Artificial Intelligence in Pathology: A Simple and Practical Guide

Keluo Yao, MD,* Amol Singh, BS,† Kaushik Sridhar, MS,* John L. Blau, MD,‡ and Robert S. Ohgami, MD, PhD*

Abstract: Artificial intelligence (AI) is having an increasing impact on the field of pathology, as computation techniques allow computers to perform tasks previously performed by people. Here, we offer a simple and practical guide to AI methods used in pathology, such as digital image analysis, next-generation sequencing, and natural language processing. We not only provide a comprehensive review, but also discuss relevant history and future directions of AI in pathology. We additionally provide a short tabular dictionary of AI terminology which will help practicing pathologists and researchers to understand this field.

Key Words: machine learning, digital pathology, artificial intelligence, natural language processing, next-generation sequencing

(*Adv Anat Pathol* 2020;27:385–393)

While the role of artificial intelligence (AI) in pathology has increased rapidly in the past several years, examples of its use go back for many years. In the 1970s, using classic computing principles, researchers designed a computing platform for AI, called expert systems. Some early examples of AI in medicine were built on expert systems, which relied on handcrafted rule-based algorithms. These translated into promising systems like MYCIN, which could identify bacteria and guide therapy based on clinical data,[1] and Pathology Expert Interpretative Reporting System (PEIRS), an automated chemical pathology report interpretation system.[2] The expert system's handcrafted rules required *domain knowledge* (ie, individual or group expertise in a certain field) during design, and faced significant challenges for tasks that required interpretations of sensory information, such as images.

In the 1990s, cytopathology and hematopathology were in the forefront of solving the issues with image processing. These later systems were designed based on feature engineering principles, which utilized domain knowledge in constructing algorithms to extract informative features from raw data.[3] PAPNET System, AutoPap 300 QC System, and later Thin-Prep Imaging System were cell morphology image analysis systems that received Food and Drug Administration (FDA) approval for clinical use during this period.[4–6] PAPNET pioneered the use of an artificial neural network, a form of machine learning (ML), to supplement the algorithm design. A major advance in computing, *ML* is a set of techniques to automatically create algorithms from data without reliance on domain knowledge (Fig. 1). PAPNET and AutoPap systems were designed as primary screening systems, while the Thin-Prep system was designed to assist cytotechnologists in the screening process by identifying cells of interest. In the early 2000s, semiautomated peripheral blood smear analysis, such as Cellavision DM96, appeared on the market. Similar to the cytology systems for the Pap test, the hematopathology systems aimed to assist laboratory technicians to identify and classify cells; in this instance, white blood cells.

Most systems did not survive commercially to the present day due to limitations of the earlier technologies and economic factors. Some, such as the ThinPrep Imaging and Cellavision systems have managed to be commercially viable on a large scale into the present day.

AI, powered by the various methods described, will continue to rapidly evolve and can be expected to grow exponentially in importance given its proven utility in numerous other industries (eg, self-driving cars, amazon recommendation systems, etc.). Thus understanding AI in the context of pathology is paramount to both academic researchers and practicing pathologists. We will provide a practical global overview of AI in pathology, in both clinical and relevant research settings.

COMPONENTS OF ARTIFICIAL INTELLIGENCE

ML is a subset of AI in which the creation of algorithms for data analysis, previously a human performed task, is being performed by computers. ML differs from classic computing in that ML algorithms are used to create a model from training data that can then be used to analyze new data sets (Fig. 1). Essentially, ML algorithms "learn" from exposure to vast amounts of
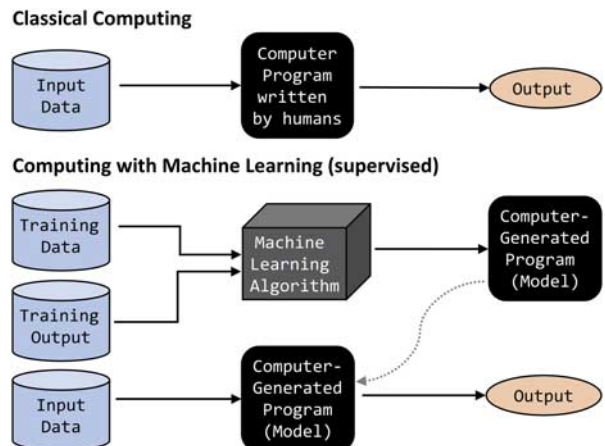
FIGURE 1. Comparison between classic computing and computing with machine learning. Please see this image in color online.
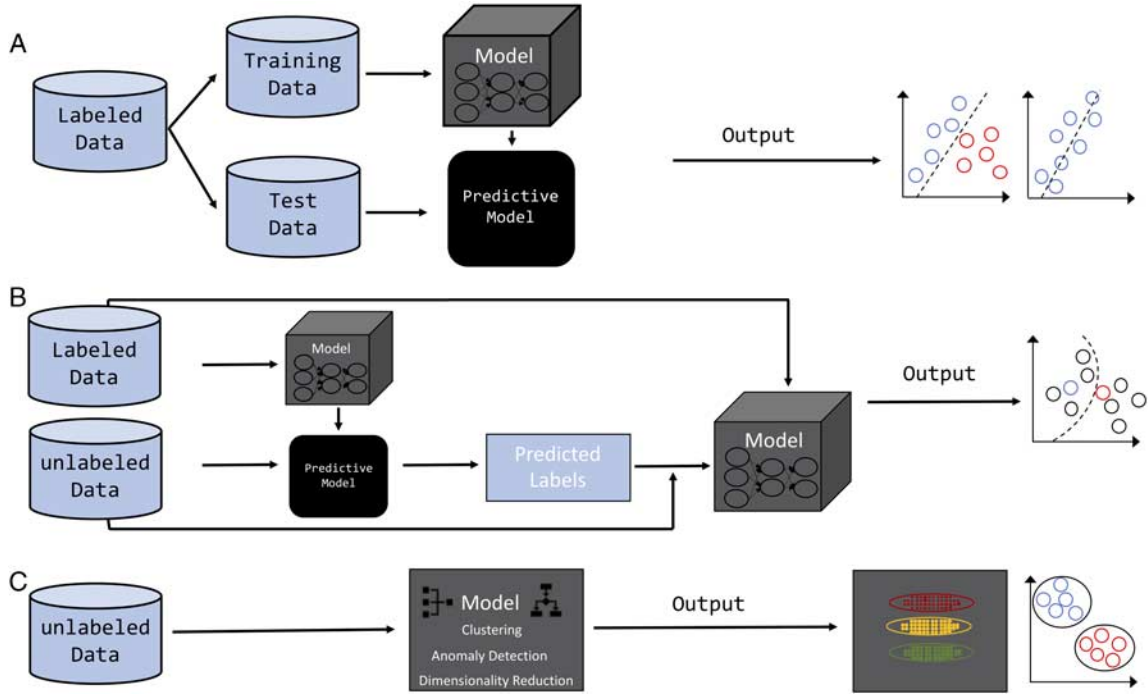
From the *Department of Pathology, University of California, San Francisco, San Francisco; †Department of Computer Science, Stanford University, Stanford, CA; and ‡Department of Pathology, University of Iowa, Iowa City, IA.

