CHAPTER

11 Overview of the role of artificial intelligence in pathology: the computer as a pathology digital assistant

John E. Tomaszewski, MD

SUNY Distinguished Professor and Chair, Pathology and Anatomical Sciences, University at Buffalo, State University of New York, Buffalo, NY, United States

Introduction

As this chapter goes to press, two great waves of change are sweeping through the science and the practice of medicine. These are the molecular revolution and computational analytics. These two scientific movements are radically altering the ways in which we approach all aspects of human biology. Previous chapters have discussed machine learning strategies in detail, with an emphasis on deep learning via various kinds of neural networks, ranging from basic convolutional models to advanced architectures such as graphic and capsule networks. In this chapter, I will describe the overall framework in which digital and computational pathologies are embedded, summarize these various strategies, and provide examples from the literature that illustrates many of the topics covered. This overview is designed to show how pathology is evolving and will continue to evolve in the age of artificial intelligence (AI). One major thesis is that human pathologists will not be replaced by computers. Instead, the computer will serve as a highly capable "digital assistant," and this synergistic pairing of human intelligence (HI) and AI will ensure that pathology will be a major driver of personalized medicine.

Computational pathology: background and philosophy

The current state of diagnostics in pathology and the evolving computational opportunities: "why now?"

Diagnostic pathology in the early 21st century is still very much an analog enterprise which uses a work process that was developed in the early 20th century and which has been substantially carried forward into the early 21st century. The processes of chemically treating and processing tissue, microtomy, staining, and the creation of

Artificial Intelligence and Deep Learning in Pathology. https://doi.org/10.1016/B978-0-323-67538-3.00011-7

glass slides which are viewable with a wide-field optical instrument (the microscope) are essentially the same processes established by early surgical pathologists. The data obtained from these processes yield continuous variables which a human pathologist diagnostician uses to drive classification systems. The output of these classifiers is in turn used by clinicians to support therapeutic decisions. This analog process is cost-effective in that it yields core actionable information in many cases and consumes only a very small fraction of total healthcare costs. These processes, however, have been fundamentally the same for 100 years and are not adapted to the rapidly changing diagnostic environment. There is a growing need for precision diagnostics. High-cost, high-risk therapeutic interventions require diagnostic and theranostic systems which model the most effective therapeutic choices with a high degree of precision and accuracy. The efficiency of the diagnostic/theranostic system has a direct and major impact on the global cost-effectiveness of the entire clinical care system.

Several new opportunities are now converging to make precision diagnostics a 21st century reality. Big data is becoming available in multiple modes. The molecular revolution which entered the practice of pathology and laboratory medicine in the 1980s is now becoming routinely available. Genomics, epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics all offer digital "multi-omic" big data opportunities in research and practice [1]. Likewise, the extension of the electronic medical record to all aspects of medical practice through the US federal "meaningful use" programs in health information technology has opened digital access to the big data of the clinical record through synoptic reporting and natural language processing. Whole slide images of tissue sections are typically multigigapixel in size and as such are a third modal window providing digital access to the big data of complex medical conditions. In tissue-based diagnostics, advances in whole slide imaging (WSI) technology are poised to cross the threshold from purely education and research use to diagnostics. The recent FDA approval of WSI systems opens up enormous opportunities for computer-aided pathological diagnosis. WSI and image analytics are poised to become mainstream methods of histopathological interpretation yielding faster, more accurate, and cheaper diagnoses, prognoses, and theranostic predictions of important diseases including cancer.

Access to high-performance computing in the medical field has also dramatically increased over the past decade. Graphical processing units and tensor processing units are now available in medical settings. Computing jobs which would previously take days or even weeks to perform in a research setting are now available through cloud computing or even at the desktop. Storage costs have seen dramatic reductions over the past 5 years.

Most importantly, our approaches to and understanding of machine learning have made a dramatic turn since 2012 when AlexNet won the ImageNet Large Scale Visual Recognition Challenge by a large margin using a convolutional neural network (CNN) [2]. With the resurgence of interest in and the success of modern neural networks and data-driven deep learning approaches, the current decade has seen a dramatic acceleration in our opportunities to learn from massive quantities

of data with limited labels. The analytics which can be applied is now beginning to match the big data of biology and medicine.

Digital pathology versus computational pathology

It is important to distinguish between the terms *digital pathology* and *computational pathology*.

Digital pathology is best understood as the digital workflow tools which are more focused on the front end of the process. Digital pathology is a dynamic, image-based environment that enables the acquisition, management, and interpretation of pathology information generated from a digitized glass slide. Healthcare applications include primary diagnosis, diagnostic consultation, intraoperative diagnosis, medical student and resident training, manual and semiquantitative review of immunohistochemistry (IHC), clinical research, diagnostic decision support, peer review, and tumor boards. Digital pathology is an innovation committed to the reduction of laboratory expenses, an improvement of operational efficiency, enhanced productivity, and improving treatment decisions and patient care.

Broadly written, the approaches to the computational modeling of complex systems of human disease from big data are termed "*computational pathology*." In 2011, Fuchs and Buhmann [3] opined that "Computational Pathology investigates a complete probabilistic treatment of scientific and clinical workflows in general pathology, i.e., it combines experimental design, statistical pattern recognition, and survival analysis within a unified framework to answer scientific and clinical questions in pathology." Computational pathology focuses on the analytics of extracting data from the output of sensors and turning those data into information and knowledge. This might include both prognostics and outcome prediction or simply a quantitative cataloging of tissue morphologies represented by computation.

Data on scale

The scientific understanding of human structure has stagnated as a largely descriptive endeavor for many decades. Structural science learning related to cells, organelles, and tissues is central to the disciplines of cell biology and pathology. With the advent of machine vision and machine learning, major advances in computing and network power, and the rapid evolution of advanced biophysical systems for cellular imaging, these disciplines are poised to embrace a quantitative paradigm for their work. The data embedded in a cellular or tissue image are deep and full of complex relationships. Thousands of image metrics derived by machine vision can be captured from a field of view and used in the modeling of complex biological or disease system outcomes. Computational advances now offer the promise of enabling the quantitative analysis of human structural data.

Biological data lives on scale. Modern imaging, image analytics, and computational methods provide tools with which to quantitatively mine the data within macroscopic (10^{1}), microscopic (10^{-7}), and submicroscopic (10^{-10}) images. The

ability to mine "subvisual" image features from digital pathology images using machine vision can develop feature data which may not be visually discernible by a pathologist.

There has also been recent substantial interest in combining and fusing radiologic imaging, and proteomics and genomics-based measurements with features extracted from digital pathology images for better prognostic prediction of disease aggressiveness and patient outcome. *Data fusion* is accomplished by computationally combining image data at these levels of scale with data from other modes of examination such as cell and molecular biological, biomechanical, and biophysical analyses. Data fusion offers the opportunity to quantitatively model complex human systems from massive multivariable statements.

We assert that *in the future, medical decision-making will rely on large datasets, collected across scale from molecular and imaging analyses, integrated by machine vision and machine learning, and that these integrated data will be used to model complex biological and disease systems. These complex computationally supported systems of data collection, integration, and modeling will be vetted by teams of humans with domain knowledge expertise.*

Machine learning tools in computational pathology: types of artificial intelligence

AI is not really a new concept. The term AI was first used by John McCarthy in 1955 [4]. He subsequently organized the Dartmouth conference in 1956 which started AI as a field. The label "AI" means very different things to different observers. For example, some commenters recognize divisions of AI as "statistical modeling" (calculating regression models and histograms) versus machine learning (Bayes, random forests, support vector machines [SVMs], shallow neural networks, or artificial neural network) versus deep learning (deep neural networks and CNNs). Others recognize categories of *traditional AI* versus *data-driven deep learning AI*. In this comparison, *traditional AI* starts with a human understanding of a domain and seeks to condition that knowledge into models which represent the world of that knowledge domain. When current lay commentators refer to "AI," however, they are usually referring to *data-driven deep learning AI*, which removes the domain knowledge inspired feature extraction part from the pipeline and develops knowledge of a domain by observing large numbers of examples from that domain.

