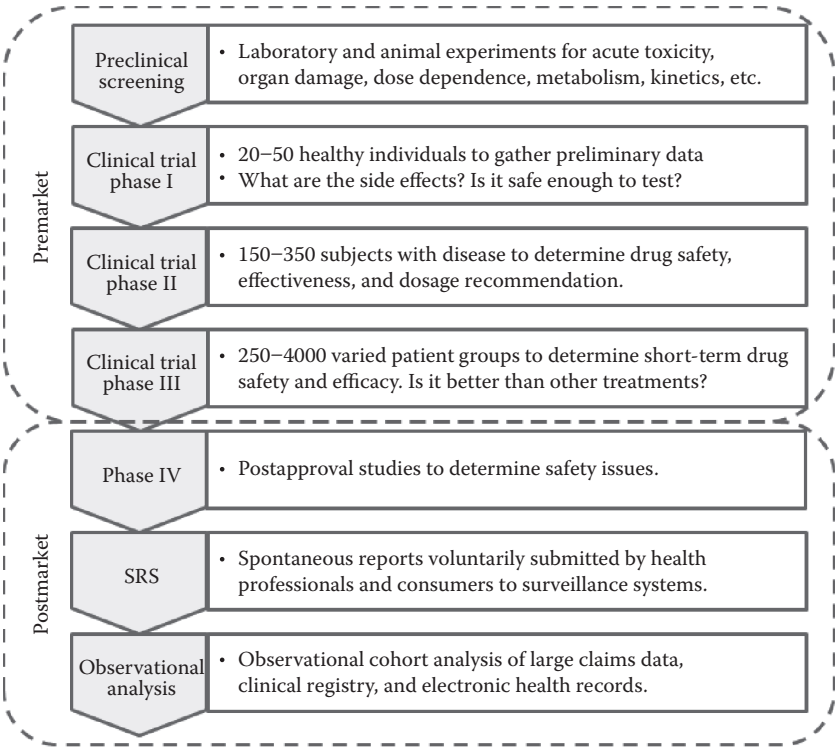


FIGURE 27.2
Pharmacovigilance at different stages of drug development. Different types of data are generated at each stage and can be used for drug safety surveillance. SRS, spontaneous reporting system.

At the premarketing stages of a drug, computational PhV efforts primarily focus on predicting potential ADRs using preclinical characteristics of the compounds (e.g., drug targets, chemical structure), or screening data (e.g., bioassay data). At the postmarketing stage, PhV has traditionally been involved in mining spontaneous reports submitted to national surveillance systems. In recent years, the postmarketing PhV research efforts have shifted toward the use of data generated from other platforms outside the conventional framework such as EMRs, clinical registries, biomedical literature, and patient-reported data in online health forums. Furthermore, an emerging trend of PhV is to link preclinical screening data from the experimental platform with human safety information observed in the postmarketing phase. The following sections provide a general overview of the current computational methodologies applied for PhV utilizing data accumulated at different stages of drug development.

27.3 Premarketing Surveillance

Many PhV efforts at the premarketing stage have been devoted to predicting or assessing potential ADRs during preclinical tests. Before a candidate compound can be approved to be tested on humans in clinical trials, it must pass a preclinical safety screening. One of the

fundamental screening methods is to apply broad-scale *in vitro* pharmacology profiling to test new chemical compounds with biochemical and cellular assays [15]. The hypothesis is that if a compound binds to a certain target, then its effect may translate into a possible occurrence of an ADR in humans. However, experimental detection of ADRs remains challenging in terms of cost and efficiency [15]. A large amount of academic research activities have been devoted to developing computational approaches to predict potential ADRs using pharmacological properties of the compounds or bioassay screening data. The existing research can be roughly categorized into protein target-based and chemical structure-based approaches, while others have also explored an integrative approach.

27.3.1 Protein Target-Based Approach

Drugs typically work by activating or inhibiting the function of a protein that plays an important role in the disease mechanism, which in turn results in therapeutic benefits to a patient. Thus, drug design essentially involves the design of small molecules that have complementary shapes and charges to the protein target with which they can bind and interact. ADRs are complex phenomenological observations on drugs that have been attributed to a variety of molecular scenarios, such as unexpected interaction with the primary targets or off-targets, downstream pathway perturbations, and kinetics [16]. Many researchers believe that direct interaction with proteins is one of the most important scenarios [15,17]. Table 27.1 provides a summary of the key articles.

Through hierarchical clustering of biological activity spectra and adverse event data on 1045 prescription drugs and 92 ligand-binding assays, Fliri et al. [18] have shown that drugs with similar *in vitro* protein binding profiles tend to exhibit a similar array of side effects. This concept was further illustrated by Campillos et al. [19], who extrapolated new drug targets by analyzing the likelihood of protein target sharing for 277,885 pairs of 746 marketed drugs using drug side-effect similarities. Furthermore, Scheiber et al. [20] demonstrated the concept by comparing pathways affected by toxic compounds versus those affected by nontoxic compounds. Fukuzaki et al. [21] proposed a method to predict ADRs

TABLE 27.1

Protein Target-Based Approaches to PhV for Preclinical Safety Screening

Concept	Method	Article
Similar *in vitro* protein binding profiles tend to exhibit similar ADRs.	Clustering	Fliri et al. [18] Campillos et al. [19]
Compounds with similar toxicity profiles may have common pathway activation conditions.	Bayesian model Network analysis	Scheiber et al. [20] Fukuzaki et al. [21]
ADRs may be explained by a drug's protein–ligand binding network where many off-targets can bind to the drug and lead to unexpected ADRs.	Statistical models for binding site alignment	Xie et al. [22]
ADR similarity could be caused by their target proteins being close in the protein interaction network.	Statistical measure to assess the distance between two drugs	Brouwers et al. [23]
Molecular actors of ADRs may involve interactions detectable using compound bioactivity screening data.	Logistic regression	Pouliot et al. [24]
In vitro drug target activities are associated with ADRs.	Disproportionality analysis	Lounkine et al. [25]

using subpathways that share correlated modifications of gene-expression profiles in the presence of the drug of interest. In order to find the "cooperative pathways" (pathways that function together), they developed an algorithm called Cooperative Pathway Enumerator (CREPE) to select combinations of subpathways that have common activation conditions.

Xie et al. [22] developed a chemical systems biology approach to identify off-targets of a drug by docking the drug into binding pockets of proteins that are similar to its primary target. Then the drug–protein interaction pair with the best docking score was mapped to known biological pathways to identify potential off-target binding networks of the drug. Unfortunately, scalability of the method is hindered by its requirement for protein 3-D structures and known biological pathways.