K.Y., A.S., K.S., and J.L.B. contributed equally.

FIGURE 2. Workflow for 3 types of machine learning algorithms. A, A workflow for a supervised learning pipeline. A model is trained on a set of labeled data and then evaluated on the testing data to determine the model's predictive strength. A regression or classification model can be trained by this pipeline. B, Semisupervised learning utilizes labeled and unlabeled data to generate a predictive model. Through training a model on the labeled data and using it to generate labels for the unlabeled data, the model creates a new and larger data set to train a final model on. C, Unsupervised learning pipelines are trained on unlabeled data. The pipelines can be trained for clustering, anomaly detection, and dimensionality reduction. Please see this image in color online.

data, while classic computing projects the expert domain knowledge onto a set of data.

The design phase of *ML* algorithms are frequently referred to as "training" and the methodology of ML is largely dominated by 3 main categories: *supervised learning*, *unsupervised learning*, and *semisupervised learning* (Fig. 2). Producing a useful model involves training and evaluating steps and is done using data that can be divided into *training*, *validation*, and *test data sets*.

Supervised learning is based on training algorithms using labeled and well-defined data sets to create a predictive model. This approach is akin to having a "teacher" supervise the computer algorithm to build an accurate model by providing a well-annotated data set. For instance, if one wants to identify images of cats versus dogs, supervised learning would use many thousands of labeled cat images and dog images to train the computer algorithm. Through establishing relationships and patterns between the input data (known images of cats and dogs) and *ground truth* output data (the category label of cat and dog), the algorithm generates a predictive model. The strength of the model is assessed with a validation data set (a new set of unknown images), which includes input-output data point pairs that were outside the original training data. The predicted output value from the model is compared with the expected output value to estimate the generalizability and strength of the model. The generalizability of a satisfactory model is further tested on an orthogonal test data set, preferably further separated from the training and validation data. Supervised learning is built for 2 major types of ML problems: classification and regression. *Classification* models are used to identify which category an input belongs to (eg, adenocarcinoma verse

squamous cell carcinoma).[7] *Regression* models are used to determine what output value an input data value should be assigned for continuous dependent variables (eg, predicting a patient's length of stay in a hospital).[8] Some examples of supervised learning algorithms are *artificial neural networks*, *decision trees*, *k-nearest neighbors*, and *linear regression*. One common aspect of supervised ML algorithms is that the accuracy and generalizability of the model can be increased by training it with a larger training data set. The number of data points necessary to achieve a useful model depends on the type of ML algorithm used, and type and quality of data analyzed. Image analysis algorithms, for example, typically require many thousands of labeled example images to train models to human levels of expertise. This can be a barrier for training models being used to recognize rare pathologies for which fewer than "many thousands" of cases may exist worldwide, let alone in a digitized training data set.

Unsupervised learning, on the other hand, identifies patterns in unlabeled input data. This has enabled researchers to generate analytic models in cases where there is a lack of ground truth. Essentially it is pattern recognition without regard to the meaning behind the pattern. Unsupervised learning is most commonly used for clustering data, anomaly detection, and data dimensionality reduction, and is useful in exploring new data sets. Unsupervised learning techniques such as cluster analysis only describe the features and patterns of the data; this is analogous to the identification of a new classification of a tumor by noting it is sufficiently morphologically different from other tumors.

Dimensionality reduction techniques, such as principle component analysis, are techniques for simplifying a predictive

TABLE 1. Terminology Central to Artificial Intelligence (AI) in Pathology

| Term	Definition
Artificial intelligence	Field of study that relates to the development and ability of computer systems to interpret information or perform tasks that would otherwise require human intelligence
Artificial neural networks	A machine learning model inspired by biological neural network
Bayesian networks	A machine learning model that updates the probability for a classification as more information becomes available
Classification	The identification of a category an input belongs to (eg, adenocarcinoma verse squamous cell carcinoma)
Computer vision	Field of science that involves the ability for computers to identify and interpret images and videos in the way that humans interpret them
Convolutional neural network	Type of machine learning algorithm that integrates feature extraction with neural network and is primarily used to analyze image data
Data augmentation	A collection of methods to increase the amount of training data using only the existing data by introducing noninterfering variations (eg, flipping, rotation, or resizing of digital images)
Decision trees	A machine learning model based on decision tree
Deep learning	A machine learning method, usually a variant of convolutional neural network that uses a very complex and deep networks to achieve a high predictive performance
Digital pathology	Subfield of pathology that involves the use of computers to store pathologic data and aid in the analysis of specimens
Dimensionality reduction	Techniques for simplifying a predictive model to its most important inputs
Domain knowledge	Individual or group expertise in a certain field
Ground truth	Ideal expected result used to train supervised or semisupervised machine learning algorithms
Machine learning	Field of science where computers are used to develop and implement mathematical models and algorithms to perform a task without explicit instructions
Narrow AI	Also known as weak AI, an implementation of AI that can only perform a narrow sets of tasks
Natural language processing	Field of AI where computational methods are used to interpret the human languages
Regression analysis	A machine learning model that estimates the relationship between input variables and outputs
Segmentation	Typically used in the context of digital image analysis. An image is "segmented" into different components that are meaningful
Semisupervised learning	Category of machine learning that uses a combination of labeled and unlabeled data to develop the algorithm
Strong AI	An implementation of AI that can solve general problems. A strong AI may be a collection of narrow AIs working together
Structured query language	A language designed to manage data in a relational database