The design approaches of traditional AI versus data-driven deep learning AI are quite different. The architects of *traditional AI learning systems* focus on building generic models. They often begin with a human understanding of the world through the statement of a prior understanding of the domain (see Fig. 11.1), develop metrics representing that prior, extract data using those metrics, and ask humans to apply class labels of interest to these data. These labels are then used to train the system to learn a hyperplane which separates one class from another. Traditional AI learning

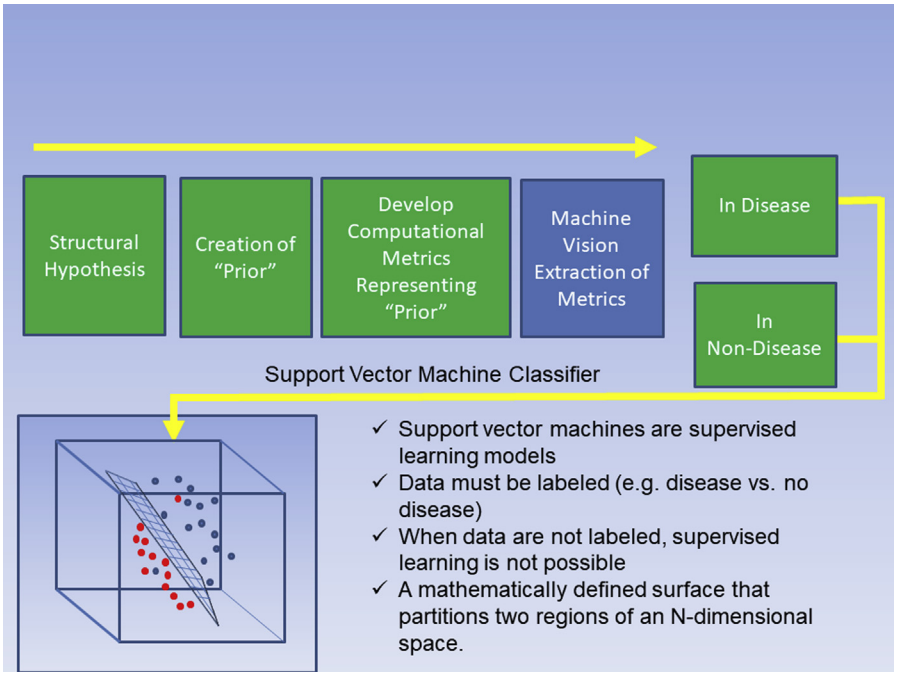

FIGURE 11.1 Traditional AI Learning Pipeline

Traditional AI machine learning in image analytics begins with the creation of a hypothesis about the structures under study. The investigators, using domain knowledge, craft their prior understanding about the structures and develop computational metrics to examine that prior understanding. Features are extracted from the image using these metrics and labeled as coming from disease vs. not-disease. A machine classifier (e.g. support vector machines) is then used to create the best hyperplane separating the classes of disease vs. non-disease.

systems will often be ineffective in capturing the granular details of a problem, and if those details are important, a traditional AI learning system may model poorly.

A *data-driven deep learning machine learning system* on the other hand can capitalize on the capture of fine details of a system, but it may not illuminate an understanding of the big picture of the problem. Data-driven models are sometimes characterized as "black box" learning systems which produce classifications or transformed representations of real-world data but without an explanation of the factors that influence the decisions of the learning system. Traditional and deep learning models are compared in Table 11.1.

Data-driven deep learning AI approaches have limited human−machine interactions constrained to a short training period from human annotated data and human verification of the classifier output of the learning system. In contrast, in traditional AI learning systems, human experts can provide actionable insights and bring these

Table 11.1 A comparison of traditional artificial intelligence (AI) versus data-driven deep learning AI.

	Traditional AI	Data-driven deep learning AI
Positives	Traditional AI is focused on conditioning the knowledge possessed by humans into models representing the real world.	Data-driven AI involves abstracting knowledge of a target domain simply by observing a large number of examples from the domain.
	Human experts can provide actionable insights and may bring these rich understandings in the form of a "prior" understanding, which can function as an advanced starting point for an AI system.	Data-driven models capitalize on capturing the fine details from the data samples including important latent variables.
Negatives	Traditional AI, with its focus on building generic models, may be ineffective in capturing fine-grained details of the problem from important unrecognized features.	Human interaction with data-driven AI models is often limited to training or verifying predicted outcomes.
		Data-driven AI may function as a black box and sacrifice an understanding of the factors which influence decisions.

rich understandings to the learning system in the form of a "prior" understanding of the domain. A prior can function as an advanced starting point for a deep learning AI system. The broad understanding of the world that humans possess with their reasoning and inferencing abilities, efficiency in learning, and the ability to transfer knowledge gained from one context to other domains is not very well understood. Framing data-driven deep learning systems with the human understanding of "what is" offers a way forward for creating partnerships between HI and AI in advanced learning systems. There is need for "explainable AI (XAI)" which can explain the inferences, conclusions, and decision processes of learning systems. There is much work that needs to be done to bridge the gap between machine and HI.

The need for human intelligence–artificial intelligence partnerships

Pathologists, confronted with a future filled with massive and ever-expanding computing power, properly ask the question "…will a machine diagnostician replace a human diagnostician and if so when?". No one knows the answer to this question but a glimpse into the world of chess offers some useful guideposts.

The chess grandmaster Garry Kasparov describes his years as a grandmaster as a time when computational chess was quickly developing and challenging human player expertise [5]. In the 1970s and 80s, grandmasters successfully outplayed the computers of the day. In the 90s, the competitions were more matched. Kasparov

noted that the things that computers do well are often the tasks where humans are weak and that the converse is also true. This concept caused Kasparov to ask whether the goal of computational chess play was to pit man against computer or rather was it to play the highest level of chess possible as a partnership of man and machine? Kasparov went on to observe a chess playing experiment in which the types of machine–human pairings varied. He saw that a weak human player paired with a machine with a better process was superior to a strong computer alone, and remarkably, also superior to a strong human player working with a machine which had an inferior process.

In the evolving world of computational pathology, the fear of artificial machine diagnostic intelligence displacing HI must be balanced with the opportunity of providing patients with more precise and accurate diagnostic, prognostic, and theranostic statements. What if, as in Kasparov's chess playing experience, that in computational pathology, a partnership of weak human diagnosticians + machines + better processes were better than strong human diagnosticians + machines + worse processes? The field of computational pathology would be immeasurably extended in its global availability and somewhat more democratized in its practice. The key to next generation diagnostics may, in fact, be getting machines and humans to interact using the best processes. I would opine that this system approach is native to the practice of pathology and laboratory medicine and that our specialty is the best positioned in all of medicine to explore the AI revolution by developing robust HI–AI partnerships in the domains of diagnostics, prognostics, and theranostics. Dr. Scott Doyle a computational pathology investigator at the University at Buffalo, State University of New York opines that "The true value of human diagnosticians is not that they are valuable because of their ability to do poorly what artificial intelligence robots can do easily. In fact, they are valuable because of all the things they can do that robots cannot do such as disparate data integration, ingenuity, serendipity, learning, experimentation, and value judgments. These are all central parts of the practice of pathology and laboratory medicine. In the end pathologists should/will focus on these things and leave the robotic tasks to the robots" (unpublished communication).