Brouwers et al. [23] quantified the contribution of a protein interaction network neighborhood on the observed side-effect similarity of drugs. Their fundamental idea is that side-effect similarity of drugs can be attributed to their target proteins being close in a molecular network. They proposed a pathway neighborhood measure to assess the closest distance of drug pairs according to their target proteins in the human protein–protein interaction network. The authors observed that network neighborhoods only account for 5.8% of the side-effect similarities, compared to 64% from shared drug targets.

Pouliot et al. [24] applied logistic regression (LR) models to identify potential ADRs manifesting in 19 specific system organ classes (SOCs), as defined by the Medical Dictionary for Regulatory Activities [26], across 485 compounds in 508 bioassays in the PubChem database [27,28]. The models were evaluated using leave-one-out cross-validation. The mean area under the receiver operating characteristic curve (AUC) ranged from 0.60 to 0.92 across different SOCs.

Lounkine et al. [25] recently systematically evaluated the potential clinical relevance of protein targets with ADRs for 2760 drugs. The authors used disproportionality analysis (DPA) in conjunction with a chi-squared test to assess the associations for all target–ADR pairs and identified a total of 3257 significant target–ADR association pairs.

27.3.2 Chemical Structure-Based Approach

The chemical structure-based approach attempts to link ADRs to their chemical structures. Table 27.2 summarizes the key articles.

As a proof-of-concept, Bender et al. [29] explored the chemical space of drugs and established its correlation for ADR prediction; however, the positive predictive value was quite low, under 0.5. Thereafter, Scheiber et al. [30] presented a global analysis that identified

TABLE 27.2

Chemical Structure-Based Approaches to PhV for Preclinical Safety Screening

Concept	Method	Article
Investigate correlation between chemical structures and ADRs	Bayesian model	Bender et al. [29]
	Build individual ADR model using Bayesian model and use Pearson correlation to assess similarity between a pair of ADR models	Scheiber et al. [30]
	Support vector machine (SVM)	Yamanishi et al. [31]
	Decision tree	Hammann et al. [32]
	Compared nearest neighbor (NN), SVM, ordinary canonical correlation analysis (OCCA), and sparse canonical correlation analysis (SCCA)	Pauwels et al. [33]

chemical substructures associated with ADRs, but the method was not designed to predict ADRs for any specific drug molecule. Yamanishi et al. [31] proposed a method that predicted pharmacological effects or adverse effects from chemical structures and then used the effect similarity to infer drug–target interactions.

Hammann et al. [32] employed the decision tree model to determine the chemical, physical, and structural properties of compounds that predispose them to causing ADRs in different organ systems. In the study, the authors focused on ADRs in the central nervous system (CNS), liver, and kidney as well as allergic ADRs for 507 compounds. The features used were numerical attributes computed from a compound's structure, which included elemental analysis (e.g., atom count), charge analysis (e.g., polarizability, ion charge, topological polar surface area), and geometry (e.g., number of aromatic rings, rotable bonds), as well as partitioning coefficients and miscellaneous other characteristics (e.g., indicators of hydrogen bonding) [32]. Their decision tree model was shown to produce predictive accuracies ranging from 78.9 to 90.2% for allergic, renal, CNS, and hepatic ADRs.

Pauwels et al. [33] developed a sparse canonical correlation analysis (SCCA) method to predict high-dimensional side-effect profiles of drug molecules based on the chemical structures. The authors demonstrated the usefulness of SCCA by predicting 1385 side effects in the side effect resource (SIDER) database [34] from the chemical structures of 888 approved drugs. They compared five algorithms: random assignment (random) as a baseline, nearest neighbor (NN), support vector machine (SVM), ordinary canonical correlation analysis (OCCA), and SCCA for their abilities to predict known side-effect profiles through five-fold cross-validation. The best resulting AUC scores were 0.6088, 0.8917, 0.8930, 0.8651, and 0.8932 for random, NN, SVM, OCCA, and SCCA, respectively. Their results suggest that the proposed method, SCCA, outperforms OCCA, and its performance is comparable to SVM and NN. The main advantage of OCCA and SCCA over other algorithms is their biological interpretability to understand relationships between the chemical substructures and ADRs.

27.3.3 Integrative Approach

Recently, computational approaches that integrate various types of data relating to drugs for ADR prediction have gained increasing interest. Huang et al. [35] proposed a new computational framework to predict ADRs by integrating systems biology data that include protein targets, protein–protein interaction networks, gene ontology (GO) annotation [36], and reported side effects. SVM was applied as the predictive model for heart-related ADRs (i.e., cardiotoxicity), which resulted in the highest AUC of 0.771. Soon after, Cami et al. [37] developed another ADR prediction framework by combining the network structure formed by drug–ADR relationships (809 drugs and 852 ADRs) and information regarding specific drugs and adverse events. The LR model was used as the predictive model and achieved an AUC of 0.87.

Despite the success of using chemical and biological information on drugs for ADR prediction, few studies have investigated the use of phenotypic information (e.g., indication and other known ADRs). Existing resources, such as the SIDER database [34], contain comprehensive drug phenotypic information, which has been demonstrated to be useful for other drug-related studies [19]. Liu et al. [38] investigated the use of phenotypic information, together with chemical and biological properties of drugs, to predict ADRs. Similar to the work by Pauwels et al. [33], Liu et al. conducted a large-scale study to develop and validate the ADR prediction model on 1,385 known ADRs for 832 US FDA-approved drugs in SIDER with five machine learning algorithms: LR, naïve Bayes (NB), *k*-nearest neighbor (KNN), random forest (RF), and SVM. Evaluation results showed that

the integration of chemical, biological, and phenotypic properties outperforms the chemical structured-based method (from 0.9054 to 0.9524 with SVM) and has the potential to detect clinically important ADRs at both the preclinical and postmarket phases for drug safety surveillance.

27.4 Postmarketing Surveillance

Although a drug undergoes extensive screening (Figure 27.2) before its approval by the FDA, many ADRs may still be missed because the clinical trials are often small, short, and biased by excluding patients with comorbid diseases. In general, for any potentially rare ADR with an occurrence rate less than 0.1%, it will be extremely difficult for the premarketing trials to identify the related rare ADRs due to the limitation in size [2]. In addition, premarketing trials do not mirror actual clinical use situations for diverse (e.g., inpatient) populations; therefore, it is important to continue the surveillance postmarket. Several unique data sources are available for postmarketing PhV.

27.4.1 Spontaneous Reports

Spontaneous reporting systems (SRSs) have served as the core data collection system for postmarketing drug surveillance since 1960. Some of the prominent SRSs are the Adverse Event Reporting System (AERS) maintained by the US FDA and the VigiBase managed by the World Health Organization (WHO). Although the SRSs may differ in structure and content, most of them rely on health care professionals and consumers to identify and report suspected cases of ADRs. Information collected usually includes the drugs suspected to cause the ADR, concomitant drugs, indications, suspected events, and limited demographic information. Many postmarketing surveillance analyses are based on these reports voluntarily submitted to the national SRSs, which include DPA for identifying single-drug and ADR associations and data mining algorithms for discovering combinatorial adverse effect of multiple drugs.