TABLE 1. (*continued*)

Term	Definition
Supervised learning	Category of machine learning in which the algorithm is developed using labeled data. The algorithm is then used to predict future data outputs
Support vector machine	A machine learning model that uses hyperplane(s) to divide data points into classes
Systematized nomenclature of medicine	A computer processable collection of medical terms
Transfer learning	A deep learning method that applies a pretrained model on a different but similar problem with limited training data
Transfer learning	A machine learning technique that uses previously gained knowledge and apply it on a different but related problem
Unsupervised learning	A machine learning method that takes unlabeled input data to form and generate patterns

model to its most important inputs. Using one of these techniques, one could take a set of all commonly ordered laboratory tests and patient diagnoses and find the top laboratory tests predictive of the presence or absence of a given disease. Then one could use those specific tests with a supervised ML algorithm to create a predictive model based on those tests. For instance, while fasting glucose and hemoglobin A1C would likely be found to be highly associated with having diabetes, the 2 values are highly correlated; if one has a hemoglobin A1C, a fasting glucose value adds little to our predictive power. A well-designed dimensionality reduction algorithm might then prefer another laboratory test than fasting glucose that has less direct correlation with hemoglobin A1C.

Semisupervised learning is frequently used in settings where generating a large labeled data set is difficult, a frequently occurring scenario in pathology AI application development. Semisupervised learning is currently an evolving field, but the overall goal is to incorporate unlabeled data into the training process by unsupervised methods such as clustering or to use the partially trained model to "label" the data.[9] An emerging topic within this field is multiple-instance learning, in which each of the training data points are separated into bags of instances with one label. For the bag of instances, the positive label applies to the bag when at least one of the instances in the bag has a positive label and the negative label applies when all the instances have negative labels.[10] In this manner, multiple-instance learning incorporates unlabeled data and labeled to learn patterns and build models. Another technique involves the clustering of unlabeled data with labeled data based on image characteristics and then providing the unlabeled data pseudo labels that will then be used for training. As an extension of semisupervised learning, *Self-supervised learning* is one of the newer techniques and the goal is to learn generalizable and basic image context with unlabeled data. This method generates a baseline model that can be further specialized to tasks with a small amount of labeled data. For the generation of this baseline model, pretext tasks are used, such as randomly patching an image and training a model to determine the location of the patches.[11] This trains a baseline model that gains a spatial understanding of the data. The baseline model, with the initial values learned from the pretext task, can be used for supervised learning as it has shown to improve model accuracy with less labeled data.
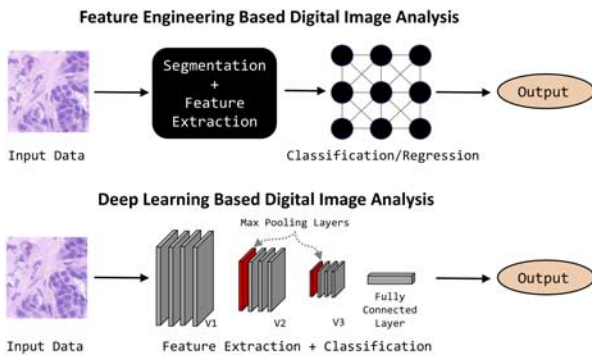
FIGURE 3. Comparison between feature engineering based digital image analysis and deep learning-based digital image analysis. Please see this image in color online.

ML has existed since the late 1950s, and numerous models have been created since then.[12] Table 1 contains descriptions of some widely used ML models.