Human transparent machine learning approaches

Human transparent machine learning systems seek to have humans understand AI cognition. This means that humans must be able to understand the dynamics of the machine learning system. In human transparent machine learning, the machine learning system may have human-like sensing inputs. Most importantly, humans are able to trust the AI output of human transparent learning systems. In the computational pathology domain, this means that the outputs of machine learning systems must be vetted against the big 360′ view of the complete clinical context. In this construct, the digital assistance which results from any computational pathology systems is used only after cross referencing them with all of the meaning understood

by human sensors. These human understandings are gleaned from the history, physical examination, clinical laboratory, radiology, surgical pathology, cytopathology, molecular pathology, and quality assurance information systems. In other words, pathologists will need to learn how to work with machine learning systems as senior partners who understand the big picture of patient care diagnostics.

Explainable artificial intelligence

The concept of XAI is that machine learning is understood by human operators and that through this understanding, a bilateral trust relationship is established between humans and machines. XAI contrasts sharply with the "black box criticism" of deep learning. XAI is very important when machine learning systems impact social systems. As diagnostics is a central function of medicine, which in turn impacts all aspects of the societal good of health and wellness, XAI is a particularly desirable form of AI in computational pathology. Regulators of clinical laboratories and the diagnostic process should have a strong interest in XAI as it offers the transparency and real-world rationales which make algorithmic outputs testable and subject to quality assurance processes. XAI guards against AI "cheating," in which a learning system tracks surrogate labels but is not truly informative. For example, a data-driven deep learning AI system that used histological images tagged with ICD10 codes could never be relied on to distinguish cancer from noncancer using pixel data from digital virtual slides. Or even more simply, if the slides with cancer in a training dataset had been marked with a red dot by a previewer and the machine learning system learned to look for a "red dot" to classify the dataset into cancer versus not-cancer, the resulting classifier might be 100% effective in the training data but 0% extensible to other test datasets.

Deep learning systems may be adapted to XAI through a technique called layer-wise relevance propagation, which determines the features in an input vector which contribute most strongly to a neural network's output [6]. Exposing the output layers or even the early activation layers of neural networks to human observers for error correction or other annotations may also be used to support XAI (Fig. 11.2).

Cognitive artificial intelligence

Cognitive computing seeks to simulate the functioning of the human brain with software and hardware which senses, responds to stimulus, and reasons. Cognitive AI systems stimulate human thought in finding solutions to broad and complex problems. Cognitive AI systems should be adaptive, interactive, iterative, and contextual. Cognitive learning systems are adaptive in that they learn in real time or near real time and they learn as information changes. Cognitive systems are interactive with users. Cognitive systems help define problems by asking multiple questions or finding additional sources of information to make the statement of a problem less ambiguous. Cognitive learning systems understand, identify, and extract contextual elements such as meaning, syntax, time, location, appropriate domain,

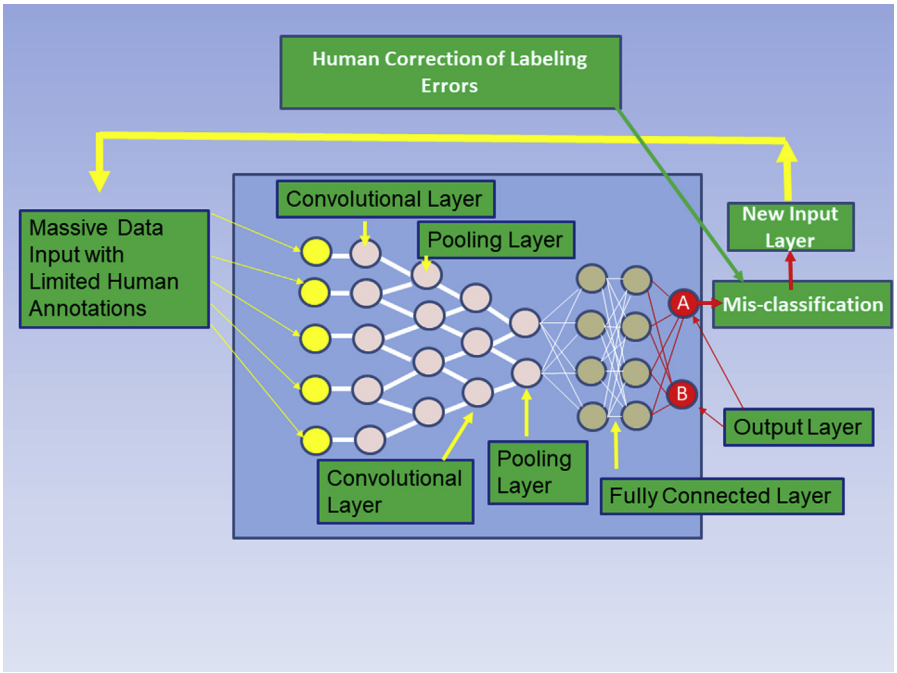

FIGURE 11.2 Opportunities for Human Interaction with Deep Learning Systems

In data driven deep learning systems the output classifications of images are derived from numerous calculations within "hidden" layers often without an explanation of the factors which influence the decisions of the learning system. Deep learning AI systems such as this convolutional neural network input large amounts of data with limited human labels. Convolutional and pooling layers may feed fully connected layers which in turn drive an output classifier. Errors in the output classifier are opportunities for human intelligence to correct the system and feed these new labels back into the input of the system. Other opportunities for human interaction with the learning system may exist in the human understanding and annotation of the activation layers of the convolutional neural network.

regulations, user's profile, process, task, and goals. Cognitive systems utilize multiple types of input data including human perceived sensory data. Cognitive AI self-learning systems weight context and conflicting evidence by using tools such as data mining, pattern recognition, and natural language processing. Cognitive AI systems in healthcare are thought of by some to be a tool for physicians to bridge the machine–doctor–patient chasm [7]. An example of cognitive computing is IBM Watson for Oncology, which has been used to derive ranked options for treatment of cancer.

Human in the loop

Human-in-the-loop learning models require human interaction. Human-in-the-loop systems allow humans to change the output of the learning systems. Human-in-the-

loop simulators always have human input as part of the simulation, and humans influence the outcomes of the simulation exercise such that the outcomes may not be exactly reproducible. Human-in-the-loop simulations, however, allow for the identification of model shortfalls which may not be apparent before testing in a real-world setting. Flight simulators may be an example of human-in-the-loop learning. Human in the loop approaches have been applied to histopathology. Lutnick et al. [8] used a human-in-the-loop strategy for data annotation and the display of neural network predictions within a commonly used digital pathology whole slide viewer. This strategy reduced the annotation burden in that the segmentation of human and mouse renal microcompartments was repeatedly improved when humans interacted with automatically generated annotations throughout the training process.

One-shot learning

Humans have the capacity to learn object classifications from limited training examples or even no examples. Humans are able to make use of information from previously learned classes to learn new ones. In a one-shot learning system, a machine uses prior learning to classify objects newly presented to it. One-shot learning emphasizes knowledge transfer from prior learning to new circumstances and as such is an approach which resonates with human learners. Knowledge transfer could be of model parameters learnt by training on multiple examples on one class and then applying them to a new object class. Knowledge transfer may be done by sharing features across classes. Mutual information among learned objects could be applied to learning a new class of object. Knowledge transfer may also occur through the transfer of context. For example, knowing the context of camera geometry used to classify one object may help classify a new object.

The computational approaches described above can be applied to any laboratory data used in the clinical setting, and these methods are opportunities to innovate in all of the disciplines in pathology and laboratory medicine. Cabitza et al. [9] outlined the opportunities of machine learning in laboratory medicine, and these authors anticipate a flood of applications in the near future. In the following sections, however, we will mostly focus on the use cases of computational pathology in cellular and tissue-based image data.