These systems are quite useful and have been in existence for many years but are subject to significant limitations in conducting postmarketing surveillance. SRSs suffer from severe underreporting. The extent of underreporting of ADRs to SRSs was investigated across 37 studies, and the median underreporting rate was estimated to be 94% [39]. Additionally, these systems do not have "denominator"-level information, which limits the types of statistical methods that can be applied to them, and reporting rates are influenced by media reports and publicity as much as by clinical and safety concerns [40]. SRSs are the most effective at detecting completely unexpected events, in which any reports of that adverse event profile are notable and suggest the need for further evaluation of the medication.

27.4.1.1 DPA for Single-Drug and ADR Associations

DPA has been the driving force behind most PhV methods involving SRS data. The first use of DPA for drug safety dates back to the early 1980s [41]. It is not our intention to exhaustively list and examine all relevant DPA work in this chapter. Rather, we aim to present the basic concepts and highlight some representative work here. DPA mainly

TABLE 27.3

Contingency Table Commonly Used in DPA Methods

	ADR Occurred	No ADR Occurred
Exposed to drug	A	B
Exposed to other drugs	C	D

TABLE 27.4

Definitions of the Frequentist Measures of Association

Common Frequentist Association Measures	Definition
Relative reporting ratio (RRR)	((a + b + c + d) * a)/((a + c) * (a + b))
Proportional reporting ratio (PRR)	(a/(a + b))/(c/(c + d))
Reporting odds ratio (ROR)	(a * d)/(c * b)

involves frequency analyses of 2 × 2 contingency tables to quantify the degree to which a drug and ADR co-occurs "disproportionally" compared with what would be expected if there were no association (Table 27.3) [42].

Straightforward DPA methods involve the calculation of frequentist metrics. Some of the widely applied frequentist measures (Table 27.4) include the relative reporting ratio (RRR) [43], proportional reporting ratio (PRR) [44] adopted by the Medicines and Healthcare Products Regulatory Agency (MHRA) in the United Kingdom, and reporting odds ratio (ROR) [45] adopted by the Netherlands Pharmacovigilance Center. Hypothesis tests of independence (i.e., chi-square test or Fisher's exact test) are typically used in conjunction with the above association estimates as extra precautionary measures.

In addition to the frequentist approaches, more complex algorithms based on Bayesian statistics were developed such as the gamma-Poisson shrinker (GPS) [46], the multi-item gamma-Poisson shrinker (MGPS) [47,48], and the empirical Bayesian geometric means (EBGMs) [49,50]. The GPS and MGPS methods are currently utilized by the FDA. Moreover, Bayesian Confidence Propagation Neural Network (BCPNN) [51–53] analysis was proposed based on Bayesian logic where the relation between the prior and posterior probability was expressed as the "information component" (IC). The IC given by the BCPNN is adopted by the WHO Uppsala Monitoring Center (UMC) to monitor safety signals in their SRSs.

Other groups have also investigated the James–Stein type of shrinkage estimation strategies in a Bayesian LR model to analyze SRSs data [54]. More recently, Ahmed et al. [55,56] proposed false discovery rate (FDR) estimation for the frequentist methods to address the limitation of arbitrary thresholds. Caster et al. [57] proposed the first implementation of shrinkage LR for large-scale pattern discovery of ADRs using the WHO global individual case safety reports database, VigiBase. Furthermore, Hopstadius and Noren [58] proposed a framework to detect local associations by employing shrinkage observed-to-expected ratios and multiple stratifications. As of now, there is no consensus on which DPA method is better due to the lack of a gold-standard data set for systematic evaluation.

27.4.1.2 Data Mining Algorithms for Multidrug and ADR Associations

The above-mentioned DPA methods are effective in detecting single-drug and ADR associations, but multidrug and ADR associations are also important because they can suggest

possible drug–drug interactions. A typical SRS database contains thousands of drugs and ADRs, so it is impractical to enumerate all combinations for statistical analysis. Thus, data mining algorithms have been employed to address this problem.

Harpaz et al. [59] applied the association rule mining algorithm to identify multi-item ADRs. Using a set of 162,744 reports submitted to the FDA in 2008, 1,167 multi-item ADR associations were identified, and among those, 67% were validated by a domain expert. Later, Harpaz et al. [60] applied the biclustering algorithm to identify drug groups that share a common set of ADRs in SRSs data. Tatonetti et al. [61] proposed an algorithm to mine drug–drug interactions from the adverse event reports by analyzing latent signals that indirectly provide evidence for ADRs. Interestingly, they discovered that coadministration of pravastatin and paroxetine had a synergistic effect on blood glucose. In contrast, neither drug individually was found to be associated with such change in the glucose levels.

27.4.2 EMRs

EMRs have emerged as a prominent resource for observational research as they contain not only detailed patient information but also copious longitudinal clinical data. Recently, investigators have begun to explore the use of EMRs for PhV. Data in EMR data bases are typically in two types of formats: (1) structured (e.g., laboratory data) and (2) narrative clinical notes.

27.4.2.1 Structured Data

Several groups have employed computational methods on structured or coded data in EMRs to identify specific ADR signals [62,63]. Jin et al. [64] proposed a new interestingness measure called residual leverage for association rule mining to identify ADR signals from health care administrative databases. Ji et al. [65] introduced potential causal association rules to generate potential causal relationships between a drug and international classification diseases, 9th revision (ICD-9) coded signs or symptoms in EMRs. Noren et al. [66,67] introduced a new measure of temporal association to contrast the observed-to-expected ratio in a time period of interest to that in a predefined control period. Schildcrout et al. [68] analyzed the relationship between insulin infusion rates and blood glucose levels in patients in an intensive care unit (ICU). Zorych et al. [69] investigated various DPA methods for PhV by mapping EMR data to drug-condition 2×2 contingency tables.

Yoon et al. [70] demonstrated laboratory abnormality to be a valuable source for PhV by examining the odds ratio of laboratory abnormalities between a drug-exposed and a matched unexposed group using 10 years of EMR data. Evaluation of their algorithm on 470 randomly selected drug-and-abnormal-lab-event pairs produced a positive predictive value of 0.837 and negative predictive value of 0.659. Liu et al. [71] also designed a study to correlate abnormal laboratory results with specific drug administrations by comparing the outcomes of a drug-exposed group and a matched unexposed group using 12 years of EMR data. The authors assessed the relative merits of six DPA methods typically used for analyzing SRS data on their EMR data by systematically evaluating the methods on two independently constructed reference standard data sets of drug–event pairs.