IMAGE ANALYSIS AND MACHINE LEARNING

Digital image analysis, and the closely related idea of computer vision, have previously been very difficult to perform with AI. Digital image analysis has benefited tremendously from an increase in affordable computational power in recent years, and from the development of improved ML algorithms such as deep convolutional neural networks (CNNs). *Digital image analysis* in its essence involves the computation of a very large array of digits, a task perfectly suited for deep CNN as it mimics biological visual systems.[13] Before the introduction of deep CNN, digital image analysis in pathology relied heavily on domain knowledge of pathologists to create feature engineering methods to come up with algorithms. Feature engineering is divided into 2 phases; segmentation and feature extraction (Fig. 3). In *segmentation*, the image is separated into "segments" by meaningful categories. For example, a simple threshold method can be used to parse out darker regions from lighter regions. More sophisticated algorithms can use a combination of color, texture, contrast, and other features to segment. Image filters (eg, sharpen, edge detection, blur, etc.) can also be mixed in. In *feature extraction*, the "segments" are extracted and computationally measured. Figure 4 shows an example of a cervical biopsy hematoxylin and eosin image being processed by a feature engineering method. The goal of feature engineering is to produce properly structured computable data for classification. In this final step, an ML-based method is the most popular approach. Any of the previously mentioned techniques can be used.
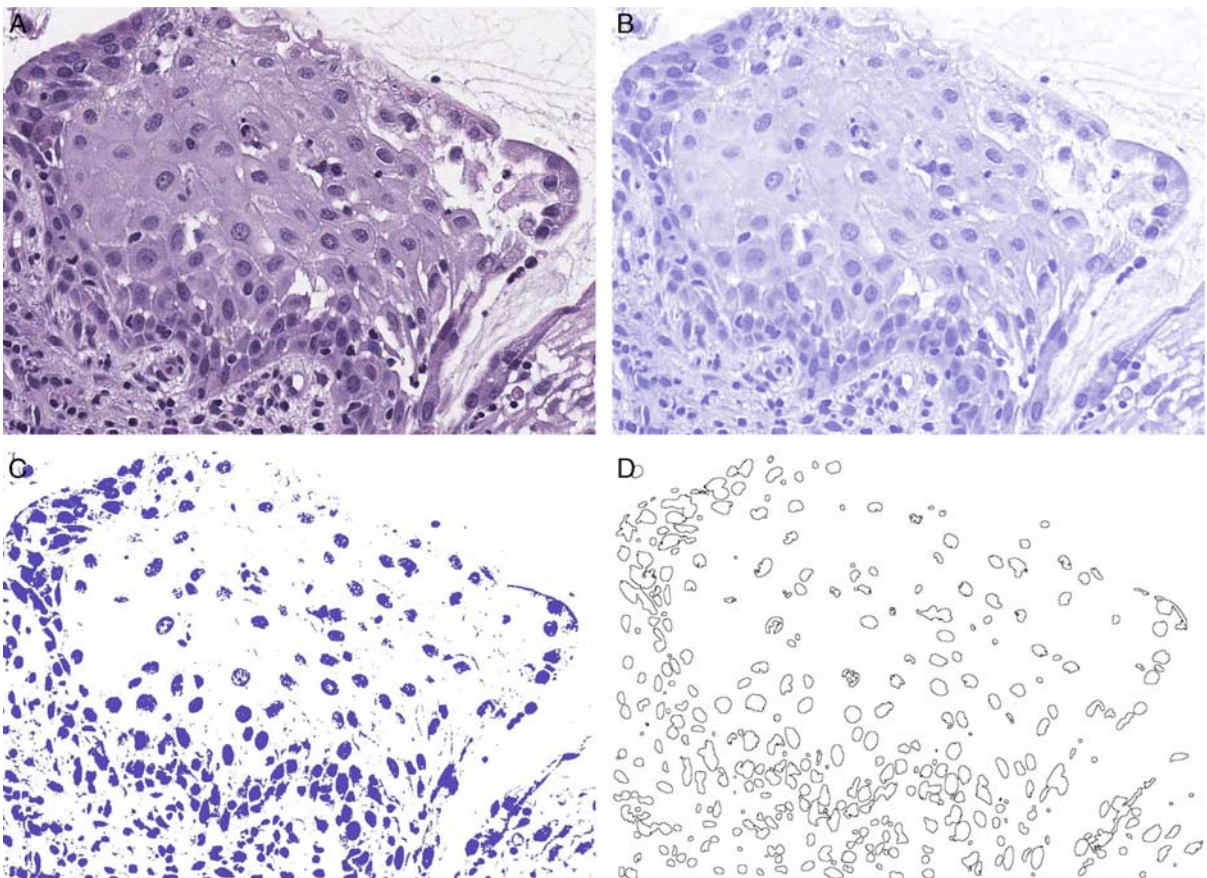
FIGURE 4. Using traditional feature engineering technique to analyze cervical biopsies: A digital image from a cervical biopsy (A) processed by a color deconvolution method that takes advantage of color contrast in H&E stain (B). C, A threshold method can be used to segment or separate the nuclei from other cell features. D, The segmented nuclei can be further extracted for their numerical properties.
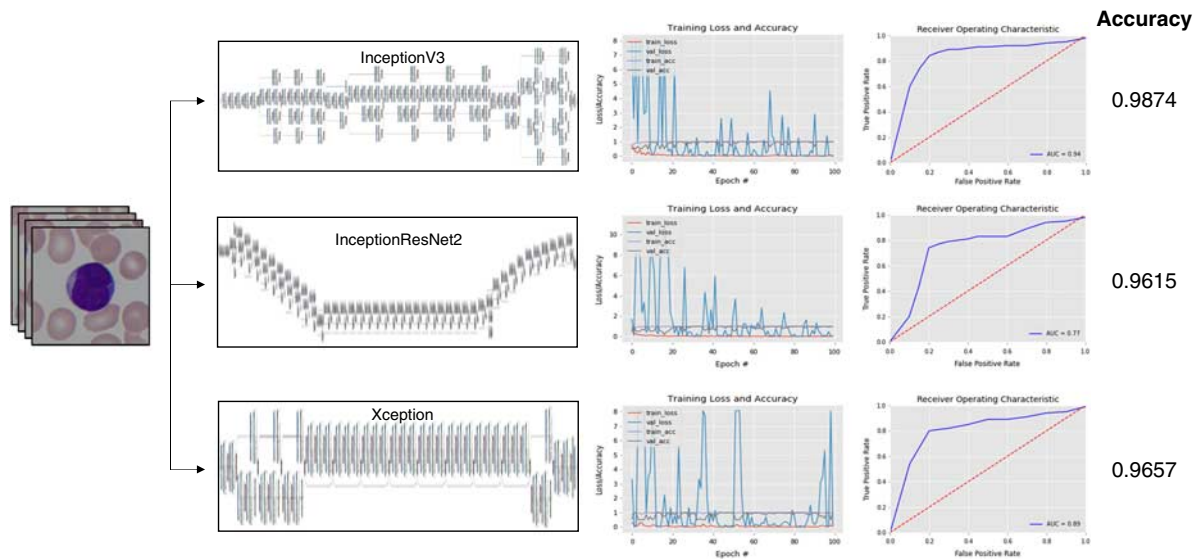
FIGURE 5. Three example workflows for an acute lymphoblastic lymphoma (ALL) classification pipeline. The InceptionV3, Inception-ResNet2, and Xception architectures (visualized above using Netron) are applied to the ALL IDB2 data set of ALL and normal cells. The training loss and accuracy plot displays the model loss over accuracy while getting trained over 100 epochs. In addition, the receiver operating characteristic curves are displayed along with the corresponding area under the curves. The InceptionV3 achieved the highest prediction accuracy on the test data. Please see this image in color online.