Image-based computational pathology

Core premise of image analytics: what is a high-resolution image?

A foundational hypothesis supporting the concept of cell and tissue image–based computational pathology is that a high-resolution image is *a self-organizing set of data* which uniquely represents all of the genes, and all of the molecules, and all of the cells in that scene at one point in time. As a self-organizing piece of data, the image can only uniquely have the phenotype which it presents. An *image is*

what it is for very specific reasons. Those reasons are the relationships among the genomics, epigenomics, proteomics, metabolomics, and all the "omics" which go into making that image. A high-resolution image is a window onto the *relationships* among all of the genes, and all of the molecules, and all of the cells in the scene at one particular point in time.

The targets of image-based calculations

The targets for algorithmic computation in image-based pathology include (1) the pixels in the image; (2) cells and organelles; (3) tissues and tissue elements and the complex relationships among pixels and cellular/tissue elements. Each of these targets may be amenable to many image analytic approaches, but some approaches are better suited for one or another of the targets. Machine learning systems driving decisions can be applied to the data extracted from any of these targets. The following are some examples of computational schemes for examining each of these targets.

Pixels: Pixels are the basic building blocks of digital images. They are a rich source of information in the color and intensity domains. Image texture features are metrics calculated from pixels that contain information about colors or intensities. Texture features can be viewed across an entire slide and leveraged as input data for machine learning systems, or they may be sorted into feature representations of cellular and tissue objects (as below). Thousands of texture features can be calculated from a histopathological image. The features can be extracted in cost-efficient unsupervised ways. Texture features data can be approached statistically. For example, an edge detection algorithm could examine the gradient magnitude and the gradient direction of the pixels and yield an "edgeness" metric for an entire image or segments of an image.
Cells and organelle segmentation: Cells and organelles are obvious biological units of interest to pathologists. Gathering quantitative information about cells is central to a computational pathologist's modern exploration of Virchow's cellular theory of disease. Segmentation of cells, nuclei, membranes, and organelles from image metrics can be accomplished by a variety of whole scene segmentation approaches, some of which include texture features, size and shape features, densitometric features, and chromatin-specific features [10].
Tissue and tissue elements and the complex relationships among pixels and cellular/tissue elements: The grand challenge for analytic approaches to computational pathology is to capture the image embedded information which is informative in modeling complex disease systems such as disease progression and the predictive modeling supporting therapeutic choices. Much of this information is embedded in the spatial relationships between cells. Spatial arrangement features can either be explicitly harvested or discovered with deep learning systems. Graph embedding theory is an example of a method which is well-adapted for exploring the relationships among pixels and cellular and tissue elements in high-resolution histopathological image data [10]. Graphs

can be constructed for modeling different states of tissue and to distinguish these different states from others by calculating metrics on the graphs and classifying their values.

First fruits of computational pathology: *the evolving digital assistant*

The concepts of AI—HI partnerships and human transparent machine learning approaches raise an aspiration for a *digital assistant* to help pathologists in their daily work. Such an assistant would reduce repetitive tasks which are subject to error and free up pathologist time for complex medical decision-making. A *digital assistant* would reduce the organization and triage work, search and identify functions, and make data quantification (particularly from images) more accessible. The *digital assistant* would generally help pathologists to practice at "the top of their licenses." This vision does not yet exist in the reality of current state clinical practice, but it is on the way. The following are just some of the anticipated facilitating functions projected from the research literature into which the *digital assistant* may evolve.

The digital assistant for quality control

All laboratories have robust and active quality control and quality assurance at the core of their workflow. In modern-day clinical chemistry and hematology laboratories, for example, the information transfer about individual or batch value comparisons to both internal and external reference values is continuous and real time. There is a conversation between instruments and the laboratory professional human operators, which is always engaged and bidirectional. Heuristics are used to flag significant deviations from acceptable norms.

In the current state, anatomical pathology, laboratory quality control, and quality assurance are much more qualitative and discontinuous. As anatomical pathology moves to a digital pathology workflow, the need for precise quantitative quality control and quality assurance systems is becoming apparent. Computational pathology algorithms, particularly in unsupervised architectures, are susceptible to interpreting the noise of artifacts as true biological signal. Quality parameters which matter can be subtle and easily overlooked. Artifacts can adversely affect the performance of computational pathology machine learning systems [11,12]. In an integrated digital pathology workflow, using WSI, continuous, real-time, and precise quality image metrics are needed to understand and normalize the preanalytic variance inherent in glass slide preparation.

Lab accreditation organizations are keenly aware of the need to address quality assurance in digital and computational pathology. The College of American Pathologists has recently performed a metaanalysis of the literature and published guidelines on the Quantitative Image Analysis (QIA) of Human Epidermal Growth Factor Receptor 2 Immunohistochemistry for Breast Cancer [13]. They recommend that "to

improve accurate, precise, and reproducible interpretation of HER2 IHC results for breast cancer, QIA and procedures must be validated before implementation, followed by regular maintenance and ongoing evaluation of quality control and quality assurance. HER2 QIA performance, interpretation, and reporting should be supervised by pathologists with expertise in QIA."

Vendors of whole slide scanning devices have also recognized that the quality of the rendered whole slide images impacts on the ability of pathologists to diagnose. When scanning the same slide on the same scanner at different times, the WSI appearance may contain subtle differences even when viewed on the same display because of variances in the scanner characteristics and/or other factors such as temperature or mechanical shifts.

Color reproduction has been approached by commercial developers through the construction of standardized slide phantoms matching the colors in the scanned images to the actual slide color [14]. The colors observed in phantoms are, however, not the same as those viewed in tissue matrix. Alternatively, commercial systems have used quantitative image quality assessment based on parameters such as sharpness, contrast, brightness, uniform illumination, and color separation to evaluate the scanner performance. In a study on HercepTest slides [15], the authors note that this computational approach is independent of image content or test specimen.

Open source approaches to the development of the digital assistant for quality assurance (QA) processes in computational pathology are also evolving. Janowczyk et al. have recently proposed an open source quality control tool for digital pathology slides which they call "HistoQC" [11]. Their system offers a modular QC application, which employs a suite of image metrics (such as color histograms, brightness, and contrast), images features detectors (for example, edge and smoothness), and supervised classifier systems (i.e., a drawing tablet with pen annotation) to facilitate QA work. The user interacts with the system through a python-based pipeline. Relevant output images of artifacts are created and presented with metadata and then sent to a data analytic tool. These authors have developed an HTML-5 user interface that allows for real-time visualization.

The digital assistant for histological object segmentation

The segmentation of histological objects by computational methods is an active area of research in pathology. The following describes some of the ways in which a variety of approaches have been employed to support clinically relevant image interpretation in pathology.

Nuclei

The enumeration and categorization of nuclei and nuclear features is a central task in many types of histopathological diagnoses. Nuclear detection methods are numerous. Xu et al. in a paper describing a deep learning strategy compiling examples of 23 different handcrafted feature-based approaches include color-based, edge-based, contextual information—based, and texture-based designs for nuclear

segmentation [16]. Nuclear shape and orientation as measured predominantly by tensor information have been used in tissue microarray images to model nuclear pleomorphism and to be independently predictive of poorer survival in breast cancer [17]. The challenge with most of these approaches is in their ability to generalize across multiinstitutional datasets encompassing the loud "noise" of the preanalytic variance which is resident in slide preparation and scanning. Recently, deep learning techniques have been applied to the problem of nuclei segmentation, feature extraction, and classification. Breast cancer nuclei classification in histopathology has been examined by comparing several supervised learning methods with deep learning systems. The deep learning systems all outperformed the supervised machine classifier algorithms [18]. Other examples of deep learning strategies applied to nuclei detection cancer images have included (1) stacked sparse autoencoders used to learn high-level features from just pixel data of nuclei in breast cancer [16]; (2) a deep contour aware network (DCAN) using multilevel contextual features in a fully CNN with an auxiliary supervision scheme in the segmentation of nuclei from glioblastoma [19]; and (3) the use of cycle generative adversarial networks (cycleGANs) deep learning to improve multiorgan nuclei segmentation to almost 95% [20].