Furthermore, Duan et al. [72] proposed a likelihood ratio model and a Bayesian network model for ADR discovery using the Observational Medical Outcomes Partnership (OMOP) [73] simulated data set. OMOP has designed and developed a procedure to generate fictional persons, drug exposure, and adverse event occurrences with predefined associations between

drugs and outcomes. Although the data are simulated, it is part of the OMOP's effort in providing data close to real clinical observational data for larger computational communities to develop and evaluate their methods. The study [73] has shown that the proposed methods performed better by 23.83% over the standard baseline algorithm, the chi-square test.

27.4.2.2 Unstructured Data

Data in narrative clinical notes is not readily accessible for data mining; thus, a natural language processing (NLP) technique is required to extract the needed information. General-purpose NLP systems (e.g., medical language extraction and encoding system [MedLEE] [74], MetaMap [75], and clinical text analysis and knowledge extraction system [cTAKES] [76]), as well as specialized systems (e.g., MedEx [77] for medication information extraction), have been developed for clinical text. Wang et al. [78] first employed NLP techniques to extract drug–ADR candidate pairs from narrative EMRs and then applied the chi-square test with an adjusted volume test to detect ADR signals. Evaluation on seven selected drugs and their known ADRs produced an overall precision and recall of 0.31 and 0.75, respectively.

Similarly, Wang et al. [79] developed other methods based on mutual information (MI) and data processing inequality (DPI) to characterize drug-and-ADR pairs extracted from EMRs. Evaluation on a random sample of two drugs and two diseases indicated an overall precision of 0.81. Furthermore, Wang et al. [80] investigated the use of filtering by sections of reports to improve the performance of NLP extraction for clinically meaningful drug-and-ADR relations. Their evaluation indicated that applying filters improved recall from 0.43 to 0.75 and precision from 0.16 to 0.31.

Sohn et al. [81] proposed a hybrid system to extract physician-asserted ADRs from clinical narratives of psychiatry and psychology patients. They first developed a rule-based system to recognize relationships between individual side effects and causative drugs. The C4.5 decision tree model was then applied to extract sentences containing pairs of side effects and causative drugs using the keyword features and expression patterns of side effects. The hybrid system had an *F*-score of 0.75 in identifying side-effect sentences.

27.4.3 Nonconventional Data Sources

27.4.3.1 Biomedical Literature

Biomedical literature can be used as a complementary resource for prioritizing drug–ADR associations generated from SRSs. Shetty and Dalal [82] retrieved articles (published between 1949 and 2009) that contain mentions of a predefined list of drug-and-ADR pairs (38 drugs and 55 ADRs) from PubMed. The authors then constructed a statistical document classifier to remove irrelevant articles with mentions of treatment relations. Finally, DPA was applied to identify statistically significant pairs from the thousands of pairs in the remaining articles. Evaluation showed that the method identified true associations with 0.41 precision and 0.71 recall.

27.4.3.2 Health Forums

Data posted by users on health-related Web sites may also contain valuable drug safety information. Leaman et al. [83] described a system to mine drug-and-ADR relationships as reported by consumers in user comments to health-related Web sites like DailyStrength (http://www.dailystrength.org/). System evaluation was conducted on a manually

annotated set of 3,600 user posts corresponding to 6 drugs. The system was shown to achieve 0.78 in precision and 0.70 in recall.

Chee et al. [84] explored the use of an ensemble classifier on data from online health forums to identify potential watch-list drugs that have an active FDA safety alert. The authors aggregated individuals' opinions and reviews of drugs and used an NLP technique to group drugs that were discussed in similar ways. Interestingly, withdrawn drugs were successfully identified based on messages even before they were removed from the market.

27.4.4 Integrative Approach

Traditionally, postmarketing ADR detection methodologies have focused on data from a single source, but the emerging belief is that combining information across sources can lead to more effective and accurate ADR signals. Matthews et al. [85] combined SRS data and literature findings to build quantitative structure–activity relationship models for several serious ADRs. Vilar et al. [86] proposed an approach to prioritize ADRs generated from SRSs based on chemical structure similarity. Harpaz et al. [87] leveraged EMRs and explicitly combined them with the SRS data to facilitate uniformed, hypothesis-free signal detection, while Vilar et al. [88] attempted to enhance the ADR signals derived from EMRs by molecular structure similarity.

27.5 Available Resources

Despite the progress made toward PhV, there is still little empirical evidence to support the use of one method or data source over another. As a result, research initiatives have been implemented so that methods and data sources can be assessed on a solid scientific footing. The following sections cover various initiatives and public resources useful for method development.

27.5.1 National Initiatives

27.5.1.1 The Sentinel Initiative

In 2007, the US Congress passed the FDA Amendments Act (FDAAA) mandating the FDA to establish an active surveillance system. In response, the FDA implemented the Sentinel Initiative [89] to develop a proactive system that will complement existing systems that FDA has in place to monitor the safety of its regulated products. The Sentinel System enables the FDA to query diverse health care databases such as EMR and administrative and insurance claims data bases so that safety issues of medical products can be evaluated quickly and securely. Since the announcement of the initiative, the FDA has supported a broad public forum to examine a range of subjects and granted pilot projects to explore many technical and policy issues in implementing such surveillance system.

27.5.1.2 OMOP

OMOP [73] is a public–private partnership launched by the Foundation for the National Institutes of Health with Pharmaceutical Research and Manufacturers of America (PhRMA) and the FDA, aiming to identify the most reliable methods for analyzing huge volumes

of data drawn from heterogeneous sources. OMOP has designed experiments to test a variety of analytical methodologies in a range of data types to look for well-known drug impacts. In 2010, OMOP organized a competition for developing signal detection methods based on simulated data. A key finding from the 2010 OMOP Cup is that the heterogeneity of data sources and methods strongly affects results. Therefore, robust methods that can be applied to multiple data sources are desired.

27.5.1.3 National Centers for Biomedical Computing

The National Centers for Biomedical Computing (NCBC) are centers funded under cooperative agreement awards with multiple NIH agencies that are intended to be the core of networked national efforts to build computational infrastructure for biomedical computing [90]. There are currently eight such funded centers, and two of these large initiatives are particularly relevant to medication surveillance.

27.5.1.4 Informatics for Integrating Biology and the Bedside

While the focus of Informatics for Integrating Biology and the Bedside (I2B2) is on data harmonization into the common I2B2 schema and information extraction with NLP and other tools, this type of infrastructure has been adopted by the Clinical and Translational Science Award (CTSA) program that currently is in operation at 60 academic medical institutions across the United States. This center and the participating consortium, particularly as the focus of integrating genomic and phenomic data into a common data platform across a large number of institutions, is providing a powerful source of data for conducting postmarketing surveillance [91].

27.5.1.5 Integrating Data for Analysis, Anonymization, and Sharing

Integrating Data for Analysis, Anonymization, and Sharing (iDASH) is the newest NCBC, funded in 2010 by the NIH [92]. The iDASH center aims to develop algorithms and tools for sharing data in a privacy-preserving manner to enable global collaborations anywhere and anytime. It will provide the biomedical researchers a needed data resource for new hypothesis generation and testing in areas like postmarketing surveillance.