CNNs have revolutionized digital image analysis by reducing the need for domain knowledge dependent feature engineering, though domain knowledge is still required to provide the labeled training data sets. Since 2012, most successful algorithms from Large Scale Visual Recognition Challenge (ILSVRC) were different architectures of CNNs. Most of them shared a similar design that involved "convolving" the image layers using image filters to produce feature signals to be processed by artificial neurons. Not unlike biological neural networks, CNN can detect and process sophisticated image features with the right number of layers and complexity. As an extremely simplified example, the successive layers of image processing can be said to mimic the human visual cortex in which successive layers of convolutional layers and neurons (V1, V2, V3, etc.) further abstract encoded data from the raw visual input. V1, for instance, detects differences in contrast (edge detection), while V2 specializes in binocular disparities that help determine depth of field, and V3 uses input from previous levels to recognize characteristics such as color, spatial frequency and orientation, and simple geometric shapes. Subsequent levels use condensed inputs from the previous levels through max pooling layers to recognize complex features. The final fully connected layer, consisting of artificial neurons, can determine if the extracted features from previous layers are sufficient to classify (eg, a certain Gleason Score) (Fig. 3). Some deep CNN models can perform segmentation, feature extraction, and classification all at once. The performance of a CNN rests on the layer design, hyperparameter settings, and training data. The training of a deep CNN can be extremely computationally expensive but it can achieve an extremely high level of performance (Fig. 5). Beyond that, recent advances are giving us a way to visualize the previous deep learning "black box" (Fig. 6). Such visualization techniques may allow us to use deep learning algorithms as a computational method to gain reproducible knowledge from the traditionally subjective morphologic features, transforming the study of histology into a computational discipline.

ARTIFICIAL INTELLIGENCE APPLICATION DEVELOPMENT AND DEPLOYMENT

AI is a maturing technology. As such, publications tend to be research and technology prototypes. Pioneering laboratories may have to validate and deploy these technologies as a laboratory-developed test for diagnostic purpose. Therefore, an understanding of the tools/prototypes used in the development of AI application development is essential. Appropriate validation guidelines are available from Centers for Medicare and Medicaid Services (CMS) and the College of American Pathologists (CAP).[14,15] The advances we have made so far are for *narrow AI* applications or applications that are only suited for a limited set of tasks. Successful projects in the past (eg, ThinPrep Imaging System) all tried to meet a specific need in the clinical practice or research workflow.

A key requirement for developing generalizable ML-based AI applications is high quality annotated data of sufficient quantity. Having access to clinical pathology images accurately annotated by practicing pathologists specific to an institutional need is a luxury. Methods such as self-supervised learning, semisupervised ML, data augmentation, and transfer learning can provide some relief. For the researchers and developers who lack access to large institutional data sets, there are publicly available image repositories such as the Cancer Genome Atlas Program (TCGA), Acute Lymphoblastic Leukemia Image Database for Image Processing (ALL-IDB), and Breast Cancer Wisconsin (Diagnostic) Data Set. The difficulty of procuring real medical images in sufficient quantity and quality for ML research has generated some interest in synthetic data. Other industries such as nuclear energy, aviation, and autonomous vehicle researches all have successfully adapted it to negate the danger of using real data.[16] As artificial data created from scratch, or generated using data manipulation techniques, properly created synthetic data can be used to for ML training without the burden of maintaining patient confidentiality and data security.

AI applications can be developed from a wide array of tools. ImageJ, a portable application and its derivatives (eg, Fiji)
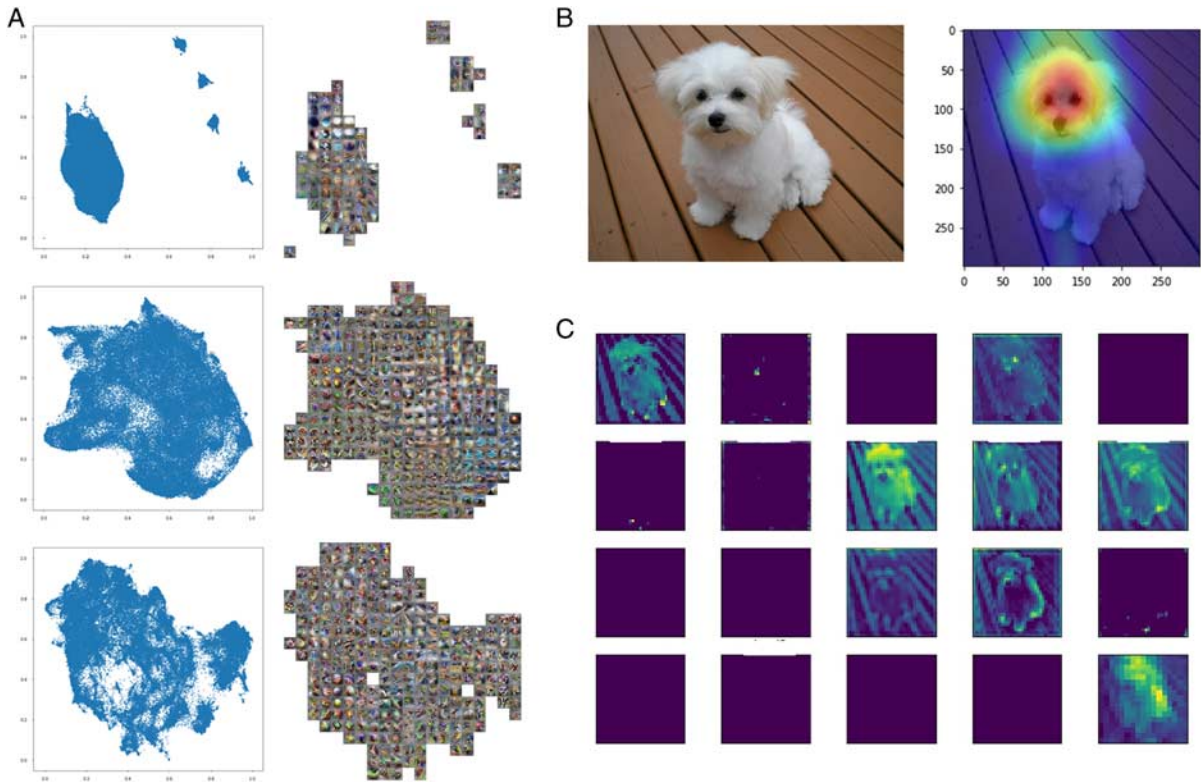
FIGURE 6. Feature visualization using deep convolutional neural network. A, Activations visualized using UMAP to aggregate and cluster similar activations for the InceptionV1 architecture. B, A class activation map of the InceptionV3 that depict the most important pixels and areas on the image for generating the classification output. C, The layer wise output visualization of the InceptionV3 that portrays the results of the intermediate layers of the convolutional neural network. Please see this image in color online.