Mitoses

The recognition, enumeration, and classification of mitotic figures in histopathological images are central tasks in many histopathological grading systems. Pathologists often use the data from multiple z planes of focus in combination to distinguish mitotic figures from other condensed chromatin structures. As WSI often does not provide data from multiple z planes, the detection of mitotic figures in standard 2D whole slide images has become a high profile contest among investigative groups in digital pathology. Many competitive grand challenges have compared the effectiveness among a variety of computational approaches to the detection of mitoses. The winner of one of these grand challenges captioned in the paper by Ciresan [21] used a patch approach at 40× resolution with deep neural network learning. Janowczyk and Madabhushi [22] have adapted this approach, used a smaller patch size at 20× resolution, and leveraged segmentation of region of interests (ROIs) using a blue ratio segmentation protocol to enrich for ROIs containing mitotic figures. Wang et al. [23] have sought to address mitosis detection by presenting a convergent approach to combine domain-inspired features with deep learning features to detect mitoses. They showed that this integrated approach yielded superior detection accuracy compared to deep learning or handcrafted feature-based strategies alone.

The recognition, enumeration, and classification of mitotic figures are likely to be important tasks in diagnostic histopathology which will require the partnership of AI and HI to bring an efficient and safe computational tool to the clinical diagnostic market.

Tumor-infiltrating lymphocytes

In a previous chapter, Saltz and colleagues provided a detailed rationale for the importance of tumor-infiltrating lymphocytes (TILs) and gave examples involving

the classification characterization of functionally relevant subpopulations of these lymphocytes and their positive and/or negative roles. For the reasons described, the detection, enumeration, and relationships among TILs are topics of great interest, particularly, for investigators interested in immuno-oncology. Klauschen et al. [24] have recently reviewed many computational approaches, which have been used to analyze TILs. TILs have been most commonly enumerated using semiquantitative scoring systems. Standardized TIL scoring has been used to good effect in clinical trials of breast cancer. Scoring systems, however, are subject to interobserver variance. The precise and time-efficient quantification of TILs is a needed tool for both research and clinical work. Computational approaches to the tasks of TIL characterization are required to enable the conduct of precision medicine clinical trials in oncology and autoimmune disease.

Traditional image analysis approaches to TIL identification rely on supervised or semisupervised systems. Through the identification of edges, cellular objects of interest are segmented and features are extracted from these objects which allow for the classification into different cell types. Thresholding, watershed, level set, color-space clustering, and morphological shape—based algorithms have all been used for the identification of TILs. These computations often require a priori knowledge of lymphocyte morphological variance, but their workings are explainable to the investigators. Relatively small datasets are used to train these handcrafted systems. Tuning the algorithmic parameters of these computational systems to handle the many cell types which need to be separated from lymphocytes, however, can be a daunting task.

Deep learning has recently been used in the detection of TILs. Janowczyk and Madabhushi [22] have used CNNs and patch classification schemes to identify lymphocytes in standard H&E histology. Larger training sets are needed to subsume the morphological variance of scenes and preparative variance, but the deep learning system in essence tunes itself. The exact nature and biological significance of the learned features may, however, be opaque. The understanding of the immune subtypes resident in host response networks may be critical to choosing effective immune therapies. Other studies have employed deep learning methods using different patch classification, boundary detection, and segmentation algorithms on tissue sections labeled immunohistochemically for lymphocytes and lymphoid subsets. The combination of molecular labels with deep learning systems offers the potential for the understanding of biologically explainable networks. Such methods could be adapted to assays designed for the examination of clinically relevant immune networks such as the relationships of immune checkpoint inhibitor molecules on host response and tumor cells.

Glands and acini

The histopathological definition of "gland" varies with the organ. Morphological deviation from normal gland structure is used by pathologists to define reactive conditions, dysplastic precancer, cancer, and cancer grade. Obviously, the segmentation of glandular objects is the first step when considering the computational analysis diseases which are defined by gland disorganization.

Gland segmentation in colon histopathology has a significant literature. The gland segmentation in colon histology images (GlaS) grand challenge [25] in computational pathology asked participants from the computer vision and medical imaging research communities to develop and compare gland segmentation algorithms of benign colonic glands and colon cancer glands. Accuracy of the detection of individual glands; volume-based accuracy of the segmentation of individual glands; and the boundary-based similarity between glands and their corresponding segmentations were the metrics for comparisons. Gland detection accuracy was performed using the F1 score ($2\times$ precision $\times$ recall/precision $+$ recall). Volume-based segmentation accuracy was evaluated using the dice index which is a measure of agreement or similarity between two sets of samples. In this competition, these sample sets included a set of pixels corresponding to a ground truth object (G) and a set of pixels belonging to a segmented object (S). Boundary-based segmentation accuracy between ground truth (G) and segmented object (S) was evaluated using object-level Hausdorff distances. In general, two broad approaches to gland segmentation were employed by the competitors. The first approach was to start by identifying pixels belonging to glands which were then grouped to form a separate spatial glandular object. The second approach began with candidate objects which were then classified as glands or nonglands. All of the methods which were based on CNNs used the first approach. Only one method followed the second approach whereby candidate objects forming part of the gland, namely lumens or epithelial boundaries, were identified first and then classified into different types followed by full gland segmentation. Overall, the pixel-based CNN entries had the strongest performances. A novel DCAN [19] (Chen et al., read also below) was the overall winner of the challenge.

The segmentation of prostatic acini is a problem analogous to that of colonic gland identification. Traditional handcrafted image analysis techniques begin with a prior concept of acinar structure. Monaco et al. [26] conceived of a prostatic acinus as a lumen surrounded by a collar of epithelium. To distinguish between benign prostatic glands and prostatic adenocarcinoma, a Markov iteration was used to group and separate microacini from benign glands. Separation of Gleason patterns 3 versus 4 was facilitated by this schema. Singh et al. [27] used a combination of pixel and object-level classifiers which incorporated local and spatial information for consolidating pixel-level classification results into object-level segmentation of acini. As alluded to above, Chen et al. [19] have described a DCAN which provides a unified multitask learning framework examining contextual features by using a fully convolutional network. The fully convolutional network takes an image as input and outputs a probability map in one forward propagation. In this system, object features (for example, texture and color) are integrated with contour information. The result is a system which addresses the difficult problem of separating touching objects. The system is extensible to many types of objects including benign and malignant colonic glands and nuclei.

The digital assistant in immunohistochemistry

The in situ IHC analysis of tissues for molecular markers is central to the moderate practice of pathology. IHC biomarkers are used for tissue classification tasks, as grading tools and as predictive markers. The qualitative/semiquantitative analysis of IHC markers can have significant gaps in precision (TP/TP + FP) and recall (TP/TP + FN). Computational methods have been explored for many years as possible tools for improving the operating characteristics of IHC tests. Gavrielides et al. [28] examined the use of computationally assisted image evaluation in the scoring of HER-2 IHC. These authors found that a computer-assisted mode which provided a HER-2 reference image along with a corresponding feature plot of membrane staining intensities and membrane staining completeness improved both interobserver and intraobserver agreement when scoring HER-2.