27.5.1.6 European Union—Adverse Drug Reaction

European Union—Adverse Drug Reaction (EU-ADR) [93] is a Research AND development project initiated in Europe by the European Commission. Similar to OMOP, EU-ADR, aims to develop computerized systems to detect ADRs in supplementing existing SRSs by exploiting EMRs of over 30 million patients from several European countries. The project employs a variety of text mining, epidemiological, and computational techniques to analyze the clinical data to detect ADR signals. These new developments rely on the expanded secondary use of electronic health care data that typically contain time-stamped interventions, procedures, diagnoses, medications, medical narratives, and billing codes.

27.5.1.7 MLI

The MLI is motivated by two NIH missions: (1) develop new approaches to determine function and therapeutic potential of all genes in the newly sequenced human genome

and (2) accelerate the translation of basic research discoveries to new therapeutics for the benefits of public health [14]. The MLI research focuses on screening, cheminformatics, and technology development, and the small-molecule research tools produced would accelerate validation of new drug targets and enable new drug development.

27.5.2 Public Databases

27.5.2.1 DrugBank

The DrugBank database [94–96] contains detailed drug (i.e., chemical, pharmacological, and pharmaceutical) data and comprehensive drug target (i.e., sequence, structure, and pathway) information. Table 27.5 shows the contents in DrugBank (accessed October 2012).

27.5.2.2 PubChem

PubChem is part of the NIH's Molecular Libraries Roadmap Initiative [27,28] to provide information on biological activities of small molecules. It is organized as three linked databases—PubChem Substance, PubChem Compound, and PubChem BioAssay—within the National Center of Biotechnology Information (NCBI) Entrez information retrieval system. The PubChem Substance database contains substance information submitted by depositors, which include any chemical structure information, chemical names, and comments. The PubChem Compound database is composed of a nonredundant set of standardized and validated chemical structures. The PubChem BioAssay database contains bioactivity screens of chemical substances. All this information can be used in various ways to develop ADR detection methods in the preapproval stage.

27.5.2.3 SIDER

SIDER is a side-effect resource [34] developed to aggregate dispersed public information on drug side effects. It contains information on marketed medicines and their recorded ADRs extracted from public documents and package inserts. In the version released in October 2012, SIDER contains 996 drugs, 4,192 side effects, and 99,423 drug–ADR pairs.

27.5.2.4 Pharmacogenomics Knowledge Base

The Pharmacogenomics Knowledge Base (PharmGKB) [97] is an interactive tool for researchers to investigate how genetic variation affects drug responses. PharmGKB

TABLE 27.5

DrugBank Statistics, Accessed in October 2012

Entries	Statistic
Drug entries	6711
FDA-approved small-molecule drugs	1447
FDA-approved biotech (protein/peptide) drugs	131
Nutraceuticals	85
Experimental drugs	5080
Nonredundant protein (i.e., drug target, enzyme, transporter, carrier)	4227

captures complex relationships between genes, variants, drugs, diseases, and pathways. Data in PharmGKB have been curated from multiple sources and can be used to facilitate genome-wide pharmacogenomics (PGx) studies, predict gene–drug relationships and support data-sharing consortia investigating clinical applications of PGx.

27.5.2.5 Connectivity Map

The Connectivity Map [98] is a comprehensive research effort for using genomics in a drug discovery framework by generating a detailed map that links gene patterns associated with disease to corresponding patterns produced by drug candidates. The ultimate goal is to connect human diseases with the genes that underlie them and drugs that treat them. The Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard brought together specialists in biology, genomics, computing science, pharmacology, chemistry, and medicine to build the Connectivity Map.

27.5.2.6 NIH Chemical Genomics Center Pharmaceutical Collection

The NIH Chemical Genomics Center (NCGC) was established to create a national resource for chemical probe development. The NCGC Pharmaceutical Collection is a comprehensive repository of approved and investigational drugs for high-throughput screening and provides a valuable resource for both validating new disease models and better understanding the molecular basis of disease pathology and intervention [99].

27.5.2.7 Therapeutic Target Database

The Therapeutic Target Database (TTD) was developed to provide comprehensive information about known and explored therapeutic protein and nucleic acid targets, targeted disease, pathway information, and corresponding drugs directed at each of these targets to facilitate target-oriented drug discovery [100]. The database, accessed in November 2012, contains 2025 targets and 17,816 drugs.

27.6 Future Perspectives and Challenges

In this chapter, we have provided a general overview of the rich and diverse applications of AI approaches with respect to different perspectives of PhV. We have seen more and more opportunities emerging as a result of new data generated from various platforms, including high-throughput experiments, EMRs, literature, and self-reported health forums.

It is evident that a new trend of computational approaches for PhV is to link preclinical data from the experimental platform with human safety information observed in the postmarketing phase [101]. From the systems biology perspective, drugs are small molecules that can induce perturbations to biological systems, which involve various molecular interactions such as protein–protein interactions, signaling pathways, and pathways of drug action and metabolism. The body's response to a drug is a complex phenomenological observation that includes both favorable and unfavorable reactions. When a drug is absorbed into the body and interacts with its intended targets, favorable effects are expected. However, a drug often binds to other protein pockets with varying affinities

(off-target interactions), leading to observed side effects. Hence, to understand ADRs, it is desirable to incorporate various data sources into one framework.

Moreover, it is critical to identify multi-item ADR associations as they may suggest drug interactions. For instance, if a patient is taking two drugs at the same time and one increases the effect of the other, the patient may experience an overdose. Similarly, if the action of a drug is inhibited by the other, the intended therapeutic effect may be reduced. Drug interactions may also increase the risk of ADRs. Statistical analysis works well with the identification of single drug-and-ADR signals but is not suitable for drug interaction identification. Alternatively, data mining algorithms such as *a priori* algorithms and clustering algorithms are applicable and useful. This provides an excellent opportunity for computer scientists to develop new algorithms for drug interaction detection.

Furthermore, EMRs have become an obvious data choice for PhV. However, many challenges exist in mining EMRs for ADR prediction. Much detailed and useful information is embedded in the narrative notes, making data extraction difficult. There have been studies using NLP techniques to extract drug and ADR concepts from narrative notes for association analysis. Wang et al. [80] have shown that filtering information based on note sections improves the identification of drug-and-ADR relations. Despite the current success, further investigation of other methods, for example, more sophisticated statistical methods and temporal models, is needed.