are perhaps the best known digital image analysis tool in the field. It is built specifically for biological image analysis problems and supported by a large contributing community and numerous plugins.[17] Other comparable tools include CellProfiler, Bio-ImageXD, Icy, iLastik, Vaa3D, and QuPath. Some of these solutions (QuPath and ImageJ) offer capabilities to manipulate whole slide images.[18] One key disadvantage of most open-source image analysis platforms is the lack of a complete set of tools to complete the project from image segmentation to classification. For example, ImageJ currently does not have a built-in classifier and the end-user must provide their own or use a plugin.[17] Newer platforms such as QuPath has a similar approach.[18] Vendor-supported solutions often offer "the complete pipeline" in this regard with an upfront cost. Examples of vendor-supported solutions include Definiens tissue studio and Viziopharm histopathology digital image analysis. Some solutions even require the user to purchase vendor-specific hardware as the Ventana iScan systems and the companion software. The latter example has an extremely high upfront cost but offers vendor-supported, validated, and deployable solutions right out of the box.

Some of the tools for ML algorithm development do not even require one to know how to code. Beginners can experiment with ML algorithms without any prior experience in programming by using a graphical user interface to construct their own algorithms. Some can be downloaded onto a desktop and run as a locally installed application (eg, Orange). Others take the data from the user and perform the processing in the cloud (eg, Datarobot). Most software packages are free to use but some cloud-based solutions such as the DataRobot operates using the "freemium" model.

To create a finely tuned ML algorithm with the intention for clinical or research deployment, a true programming platform, which can be open source or proprietary, is probably needed. Python and R are good examples of open-source programming languages where ML libraries are well developed and supported.[19,20] Python emphasizes code readability and is a true objective-oriented programming language with capabilities extending well beyond ML applications.[19] R has been the preferred programming language for data scientists.[20] Both options now have deployment options as web applications. Python, being a general-purpose programming language, has the advantage of the ease of integration with other languages which allows more flexible deployment.[19]

With a long-track record of committed users from academia and industry, MATLAB has been one of the well-maintained heavyweight data analytics platforms. Licenses can be expensive for independent users but well-funded academic and industrial institutions usually make them available to researchers. However, being a proprietary platform, the end-users trade a polished development experience for an open-source ecosystem.

NONIMAGING RELATED MACHINE LEARNING TECHNOLOGIES

Next-generation Sequencing and Artificial Intelligence

In addition to imaging applications, specialized ML technologies have been developed and deployed to aid precision medicine by providing better tools for *next-generation*
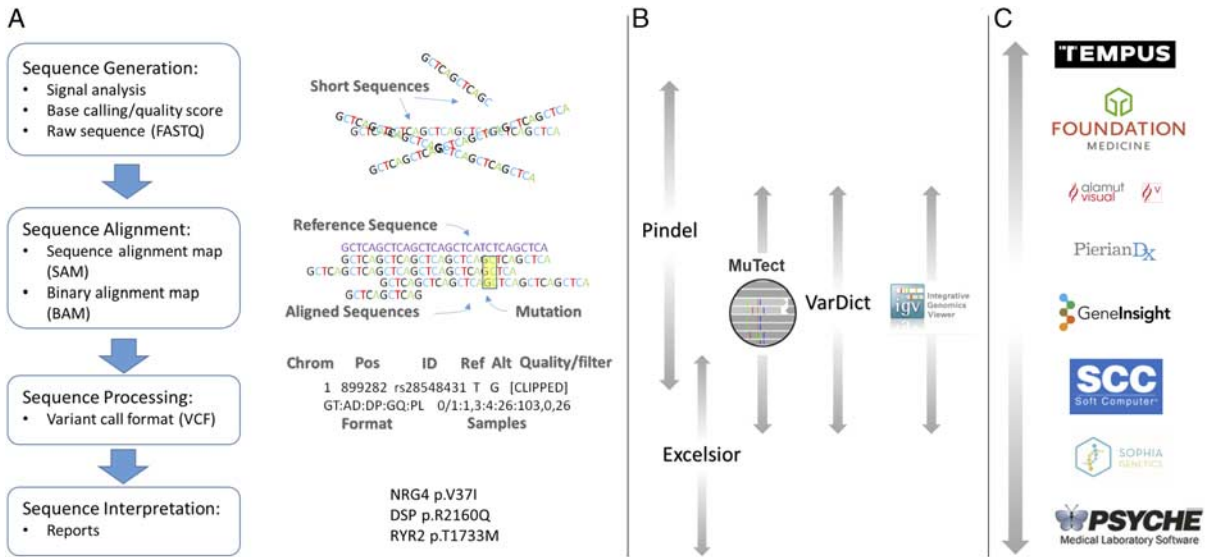
FIGURE 7. Next-generation sequencing data pipeline. A, The pipeline consists of sequence generation (FASTQ), alignment (SAM or BAM), processing (VCF), and interpretation. B, Various open-source tools (Pindel, MuTect, and VarDict) and laboratory-developed applications (Excelsior, University of Michigan) are designed to handle specific portion of the pipeline. C, A noncomprehensive list of current commercial companies that can process the entire pipeline. Please see this image in color online.

sequencing (NGS). The highly multiplexed nature of NGS has helped to alleviate some of the problems associated with traditional Sanger sequencing, namely cost, reaction time and sensitivity to low-frequency variants.[21,22] NGS methods generate a massive amount of data per sequencing run and manual identification of pathologically and clinical significant mutations is not feasible and must be accomplished by machine assistance. On the basis of the iterative nature of these processes most laboratories have adapted a data pipeline style of workflow.[23]

A typical data pipeline consists of sequence generation, alignment, processing, and interpretation, producing mostly standardized structured files at each step (Fig. 7). Sequence generation produces the raw short sequences, which are then aligned algorithmically to produce a sequence alignment map or binary alignment map. Sequence processing takes the aligned map files and makes variant calls by comparing the aligned sequences to a reference genome, producing a file called variant call format. Variant calling may produce erroneous results based on the sample type, preparation, sequencing platform, and errors in sequencing and alignment.[24] An additional error can occur from manual curation of sequencing data, which can be required based on the filters used and random biases in the sequencing data.[25,26] Sequence interpretation utilizes the variant call format file and renders an interpretation, which is heavily dependent on literature mining.[27]