Machine learning systems are being developed to analyze IHC labels. Chang et al. [29] have addressed image cytometry multiplex IHC data by comparing manual gating of lymphoid subsets to clustering algorithms and sparse representations to yield biologically interpretable subset populations.

Vandenberghe et al. [30] compared handcrafted features and machine classifiers (SVMs and random forest) to a deep learning CNN (ConvNets) in the analysis of HER-2/neu staining in breast cancer. Overall the accuracy of the neural network deep learning methods was somewhat better than the handcrafted feature classifiers. Interestingly, none of the methods were particularly good in their accuracy of scoring 2+ cells, where interobserver variance and clinical significance are known to be great.

Khosravi et al. [31] have compared six different deep learning systems using transfer learning for the scoring of IHC biomarkers in bladder and breast cancer. Transfer learning is the pretraining of a network architecture on a very large dataset and the use of that trained model for new classification tools for a data asset with limited size. In this study, the transfer learning strategies included pretraining of the network as a feature extractor and the fine-tuning of a pretrained network. The accuracy retrieval curves for the prediction of biomarker score ranged from 72% to 99%.

The digital assistant in tissue classification

The identification of cancer and its separation from reactive conditions are fundamental high-order functions of histopathologists. The two sequential activities of the pathologist workflow in executing these functions are the identification of candidate regions of interest and the subsequent classification of those regions into meaningful clinicopathological classes. These processes are true in any subspecialty of oncological surgical pathology. Breast cancer has been of particular interest to investigators studying computational pathology approaches to cancer diagnosis.

Fondon et al. [32] reviewed the use of traditional machine learning in breast cancer diagnosis. These authors approached the problems of automatically identifying and classifying breast histopathological digital images into one of the four classes: normal, benign lesion, in situ carcinoma, and invasive carcinoma. The functions of preprocessing, feature extraction from nuclei, regions, textures, and the use of these features to drive a variety of machine classifier tools are described. The SVM classifiers were most efficient with accuracy of up to 76%.

Deep learning approaches to breast cancer diagnosis using CNNs are being vigorously explored. Araujo et al. [33] have used CNNs on digital images from the Bioimaging 2015 breast histology classification challenge to separate into the same four classes of normal, benign lesion, in situ carcinoma, and invasive carcinoma. The overall sensitivity of the CNN in finding the carcinoma classes was 80%.

Cruz-Roa et al. [34] have used CNN deep learning to automatically identify invasive breast cancer on whole slide images from the Cancer Genome Atlas and from whole slide images collected from three contributing institutions. They found a positive predictive value of 72% and a negative predictive value of 97% for identifying invasive breast cancer. CNN classifiers outperformed visual feature classifiers (color and intensity, color histograms, shape index histograms, Haralick features, and graph-based features) in this study. Bejnordi et al. [35] have extended the CNN approach on whole slide breast images by using context-aware stacked CNNs for a similar task of separation of the classes normal/benign, ductal carcinoma in situ, and invasive carcinoma. In their "context-aware" approach, these authors first trained a CNN using high-pixel resolution information to classify the tissue into different classes. To incorporate more context, they fed much larger patches to this model at the time of testing. The output of this first system is then the input to a second stacked system, which uses the compact informative representations of the first model together with the information of the surrounding context to learn global independence of structures in different lesion categories. In this 3 class task, they achieved an accuracy of 0.812.

The digital assistant in finding metastases

The tasks of searching for and identifying cancer metastases in lymph nodes are critical but time-consuming work for diagnostic pathologists. In breast cancer, the definition of nodal status by histopathology is ground truth for the assignment of nodal status. Current clinical guidelines codify the sizing of lymph node metastases into pN 0 (i-): No regional lymph node metastases histologically, negative IHC; pN0 (i-): Malignant cells in regional lymph node(s) no greater than 0.2 mm and no more than 200 cells (detected by H&E or IHC including isolated tumor cells [ITCs]); pN1mi: Micrometastases (greater than 0.2 mm and/or more than 200 cells, but none greater than 2.0 mm); and pN1 macrometastases at least 1 tumor deposit greater than 2.0 mm. Steiner et al. [36] have used a scoring system for gauging how difficult it is for a pathologist to distinguish among these classes which they term "obviousness scores." In their study, it was found, not surprisingly, that the

obviousness score is high for macrometastases and low for identifying isolated tumor cells. These authors also examined the impact of deep learning algorithmic assistance, which they named the Lymph Node Assistant (LYNA), on the obviousness score. LYNA improved the obviousness score for micrometastasis detection. The impact on assignment of the other classes is was less apparent. The average review time was also decreased for micrometastasis detection with LYNA, marginally improved for a negative class assignment, and not significantly impacted for the assignment of macrometastases or isolated tumor cells.

A recent competition (CAMELYON16) [37] to develop algorithmic solutions for the automated detection of sentinel lymph node metastases had 32 algorithms submitted. The areas under the curves (AUCs) for the receiver operating characteristic curves ranged from 0.556 to 0.994. The top five algorithms had a mean AUC comparable to the study pathologists' interpretations in the absence of time constraints.

Such studies offer the real possibility that the algorithmic digital assistant may provide a valuable prescreening function in histopathology in future diagnostic workflows.

The digital assistant in predictive modeling and precision medicine

Apart from substantially aiding the pathologists in decision-making, the use of computational imaging tools could enable the creation of digital imaging—based companion diagnostic assays that would allow for improved disease risk characterization. Unlike expensive molecular-based assays that involve destroying the tissue and invariably capture genomic or proteomic measurements from a small part of the tumor, these digital imaging—based companion diagnostic tests could be offered for a fraction of the price and could enable characterization of disease heterogeneity across the entire landscape of the tissue section. Both traditional AI and deep learning approaches for the modeling of patient outcomes and response to therapy have been explored in a variety of neoplastic and nonneoplastic conditions.

Lu et al. [17] have studied the prediction of survival in early stage estrogen-positive breast cancers using a traditional AI approach. These authors examined tissue microarrays for 615 features related to nuclear shape, texture, and orientation. They identified the top 15 quantitative history morphometric features which in combination modeled, using a linear discriminate analysis classification, the probability of long-term versus short-term disease-specific survival. Features relating to the heterogeneity of nuclear orientation dominated in this model. The authors in multivariate analysis controlled for tumor stage found that their model was strongly independently predictive of patient's survival in estrogen receptor (ER)—positive and lymph node—negative breast carcinomas.

Bychkov et al. [38] took a deep learning approach to the prognostic modeling of colon cancer outcomes from standard histopathology. These authors used an image analysis pipeline which included long short-term memory (LSTM) networks that allow the learning system to detect and memorize image tiles, which encode relevant and contributory morphologic information and disregard irrelevant image titles.

Through this architecture, the authors attempted to find units which performed biologically meaningful discriminations of tissue patterns. They hypothesized that if the system observed tiles which contained information regarding disease, then some memory cells within the network will learn this pattern, aggregate it into memory, and propagate the understanding through the network. The authors used LSTM networks to examine tissue microarrays of colorectal carcinoma. They found that the LSTM model provided a strong hazard ratio for disease prognosis as compared to visual scoring, but that their LSTM system did not outperform the Dukes staging system for colorectal carcinoma prognosis.

Deep learning classifiers have also been used in predicting outcomes of nonneoplastic conditions. For example, in the arena of cardiac pathology, Nirschl et al. [39] used deep CNNs to predict clinical heart failure from H&E-stained whole slide images with 99% sensitivity and 94% specificity.

The clinical pathology laboratories of health systems offer robust data basis of well-organized quantitative data representing the chronological patient record. Cabitza and Banfi [9] review multiple applications of machine learning to clinical laboratory data for the diagnosis, prognosis, and predictive modeling of a variety of conditions. These authors predict a deluge of investigations using clinical laboratory data and machine learning in the near future.