As yet, few studies have explored the automatic construction of large-cohort or case–control studies from EMRs for ADR prediction. There are many issues to consider in the NLP-based cohort/case–control study construction, for instance, how to extract adverse event concepts from narrative notes. It is common for multiple concepts to describe the same outcome/phenotype. Since most current practices focus on a single outcome at a time, phenotype is usually defined manually by experts. However, for large-scale ADR studies, how do we automatically define the phenotypes? Also, a key distinguishing feature of EMR-based methods is the use of temporal information to define time frames as surveillance windows to identify ADRs; for example, adverse outcomes must occur at certain time periods after drug exposure. However, to accurately determine the temporal relations between events in the narrative text remains challenging. In fact, each of the above-mentioned questions is an active area of research.

After overcoming the above hurdles in study design, one must keep in mind the confounding problem during analysis. For instance, the basic concept behind the cohort design is to partition a population into those who are "exposed" (taking a specific drug) and "unexposed" (taking a comparator drug or not taking a specific drug). A drug is determined to be associated with a specified outcome when the outcome occurs more often in the exposed group than in the unexposed group. Since the group assignment is not random, increased attention must be given when selecting the "unexposed" group. A common technique to minimize the issues caused by confounding and bias is to match patient groups based on a set of basic covariates such as gender, age, and comorbidities. On the other hand, case–control designs divide the study population into those who experienced the outcome ("case") and those who did not experience the outcome ("control"). If the drug exposure occurs more frequently in the cases than in the controls, the drug is said to be associated with the outcome. The same issues with confounding apply to case–control studies. Matching two groups before analysis is usually a good idea.

Lastly, it is important to note that most of the existing computational methodologies for PhV involve assessment of association between a drug and an ADR. However, association does not necessarily imply causation. Intuitively, causation requires not only correlation but also a counterfactual dependence. Inferring cause-and-effect relationships is an

intrinsically hard problem in data mining and needs to be further investigated for PhV applications.

A young but rapidly advancing field of health care is personalized medicine—customizing health care decisions and practices to individual patients. The nature of diseases (e.g., their onset and course) is usually as unique as the patients who have them. Every person may respond to drugs or interventions differently because of their genetic makeup, clinical conditions, and environmental information. Classen et al. [102] estimated that about 50% of ADRs are likely to be related to genetic factors. PGx research can impact this problem by linking inherited differences to variable drug responses. Recent success stories of PGx studies include more accurate and safe dosing of warfarin following genotyping at *CYP2C9* [103] and *VKORC1* [104] and recognizing clopidogrel resistance with *CYP2C19* variants [105–107]. Despite significant progress, our knowledge of genetic factors contributing to ADRs is still limited [108]. To accelerate PGx discovery and knowledge validation, more efficient computational approaches are required, presenting the informatics community ample opportunities and challenges.

References

1. Ledley RS. Digital electronic computers in biomedical science. *Science*. 1959 Nov 6;**130**(3384):1225–34.
2. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science*. 1959;**130**(3366):9–21.
3. Shortliffe EH. The adolescence of AI in medicine: will the field come of age in the '90s? *Artif Intell Med*. 1993 Apr;**5**(2):93–106.
4. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, Abu-Hanna A. The coming of age of artificial intelligence in medicine. *Artif Intell Med*. 2009;**46**(1):5–17.
5. Pirmohamed M, Breckenridge AM, Kitteringham NR, Park BK. Adverse drug reactions. *BMJ*. 1998;**316**(7140):1295–8.
6. Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annest JL. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA*. 2006;**296**(15):1858–66.
7. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA*. 1998;**279**(15):1200–5.
8. Moore TJ, Cohen MR, Furberg CD. Serious adverse drug events reported to the Food and Drug Administration, 1998–2005. *Arch Intern Med*. 2007;**167**(16):1752–9.
9. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y. When good drugs go bad. *Nature*. 2007;**446**(7139):975–7.
10. Leone R, Sottosanti L, Luisa Iorio M, Santuccio C, Conforti A, Sabatini V, Moretti U, Venegoni M. Drug-related deaths: an analysis of the Italian spontaneous reporting database. *Drug Saf*. 2008;**31**(8):703–13.
11. van der Hooft CS, Sturkenboom MC, van Grootheest K, Kingma HJ, Stricker BH. Adverse drug reaction-related hospitalisations: a nationwide study in The Netherlands. *Drug Saf*. 2006;**29**(2):161–8.
12. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;**9**(3):203–14.
13. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*. 2008;**4**(11):682–90.
14. Austin CP, Brady LS, Insel TR, Collins FS. NIH Molecular Libraries Initiative. *Science*. 2004;**306**(5699):1138–9.

15. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: *in vitro* safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov Today*. 2005;**10**(21):1421–33.
16. Liebler DC, Guengerich FP. Elucidating mechanisms of drug-induced toxicity. *Nat Rev Drug Discov*. 2005;**4**(5):410–20.
17. Blagg J. Structure–activity relationships for *in vitro* and *in vivo* toxicity. *Annu Rep Med Chem*. 2006;**41**(2006):353–68.
18. Fliri AF, Loging WT, Thadeio PF, Volkmann RA. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat Chem Biol*. 2005;**1**(7):389–97.
19. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science*. 2008;**321**(5886):263–6.
20. Scheiber J, Chen B, Milik M, Sukuru SC, Bender A, Mikhailov D, Whitebread S, Hamon J, Azzaoui K, Urban L, Glick M, Davies JW, Jenkins JL. Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. *J Chem Inf Model*. 2009;**49**(2):308–17.
21. Fukuzaki M, Seki M, Kashima H, Sese J. Side effect prediction using cooperative pathways. In *IEEE International Conference on Bioinformatics and Biomedicine*. Washington DC; 2009. p. 142–7.
22. Xie L, Li J, Bourne PE. Drug discovery using chemical systems biology: identification of the protein–ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput Biol*. 2009;**5**(5):e1000387.
23. Brouwers L, Iskar M, Zeller G, van Noort V, Bork P. Network neighbors of drug targets contribute to drug side-effect similarity. *PLoS One*. 2011;**6**(7):e22187.
24. Pouliot Y, Chiang AP, Butte AJ. Predicting adverse drug reactions using publicly available PubChem BioAssay data. *Clin Pharmacol Ther*. 2011;**90**(1):90–9.
25. Lounkine E, Keiser MJ, Whitebread S, Mikhailov D, Hamon J, Jenkins JL, Lavan P, Weber E, Doak AK, Cote S, Shoichet BK, Urban L. Large-scale prediction and testing of drug activity on side-effect targets. *Nature*. 2012;**486**(7403):361–7.
26. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf*. 1999;**20**:109–17.
27. Chen B, Wild D, Guha R. PubChem as a source of polypharmacology. *J Chem Inf Model*. 2009;**49**(9):2044–55.
28. Bolton E, Wang Y, Thiessen PA, Bryant SH. PubChem: integrated platform of small molecules and biological activities. Chapter 12 in *Annual Reports in Computational Chemistry*. Washington DC: American Chemical Society; 2008.
29. Bender A, Scheiber J, Glick M, Davies JW, Azzaoui K, Hamon J, Urban L, Whitebread S, Jenkins JL. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem*. 2007;**2**(6):861–73.
30. Scheiber J, Jenkins JL, Sukuru SC, Bender A, Mikhailov D, Milik M, Azzaoui K, Whitebread S, Hamon J, Urban L, Glick M, Davies JW. Mapping adverse drug reactions in chemical space. *J Med Chem*. 2009;**52**(9):3103–7.
31. Yamanishi Y, Kotera M, Kanehisa M, Goto S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics*. 2010;**26**(12):i246–54.
32. Hammann F, Gutmann H, Vogt N, Helma C, Drewe J. Prediction of adverse drug reactions using decision tree modeling. *Clin Pharmacol Ther*. 2010;**88**(1):52–9.
33. Pauwels E, Stoven V, Yamanishi Y. Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC Bioinformatics*. 2011;**12**:169.
34. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol*. 2010;**6**:343.
35. Huang LC, Wu X, Chen JY. Predicting adverse side effects of drugs. *BMC Genomics*. 2011;**12 Suppl 5**:S11.

36. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;**25**(1):25–9.
37. Cami A, Arnold A, Manzi S, Reis B. Predicting adverse drug events using pharmacological network models. *Sci Transl Med*. 2011;**3**(114):114ra27.
38. Liu M, Wu Y, Chen Y, Sun J, Zhao Z, Chen XW, Matheny ME, Xu H. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc*. 2012;**19**:e28–e35.
39. Hazell L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. *Drug Saf*. 2006;**29**(5):385–96.
40. O'Shea JC, Kramer JM, Califf RM, Peterson ED. Part I: Identifying holes in the safety net. *Am Heart J*. 2004;**147**(6):977–84.
41. Montastruc JL, Sommet A, Bagheri H, Lapeyre-Mestre M. Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. *Br J Clin Pharmacol*. 2011;**72**(6):905–8.
42. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009;**18**(6):427–36.
43. Hauben M, Madigan D, Gerrits CM, Walsh L, Van Puijenbroek EP. The role of data mining in pharmacovigilance. *Expert Opin Drug Saf*. 2005;**4**(5):929–48.
44. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;**10**(6):483–6.
45. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf*. 2002;**25**(6):381–92.
46. Ahmed I, Haramburu F, Fourrier-Reglat A, Thiessard F, Kreft-Jais C, Miremont-Salame G, Begaud B, Tubert-Bitter P. Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting. *Stat Med*. 2009;**28**(13):1774–92.
47. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat*. 1999;**53**(3):177–202.
48. Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJ, Yuen N. Novel statistical tools for monitoring the safety of marketed drugs. *Clin Pharmacol Ther*. 2007;**82**(2):157–66.
49. DuMouchel W, Smith ET, Beasley R, Nelson H, Yang X, Fram D, Almenoff JS. Association of asthma therapy and Churg-Strauss syndrome: an analysis of postmarketing surveillance data. *Clin Ther*. 2004;**26**(7):1092–104.
50. Gould AL. Accounting for multiplicity in the evaluation of "signals" obtained by data mining from spontaneous report adverse event databases. *Biom J*. 2007;**49**(1):151–65.
51. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*. 1998;**54**(4):315–21.
52. Lindquist M, Edwards IR, Bate A, Fucik H, Nunes AM, Stahl M. From association to alert—a revised approach to international signal analysis. *Pharmacoepidemiol Drug Saf*. 1999;**8 Suppl 1**:S15–25.
53. Lindquist M, Stahl M, Bate A, Edwards IR, Meyboom RH. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. *Drug Saf*. 2000;**23**(6):533–42.
54. An L, Fung KY, Krewski D. Mining pharmacovigilance data using Bayesian logistic regression with James-Stein type shrinkage estimation. *J Biopharm Stat*. 2010;**20**(5):998–1012.
55. Ahmed I, Dalmasso C, Haramburu F, Thiessard F, Broet P, Tubert-Bitter P. False discovery rate estimation for frequentist pharmacovigilance signal detection methods. *Biometrics*. 2010;**66**(1):301–9.

56. Ahmed I, Thiessard F, Miremont-Salame G, Begaud B, Tubert-Bitter P. Pharmacovigilance data mining with methods based on false discovery rates: a comparative simulation study. *Clin Pharmacol Ther*. 2010;**88**(4):492–8.
57. Caster O, Noren GN, Madigan D, Bate A. Large-scale regression-based pattern discovery: the example of screening the WHO global drug safety database. *Stat Anal Data Min*. 2010;**3**(4):197–208.
58. Hopstadius J, Noren GN. Robust discovery of local patterns: subsets and stratification in adverse drug reaction surveillance. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. Miami, FL; 2012. pp. 265–73.
59. Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics*. 2010;**11 Suppl 9**:S7.
60. Harpaz R, Perez H, Chase HS, Rabadan R, Hripcsak G, Friedman C. Biclustering of adverse drug events in the FDA's spontaneous reporting system. *Clin Pharmacol Ther*. 2011;**89**(2):243–50.
61. Tatonetti NP, Denny JC, Murphy SN, Fernald GH, Krishnan G, Castro V, Yue P, Tsao PS, Kohane I, Roden DM, Altman RB. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clin Pharmacol Ther*. 2011;**90**(1):133–42.
62. Brown JS, Kulldorff M, Chan KA, Davis RL, Graham D, Pettus PT, Andrade SE, Raebel MA, Herrinton L, Roblin D, Boudreau D, Smith D, Gurwitz JH, Gunter MJ, Platt R. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf*. 2007;**16**(12):1275–84.
63. Berlowitz DR, Miller DR, Oliveria SA, Cunningham F, Gomez-Caminero A, Rothendler JA. Differential associations of beta-blockers with hemorrhagic events for chronic heart failure patients on warfarin. *Pharmacoepidemiol Drug Saf*. 2006;**15**(11):799–807.
64. Jin HD, Chen J, He HX, Williams GJ, Kelman C, O'Keefe CM. Mining unexpected temporal associations: applications in detecting adverse drug reactions. *IEEE Trans Inf Technol Biomed*. 2008;**12**(4):488–500.
65. Ji YQ, Ying H, Dews P, Mansour A, Tran J, Miller RE, Massanari RM. A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance. *IEEE Trans Inf Technol Biomed*. 2011;**15**(3):428–37.
66. Noren GN, Bate A, Hopstadius J, Star K, Edwards IR. Temporal pattern discovery for trends and transient effects: its applications to patient records. In *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining SIGKD 2008*. Las Vegas, NV; 2008. pp. 963–71.
67. Noren GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov*. 2010;**20**(3):361–87.
68. Schildcrout JS, Haneuse S, Peterson JF, Denny JC, Matheny ME, Waitman LR, Miller RA. Analyses of longitudinal, hospital clinical laboratory data with application to blood glucose concentrations. *Stat Med*. 2011;**30**(27):3208–20.
69. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Stat Methods Med Res*. 2013;**22**(1):39–56.
70. Yoon D, Park MY, Choi NK, Park BJ, Kim JH, Park RW. Detection of adverse drug reaction signals using an electronic health records database: Comparison of the Laboratory Extreme Abnormality Ratio (CLEAR) algorithm. *Clin Pharmacol Ther*. 2012;**91**(3):467–74.
71. Liu M, McPeek Hinz ER, Matheny ME, Denny JC, Schildcrout JS, Miller RA, Xu H. Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. *J Am Med Inform Assoc*. 2013;**20**(3):420–6.
72. Duan L, Kohoshneshin M, Street W, Liu M. Adverse drug effect detection. *IEEE J of Biomed Health Inform.* 2013;**17**(2):305–11.
73. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, Welebob E, Scarnecchia T, Woodcock J. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med*. 2010;**153**(9):600–6.
74. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc*. 1994;**1**(2):161–74.

75. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc*. 2010;**17**(3):229–36.
76. Savova GK, Kipper-Schuler K, Buntrock JD, Chute CG. UIMA-based clinical information extraction system. LREC 2008: towards enhanced interoperability for large HLT systems: UIMA for NLP; 2008, 39.
77. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inform Assoc*. 2010;**17**(1):19–24.
78. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. *J Am Med Inform Assoc*. 2009;**16**(3):328–37.
79. Wang X, Hripcsak G, Friedman C. Characterizing environmental and phenotypic associations using information theory and electronic health records. *BMC Bioinformatics*. 2009;**10 Suppl 9**:S13.
80. Wang X, Chase H, Markatou M, Hripcsak G, Friedman C. Selecting information in electronic health records for knowledge acquisition. *J Biomed Inform*. 2010;**43**(4):595–601.
81. Sohn S, Kocher JP, Chute CG, Savova GK. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. *J Am Med Inform Assoc*. 2011;**18 Suppl 1**:i144–9.
82. Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. *J Am Med Inform Assoc*. 2011;**18**(5):668–74.
83. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In *Workshop on Biomedical Natural Language Processing*, Uppsala, Sweden: Association for Computational Linguistics; 2010. pp. 117–25.
84. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. In *AMIA Annu Symp Proc*. Washington DC; 2011. pp. 217–26.
85. Matthews EJ, Kruhlak NL, Benz RD, Aragones Sabate D, Marchant CA, Contrera JF. Identification of structure–activity relationships for adverse effects of pharmaceuticals in humans: Part C: use of QSAR and an expert system for the estimation of the mechanism of action of drug-induced hepatobiliary and urinary tract toxicities. *Regul Toxicol Pharmacol*. 2009;**54**(1):43–65.
86. Vilar S, Harpaz R, Chase HS, Costanzi S, Rabadan R, Friedman C. Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. *J Am Med Inform Assoc*. 2011;**18 Suppl 1**:i73–80.
87. Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, Chase HS, Friedman C. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc*. 2013;**20**(3):413–9.
88. Vilar S, Harpaz R, Santana L, Uriarte E, Friedman C. Enhancing adverse drug event detection in electronic health records using molecular structure similarity: application to pancreatitis. *PLoS One*. 2012;**7**(7):e41471.
89. Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network—improving the evidence of medical-product safety. *New Engl J Med*. 2009;**361**(7):645–7.
90. National Centers for Biomedical Computing. Available from: http://www.ncbcs.org/summary.html
91. Informatics for Integrating Biology & the Bedside. Available from: https://www.i2b2.org/about/index.html
92. Ohno-Machado L, Bafna V, Boxwala AA, Chapman BE, Chapman WW, Chaudhuri K, Day ME, Farcas C, Heintzman ND, Jiang X, Kim H, Kim J, Matheny ME, Resnic FS, Vinterbo SA. iDASH: integrating data for analysis, anonymization, and sharing. *J Am Med Inform Assoc*. 2012;**19**(2):196–201.
93. Coloma PM, Schuemie MJ, Trifiro G, Gini R, Herings R, Hippisley-Cox J, Mazzaglia G, Giaquinto C, Corrao G, Pedersen L, van der Lei J, Sturkenboom M. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf*. 2011;**20**(1):1–11.

94. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2010;**39**(Database issue):D1035–41.
95. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*. 2008;**36**(Database issue):D901–6.
96. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;**34**(Database issue):D668–72.
97. McDonagh EM, Whirl-Carrillo M, Garten Y, Altman RB, Klein TE. From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. *Biomark Med*. 2011;**5**(6):795–806.
98. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;**313**(5795):1929–35.
99. Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, Nguyen DT, Austin CP. The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. *Sci Transl Med*. 2011;**3**(80):80ps16.
100. Zhu F, Shi Z, Qin C, Tao L, Liu X, Xu F, Zhang L, Song Y, Zhang J, Han B, Zhang P, Chen Y. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res*. 2012;**40**(Database issue):D1128–36.
101. Harpaz R, Dumouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse Drug event discovery and analysis. *Clin Pharmacol Ther*. 2012;**91**(6):1010–21.
102. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *JAMA*. 1997;**277**(4):301–6.
103. Caraco Y, Blotnick S, Muszkat M. CYP2C9 genotype-guided warfarin prescribing enhances the efficacy and safety of anticoagulation: a prospective randomized controlled study. *Clin Pharmacol Ther*. 2008;**83**(3):460–70.
104. Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hortnagel K, Pelz HJ, Lappegard K, Seifried E, Scharrer I, Tuddenham EG, Muller CR, Strom TM, Oldenburg J. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature*. 2004;**427**(6974):537–41.
105. Simon T, Verstuyft C, Mary-Krause M, Quteineh L, Drouet E, Meneveau N, Steg PG, Ferrieres J, Danchin N, Becquemont L. Genetic determinants of response to clopidogrel and cardiovascular events. *New Engl J Med*i. 2009;**360**(4):363–75.
106. Mega JL, Close SL, Wiviott SD, Shen L, Hockett RD, Brandt JT, Walker JR, Antman EM, Macias W, Braunwald E, Sabatine MS. Cytochrome p-450 polymorphisms and response to clopidogrel. *New Engl J Med*. 2009;**360**(4):354–62.
107. Collet JP, Hulot JS, Pena A, Villard E, Esteve JB, Silvain J, Payot L, Brugier D, Cayla G, Beygui F, Bensimon G, Funck-Brentano C, Montalescot G. Cytochrome P450 2C19 polymorphism in young patients treated with clopidogrel after myocardial infarction: a cohort study. *Lancet*. 2009;**373**(9660):309–17.
108. Ingelman-Sundberg M. Pharmacogenomic biomarkers for prediction of severe adverse drug reactions. *New Engl J Med*. 2008;**358**(6):637–9.