At the moment, NGS data pipeline have standardization and quality control issues and ML methods can be potential solutions.[28] ML-based variant calls can potentially increase the performance and accuracy through better usage of variables such as qualities of sequencing, alignment, and tumor, potentially increasing the use of inferior quality samples. Google developed DeepVariant which treats binary alignment map files like images and thus leverages the deep learning technologies to significantly improve performance.[29] Many variations of this system are in development and high-quality reference databases

(eg, ClinVar, UniProt, gnomAD) are absolutely essential. Several earlier tools for variant annotation and classification such as PolyPen and SIFT failed to establish themselves definitively in the clinical setting. The recent release of American College of Medical Genetics (ACMG) standards and guidelines for the interpretation of the sequence variants has fostered another wave of development.[30] Companies like Fabric Genomics and Invitae have produced tools that can classify variants based on ACMG guideline with rapid turnaround time.[31,32] Nonetheless, the lack of ground truth data derived from protected health information repositories has typically been a bottleneck in evaluating the accuracy of AI applications for clinical decision making. Some institutions are exploring new collaborative ventures with companies like Sophia Genetics to overcome this limitation by analyze patient data in a secure manner.[33]

Artificial Intelligence in Natural Language Processing

The recent media buzz led by IBM Watson has created an impression that human or superhuman level of language processing and knowledge mastery has been achieved. Indeed since the introduction of IBM Watson, ML enabled natural language processing (NLP) and knowledge base has been claiming medical board certification level of competency in China and Britain with varying degree of success.[34,35] The advent of electronic health records (EHRs) to store clinical information in structured and unstructured forms has created the infrastructure and the potential use case scenarios for optimizing of clinical decision trees and biomedical research.[36]

It is important to recognize that computerized processing of clinical text has been in practice since computers were first used in clinical settings. Many earlier attempts aimed at creating rules to structure clinical text information rather than designing a system to accommodate traditional documentation style. The International Statistical Classification of Diseases and Related Health Problems (ICD) and Systematized Nomenclature of Medicine (SNOMED) were all

attempts to produce structured and granular clinical text so they can be more easily consumed by earlier computer systems with limited capabilities. In general, anatomic pathology reporting is slowly mirroring the clinical workflow which has extensively adapted the use of checkboxes and structured templates. The College of American Pathologist is advocating for electronic Cancer Checklists (CAP eCC) that can provide template-based structured input to produce standardized cancer synoptic reports. Bethesda classification systems for reporting cervical cytology, thyroid cytopathology, the Milan system for reporting salivary gland cytopathology, and the Paris system for reporting urinary cytology are similar reforms to create more structure and granularity that will ultimately make NLP easier. While pathology reports are still predominately in free-text form, it is possible to utilize context-free grammar to parse them and extract meaningful information. Nonetheless, routine pathology and laboratory workflow continues to produce large amounts of unstructured material that requires more robust NLP to translate into clinical management and research information.

The most common method for text mining for pathology reports utilizes Structured Query Language, a language designed to manage data in a relational database; the relational database is the most common type of database for a laboratory information system. As a fully-featured programming language designed to query information, one can create information extraction pipelines using Structured Query Language alone.[37] Such a pipeline would require well-engineered business logic or rules to compensate for errors in spelling and grammar as well as lingual variations. A Grammar-based parsing approach can be used on well-structured and consistent data.

The recent attempts by the technology companies to gain entry into medicine through the NLP pathway is initiating the shift from rule-based algorithms to ML or statistics-based methods. Analogous to the revolution in digital image analysis, the shift is reducing the need for domain knowledge. The centerpiece of the technical advances from IBM Watson and Amazon Comprehend are the abilities of the algorithm to extract features such as words and relationships between words to classify a blob of text, a process called text classification. Pathology is in the very early phase of investigating the use of ML-based NLP methods, but potential applications are numerous. For instance a NLP pipeline for medical billing can be constructed by parsing the necessary pathology reports from a laboratory information system, which can be mapped to Current Procedural Terminology (CPT) codes with SNOMED and perhaps further enhanced with ML-based methods to facilitate automated billing.[38] Clinical and epidemiological studies that require text mining can be implemented this way.[39]

Though the use of NLP in health care is yet to reach its full potential, companies like Amazon Comprehend Medical and Mendel AI are using their algorithms to connect patients with relevant clinical trial information based on EHR data.[40,41] The advent of AI is facilitating increased interaction between patients, pharmaceutical companies, hospitals, researchers, technology companies, and insurers, and this may provide a potential solution for accessing EHR data in a regulated way.

FUTURE DIRECTIONS

While AI applications continue to improve, most of them will continue to be in the narrow AI domain, focusing on only a single task. The current AI tools can recognize tumors and score grades, but none can do everything that a pathologist does while looking at a glass slide. We can look at other industries to help guide our path forward. Self-driving cars are materializing one component at a time; first, we had the antilock braking system, then came cruise control, which then became adaptive cruise control, finally lane-centering technologies that allowed the car to stay in the lane without driver input. This shows perhaps strong general-purpose AIs are simply a collection of narrow AIs working together.

At the time of writing, the tragic impact COVID-19 can be felt in every corner of the globe. The digital transformation of pathology, and the emphasis on working remotely, has been further accelerated due to the need for social distancing measures. The power of AI has also been felt in the pharmaceutical sector where CNNs can predict protein-ligand interactions in addition to repurposing treatment regimens based on synthetic predictions.[42] Precise prediction of the future is difficult, but it is safe to say this trend will catalyze more development of AI applications for pathology. Increased reliance on AI for clinical trials, digital pathology, genomics and synthetic biology reveals the various areas for growth in health care. Analogous to the building of more roads leading to more cars, the increased usage of digital pathology will lead to more AI digital pathology applications.

CONCLUSIONS

AI technologies has been powering pathology and laboratory medicine since computers were introduced into the workflow. Early methods relied on classic computing but few gained traction. Early image analysis tools relied on feature engineering methods and some of them thrived in clinical and research settings. ML-powered AI technologies are now dominating the field due to the method's improved performance and flexibility. Application of AI technologies in digital image analysis, NGS, and NLP will continue to evolve in the foreseeable future.