The digital assistant for anatomical simulation learning

Just as the digital revolution has required a robust infrastructure, so also does the learning from and teaching of complex 3D subjects like anatomy. Pedagogically relevant programs to develop, test, and scale a data-intensive, bandwidth-heavy, digitally enhanced, collection and visualization systems for anatomical data are critical to UME, GME, and CME training in radiology, surgery, and pathology. Simulation learning depends on computational approaches to the 3D structure of human gifts, large animal models, radiological images, and pathological specimens. 3D Slicer (http://www.slicer.org) is an open source software platform for the registration, interactive segmentation, visualization, and volume rendering of medical images and for research in image-guided therapy [40]. This sort of computational analysis of structure can integrate biological and digital phantoms within the pedagogic experience. This juxtaposition can allow for robust surgical simulation programs, which integrates the use of artificial phantoms, digital phantoms, and biological phantoms in preclinical surgical investigation, teaching, and training. This combination of the core activities of teaching and investigation in the structural sciences can offer a laboratory where students can learn in the new environment of quantitative structural science.

The digital assistant for image-omics data fusion

When trying to understand the cell biology systems which underpin all pathological interpretations, the data from different modes of examination are best not considered

in isolation. Rather, these data are streams (or perhaps torrents) which integrate and literally "fuse" into robust representations of these complex cell and tissue systems. Both image and molecular 'omic (genomic, proteomic, metabolomics, etc.) data are dense and heterogeneous. Image data are strong at probing the relationships among features. Reductionist 'omic data are strong at probing causal pathways. The frameworks in which image data and 'omic data are represented are often quite different. We know digital image data as pixels features and values and 'omic data as molecular structures, molar quantities, and frequencies. It is not intuitively evident that these very different modes of data can live in the same space. Computational pathology and AI approaches, however, now allow for the quantitative fusion of these diverse cross-modal data streams in support of the modeling of complex biological systems.

There are several approaches to combining data. One is that of combining interpretations. In this method, there is an aggregation of the decisions made from classifiers. Each classifier corresponds to one particular data stream. This is quite common in biology and medicine. A combination of interpretations may, however, not be able to leverage all the power of the combined data streams because information is often lost in the process of converting from feature vectors to class labels.

A concatenation of high-dimensionality feature vectors is an inclusive approach, which subsumes all the available features; however, this approach invokes the "curse of multidimensionality," whereby the excess of features compared to patients may result in a classifier system which is overfit. Dimensionality reduction approaches such as principle component analysis (PCA) project feature vectors into a unified low-dimensionality eigenvector space. PCA assumes that the system data are contained within a linear space. Most biological systems, however, are nonlinear. Nonlinear dimensionality reduction approaches can be of assistance in addressing such complex systems. There are numerous algorithms for manifold learning and nonlinear dimensionality reduction. Some of these algorithms simply function as visualization tools, while others map data from high-dimensionality space to low-dimensionality spaces. Data may be combined by creating transformed representations of each data stream before combining the transformed representations. Data streams are projected (embedded) into a homogeneous metaspace, where all data are represented at the same scale. Embeddings may be combined in many ways. For example, one strategy is to combine embeddings by using high-dimensionality kernels. Kernels are dot product representations of each modality, which can be combined to create a fused representation of heterogeneous data. Using another method, Lee et al.[41] examined the effectiveness of fused data to predict biochemical recurrent of prostate cancer after radical prostatectomy. These authors employed a novel technique which they termed supervised multiview canonical correlation analysis (sMVCCA). Canonical correlation analysis (CCA) is a multivariate statistical method which seeks to find a linear subspace in which correlation between two sets of variables is maximized. Using sMVCCA, Lee et al. created an integrated feature vector composed of quantitative histopathological features and proteomic features obtained from tandem mass spectrometry of prostate

cancer from radical prostatectomies. Kaplan–Meier analysis showed improved biochemical recurrence free survival prediction using the sMVCCA fused data classifier as compared to histology or proteomic features alone.

In another example, Savage and Yuan [42] presented a new tool for selecting informative features from heterogeneous data types and predicting treatment response and prognosis which they have named FusionGP. This is a Bayesian nonparametric method for integrating multiple data types. The relationships between input features and outcomes are modeled through a set of (unknown) latent functions. The latent functions are constrained using a set of Gaussian priors, and sparse feature selection is applied. The most strongly selected molecular features were evaluated in gene ontology for dominant core processes at work. These authors examined a cohort of 119 ER-negative and 345 ER-positive breast cancers to predict two important clinical outcomes: death and chemoinsensitivity by combining gene expression, copy number alteration, and digital pathology image data. For the prediction of disease-specific death, the molecular data were most informative for the ER-negative tumors. Interestingly, image features outperformed the molecular data in ER-positive tumors.

The digital assistant of the future might present the diagnostician with a variety of data fusion options to assist the HI–AI partnership team seeks in identifying the best precision diagnostic, prognostic, and predictive statements for a given patient.

Artificial intelligence and regulatory challenges

The adoption of computational methods into the diagnostic workflow will require an aligned combination of technology, economic value, and regulatory oversight to insure a safe process. The FDA employs a risk-based paradigm for medical device classification. “Premarket approval” (PMA) is required for Class III devices, which have no comparison to a predicate. A “510(K) Premarket notification” is a path to market for new devices which are substantially equivalent to predicate devices. In certifying image-based pathology systems, the FDA has broad experience in the regulation of image-based instrumentation driven by algorithms. The clinical use of algorithms in image-based assays in pathology already exists in several domains including automated hematology analyzers, chromosome analysis, FISH enumeration systems, urine sediment analysis, gynecological cytology, and IHC for predictive markers such as HER2/neu, ER, and progesterone receptor (PR).

When considering the regulation of market entry for deep learning AI systems in image-based pathology, the new regulatory challenge is to provide access to these powerful systems and yet constrain their application to use cases, where patient safety can be assured. Clearly, new approaches to the regulatory evaluation of machine learning systems are required.

The understanding of the relationship between digital image data and the human perception of image data in pathology is one challenge. Gallas and

Gravrielides [43] from the FDA are creating an evaluation program which develops information regarding which technical characteristics of an image-based system are important; learning how these technical characteristics impact pathologists' interpretations; and understanding how these technical characteristics can be measured. These authors and their colleagues have created an "evaluation environment for digital and analog pathology (eeDAP)." They anticipate that this be used as a "Clinical Outcome Assessment tool used in reader studies for A premarket submissions (PMA or 510k deNovo) to compare the accuracy or reproduce ability of pathologist evaluations of digital images on a display to those of glass slides on a microscope. The pathologist evaluations of outpatient tissue are the clinical outcomes. The accuracy or reproducibility is the clinical outcome assessment; this assessment reflects image quality." Such tools will be important in examining sources of discrepancies between pathologists for classifying different types of diseases, for developing a panel of histopathological patterns and related decision support tools for improving pathologist performance for these classification tasks, and in assessing pathologist performance with whole slide images versus traditional optical microscopy.

A second challenge is fixing on the target of regulation. Machine learning systems are endlessly changing. With each epoch, a machine learning system understands or "sees" a set of images differently and can output a different understanding of class assignments. The fundamental regulatory challenge for harnessing image-based machine learning systems as clinical decision tools is to constrain the questions asked to clinically meaningful outputs which can be tested against other forms of understanding. For example, if a machine learning system outputs an interpretation of Her2/neu IHC that seeks to classify a breast cancer as having amplification of this gene, the output of the image-based test can be validated with other modes of analysis such as PCR or FISH for Her2/neu amplification. This validation is only possible after some training and testing time of the machine learning system and during a period when the learning system is locked down and not changing. The qualification(s) of image-based machine learning systems as clinical laboratory tools are likely to address newly proposed machine learning systems in a series of locked down versions, each of which will have to be validated. Functionally, what will need to be developed are rapid validation programs with which regulatory agencies such as the FDA can quickly compare the outputs of locked down machine learning system versions against predicate testing systems.

The current literature on the use of image-based machine learning systems as clinical laboratory testing tools is primordial. It is apparent from a quick inspection that most of the studies cited above are investigational and not yet ready for use in clinical practice. The tremendous opportunities promised by machine learning technologies are, however, likely to soon propel these tools into 510(k) approval pathways. Humans in the loop architectures are likely to be deemed safer approaches for the early application of machine learning systems in pathology and laboratory medicine.

Educating machines—educating us: learning how to learn with machines

Pathological diagnostics, prognostics, and theranostic statements of cell and tissue samples are not just application spaces for AI, but they offer tremendous opportunities to learn more about HI—machine intelligence partnerships in unique ways. Researchers at the University at Buffalo, State University of New York (Doyle S, unpublished communications) are developing a novel approach to building machine learning—based systems, where classifier training is recast as a problem of pedagogy. In this approach, termed the "**AI School for Pathology**," AI agents are treated as students in a school where human agents are instructors and the learning context is that of higher order clinical phenotypes and clinical outcomes. Both AI and human agents learn the biological/medical realities and the meaning of data from tissue imaging, 'omics, electronic medical record (EMR), and their fused data products. Within this paradigm, learning researchers propose to examine various portals for AI—human agent interactions analogous to those of teachers and students in a traditional learning setting. The understandings generated from this project will help advance research in one of the key challenge areas for AI today, namely, enabling AI systems to learn as humans learn.

References

[1] Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biology 2017; 18:83. https://doi.org/10.1186/s13059-017-1215-1.
[2] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 2012:1—9.
[3] Fuchs T, Buhmann JM. Computational pathology: challenges and promises for tissue analysis. Computerized Medical Imaging and Graphics 2011;35:515—30.
[4] Website: John McCarthy Available from: https://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist). [Accessed 14 September 2019].
[5] Rasskin-Gutman D. Chess metaphors: artificial intelligence and the human mind, 2009, MIT Press, ISBN: 9780262182676.
[6] Bach S, et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 2015;10(7):e0130140.
[7] Website: Robinson S. Cognitive computing in healthcare mends doctor-patient gap. 2018. Available at: https://searchenterpriseai.techtarget.com/opinion/Cognitive-computing-in-healthcare-mends-doctor-patient-gaps. [Accessed 14 September 2019].
[8] Lutnick B, et al. An integrated iterative annotation technique for easing neural network training in medical image analysis. Nature Machine Intelligence 2019;1(2):112—9.
[9] Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? Clinical Chemistry and Laboratory Medicine 2018;56(4):516—24.
[10] Gurcan MN, et al. Histopathological image analysis: a review. IEEE reviews in biomedical engineering 2009;2:147—71.
[11] Janowczyk A, et al. HistoQC: an open-source quality control tool for digital pathology slides. JCO Clinical Cancer Informatics 2019;3:1—7.

[12] Leo P, et al. Stable and discriminating features are predictive of cancer presence and Gleason grade in radical prostatectomy specimens: a multi-site study. Scientific Reports 2018;8:14918.
[13] Bui MM, et al. Quantitative image analysis of human epidermal growth factor receptor 2 immunohistochemistry for breast cancer: guideline from the College of American Pathologists. Archives of Pathology & Laboratory Medicine 2019:1−16.
[14] Shrestha P, Hulsken B. Color accuracy and reproducibility in whole slide imaging scanners. Journal of Medical Imaging 2014;1:027501.
[15] Shrestha P, et al. A quantitative approach to evaluate image quality of whole slide imaging scanners. Journal of Pathology Informatics 2016;7:56.
[16] Xu J, et al. Stacked sparse autoendoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Transactions on Medical Imaging 2016;35(1):119−30.
[17] Lu C, et al. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. Laboratory Investigation 2018; 98(11):1438−48.
[18] Feng Y, Zhang L, Yi Z. Breast cancer cell nuclei classification in histopathology images using deep neural networks. International Journal for Computer Assisted Radiology and Surgery 2018;13:179−91.
[19] Chen H, et al. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Medical Image Analysis 2017;36:135−46.
[20] Mahmood F, et al. Adversarial U-net with spectral normalization for histopathology image segmentation using synthetic data. Proceedings of SPIE 2019;10956. Medical Imaging.
[21] Ciresan DC, et al. Mitosis detection in breast cancer histology images with deep neural networks. Medical Image Computing and Computer-Assisted Intervention 2013; 16(Pt2):411−3.
[22] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. Journal of Pathology Informatics 2016;7:29.
[23] Wang H, et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging 2014; 1(3):034003.
[24] Klauschen F, et al. Scoring of tumor-infiltrating lymphocytes: from visual estimation to machine learning. Seminars in Cancer Biology 2018;52:151−7.
[25] Sirinukunwattana K, et al. Gland segmentation in colon histology images: the GlaS challenge contest. Medical Image Analysis 2017;35:489−502.
[26] Monaco JP, et al. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Medical Image Analysis 2010;14(4): 617−29.
[27] Singh M, et al. Gland segmentation in prostate histopathological images. Journal of Medical Imaging 2017;4(2). 027501-1-02750125.
[28] Gavrielides MA, et al. Observer variability in the interpretation of HER2/neu immunohistochemical expression with unaided and computer-aided digital microscopy. Archives of Pathology & Laboratory Medicine 2011;135:233−42.
[29] Chang YH, et al. Multiplexed immunohistochemistry image analysis using sparse coding. In: Conference proceedings: annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Annual conference 2017; 2017. p. 4046−9.

[30] Vandenberghe ME, et al. Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. Scientific Reports 2017;7:45938.
[31] Khosravi P, et al. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. EbioMedicine 2018;27:317−28.
[32] Fondon I, et al. Automatic classification of tissue malignancy for breast carcinoma diagnosis. Computers in Biology and Medicine 2018;96:41−51.
[33] Araujo T, et al. Classification of breast cancer histology images using convolutional neural networks. PLoS One 2017;12(6):e0177544.
[34] Cruz-Roa A, et al. Accurate and reproducible invasive breast cancer detection in whole slide images: a deep learning approach for quantifying tumor extent. Scientific Reports 2017;7:46450.
[35] Bejnordi BE, et al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. Journal of Medical Imaging 2017;4(4). 044504-1-8.
[36] Steiner DF, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. The American Journal of Surgical Pathology 2018;42:1636−46.
[37] Bejnordi BE, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Journal of the American Medical Association 2017;318(22):2199−210.
[38] Bychkov D, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Scientific Reports 2018;8:3395.
[39] Nirschl JJ, et al. A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue. PLoS One 2018;13(4):e0192726.
[40] Fedorov A, et al. 3D slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging 2012;30(9):1323−41.
[41] Lee G, et al. Supervised multi-view canonical correlation analysis (sMVCCA): integrating histologic and proteomic features for predicting recurrent prostate cancer. IEEE Transactions on Medical Imaging 2015;34:284−97.
[42] Savage RS, Yuan Y. Predicting chemosensitivity in breast cancer with 'omics/digital pathology data fusion. Royal Society Open Science 2016;3:14051.
[43] Gallas BD, Gravrielides MA. Assessment of digital pathology. 2018. Available from: https://www.fda.gov/medical-devices/cdrh-research-programs/assessment-digital-pathology. [Accessed 14 September 2019].