REFERENCES

1. Shortliffe E. *Computer-based Medical Consultations: MYCIN.* New York, NY: Elsevier; 2012.
2. Edwards G, Compton P, Malor R, et al. PEIRS: a pathologist-maintained expert system for the interpretation of chemical pathology reports. *Pathology*. 1993;25:27–34.
3. Heaton J. An empirical analysis of feature engineering for predictive modeling. SoutheastCon: IEEE; 2016:1–6.
4. Koss LG, Lin E, Schreiber K, et al. Evaluation of the PAPNET™ cytologic screening system for quality control of cervical smears. *Am J Clin Pathol*. 1994;101:220–229.
5. Colgan TJ, Patten SF Jr, Lee JS. A clinical trial of the AutoPap 300 QC System for quality control of cervicovaginal cytology in the clinical laboratory. *Acta Cytol*. 1995;39:1191.
6. Biscotti CV, Dawson AE, Dziura B, et al. Assisted primary screening using the automated ThinPrep Imaging System. *Am J Clin Pathol*. 2005;123:281–287.
7. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24: 1559–1567.
8. Tsai PFJ, Chen PC, Chen YY, et al. Length of hospital stay prediction at the admission stage for cardiology patients using artificial neural network. *J Healthc Eng*. 2016;2016:7035463.
9. Lu MY Williamson DFK, Chen TY, et al. Data efficient and weakly supervised computational pathology on whole slide images. arXiv Prepr arXiv200409666; 2020.

10. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25:1301–1309.
11. Doersch C, Gupta A, Efros AA. Unsupervised visual representation learning by context prediction. In: Proceedings of the IEEE International Conference on Computer Vision; 2015:1422–1430.
12. Kohavi R. Glossary of terms. *Mach Learn*. 1998;30:271–274.
13. Lindsay G. Convolutional neural networks as a model of the visual system: past, present, and future. *J Cogn Neurosci*. 2020:1–15.
14. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med*. 2013;137: 1710–1722.
15. Lange H. Digital pathology: a regulatory overview. *Lab Med*. 2011;42:587–591.
16. Nikolenko SI. Synthetic data for deep learning. arXiv Prepr arXiv190911512; 2019.
17. Ferreira T, Rasband W. ImageJ user guide. *ImageJ/Fiji*. 2012; 1:155–161.
18. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7:1–7.
19. Müller AC, Guido S. *Introduction to Machine Learning With Python: a Guide for Data Scientists*. Sebastopol, CA: O'Reilly Media Inc.; 2016.
20. Bischl B, Lang M, Kotthoff L, et al. mlr: machine learning in R. *J Mach Learn Res*. 2016;17:5938–5942.
21. Sridhar K, Singh A, Butzmann A, et al. Molecular genetic testing methodologies in hematopoietic diseases: current and future methods. *Int J Lab Hematol*. 2019;41:102–116.
22. Rivas MA, Beaudoin M, Gardet A, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet*. 2011;43:1066–1073.
23. Jalali M, Saldanha FYL, Jalali M. Basic science methods for clinical researchers. basic science methods for clinical researchers; 2017.
24. Merker JD, Valouev A, Gotlib J. Next-generation sequencing in hematologic malignancies: what will be the dividends? *Ther Adv Hematol*. 2012;3:333–339.
25. Kulkarni P, Frommolt P. Challenges in the setup of large-scale next-generation sequencing analysis workflows. *Comput Struct Biotechnol J*. 2017;1:471–477.
26. Bonfield JK, McCarthy SA, Durbin R. Crumble: reference free lossy compression of sequence quality values. *Bioinformatics*. 2019;35:337–339.
27. Xu J, Yang P, Xue S, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. *Hum Genet*. 2019;138:109–124.
28. Lubin IM, Aziz N, Babb LJ, et al. Principles and recommendations for standardizing the use of the next-generation sequencing variant file in clinical settings. *J Mol Diagn*. 2017; 19:417–426.
29. Anderson C. Google's AI tool deepvariant promises significantly fewer genome errors. Clinical OMICS; 2018;5:33.
30. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–423.
31. Fraikin G. Fabric genomics announces AI-based ACMG Classification solution for genetic testing with hereditary panels. Businesswire; 2019.
32. Invitae. Invitae's interpretation and reporting process: step by step title. Available at: www.invitae.com/en/variant-classification/. Accessed June 2, 2020.
33. Camblong J. Hospital uses AI to embrace clinical genomics. Pharmaphorumcom; 2017. Available at: https://pharmaphorum.com/views-and-analysis/hospital-ai-genomics/.
34. Wu J, Liu X, Zhang X, et al. Master clinical medical knowledge at certificated-doctor-level with deep learning model. *Nat Commun*. 2018;9:1–7.
35. Oliver D. David Oliver: lessons from the Babylon Health saga. *BMJ*. 2019;365:l2387.
36. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13:395–405.
37. Glaser AP, Jordan BJ, Cohen J, et al. Automated extraction of grade, stage, and quality information from transurethral resection of bladder tumor pathology reports using natural language processing. *JCO Clin Cancer Inform*. 2018;2:1–8.
38. Kim SY, Kim HH, Lee IK, et al. Proposed algorithm with standard terminologies (SNOMED and CPT) for automated generation of medical bills for laboratory tests. *Healthc Inform Res*. 2010;16:185–190.
39. Nguyen A, Moore J, Lawley M, et al. Automatic extraction of cancer characteristics from free-text pathology reports for cancer notifications. *Stud Health Technol Inform*. 2011;168:117–124.
40. Mishra A. Machine learning in the AWS cloud. [Amazon Comprehend Medical website]. 2002. Amazon Comprehend Medical. Available at: https://aws.amazon.com/comprehend/medical/. Accessed June 2, 2020.
41. Mendel. Turbocharging clinical research by using machine precision with human intuition. Available at: www.mendel.ai/. Accessed June 2, 2020.
42. Ragoza M, Hochuli J, Idrobo E, et al. Protein-ligand scoring with convolutional neural networks. *J Chem Inf Model*. 2017;57: 942–957.