In recent years, machine learning and deep learning-based approaches, two sub-fields of artificial intelligence, have emerged as key components in biomedical data analyses (1-5). They can be applied to image segmentation, identifying insertion/deletion mutations, protein alignments, and so on. Several studies have integrated pathological image data with genomics data. Yuan et al. have quantitatively analyzed image data to better describe and validate the independent prognostic factors in estrogen receptor-negative breast cancer (6). Another study by Copper et al. also used histopathology images and genomics data to identify prognostic factors in breast cancer (7). Other types of cancers such as prostate cancer (8), renal cell carcinoma (9), low grade glioma (10), and non-small cell lung cancer (11), just to name a few, have also been studied by approaches integrating (multi-)omics data with pathology images.

The literature on deep learning methods used to assist cancer diagnosis and predict patient outcomes enables us to observe the exploding trend in this field (12-15). Massive amounts of published research and large numbers of clinical trials have illustrated the reliability and practicality of machine learning approaches, particularly deep learning. Various studies have employed deep learning methods to auto-detect and classify benign nuclei from cancer cells (1,16,17), to identify and quantify the rate and amount of mitosis (18). Deep learning has also been used for tissue origin classification, nuclear grading, precision medicine matching trials (1,19,20), classification of ancient and modern DNA (21), and drug repurposing (22).

For tissue quantification, there are two primary methods, namely handicraft features and the unsupervised approach (23). The former method consists of domain-agnostic and domain-inspired features (24,25) whereas the latter uses an automatic approach to identify distinguishing features (26). Domain-agnostic features focus on nuclear appearance, gland shape, object size, tissue texture, and architecture, while the domain-inspired features focus on certain particular domains, such as disease and organ origin (27). There have been studies that have applied these methods in prostate cancer and triple negative breast cancer (TNBC) samples (27). Gland architecture has been correlated by the domain-inspired approach (25) with aggressiveness of intermediate-risk prostate cancer. Another study calculated the number of intra-tumor lymphocytes, adjacent lymphocytes, and distant site tumor lymphocytes from TNBC (26). These studies found that these cell types and numbers can be used as independent prognostic predictors of disease-specific survival in TNBC (26). Tissue microarrays have also been used to predict colorectal cancer patient outcomes by deep learning approaches (28). The advantage of deep learning is that it is quick and seamless, although feature interpretability is missing (23).

In recent years, various tools related to pattern recognition have also been developed, and huge numbers of datasets are now readily available for public use. Many archives and databases for radiological and pathological images have been established as well, such as The Cancer Imaging Archive (TCIA) and the Cancer Digital Slide
Archive (CDSA), both of which facilitate image data analyses. Taking advantage of these databases and archives, many studies have been published with MRI and/or CT imaging incorporated with biological pathways and cellular morphology to further characterize a disease (29-34). These radiological data could potentially aid in determining the molecular subtypes of cancer.

Furthermore, radiological data could be linked to gene expression and/or mutation profiles to identify distinct cellular subtypes within the same cancer. Radiological data comprising hundreds of thousands of cells within a patient, once coupled with gene expression, could decipher the multiple dimensional features of the tumor, which is not achievable with genomic data alone (35). Consequently, the integration of radiological features with genomic data undoubtedly has a crucial role to play in improving diagnostic, prognostic, and predictive power in comparison to conventional approaches such as immunohistochemical assays. There are a couple of research papers that have applied radiological and genomic data to discriminate prostate cancer tissues from benign tissues, thereby enhancing information related to prostate cancer aggressiveness (36). Another study conducted with lung adenocarcinoma integrated CT images to predict the metastatic potency driving cells to distant organs (37).

Radiomics has been shown to be powerful in parallel with genetic markers with extracted semantic and agnostic features. Integrating multiple platforms to bridge radiomics with genomics could lead to better characterization of disease. This is of particular value for better treatment decisions and correct explanation of biological and treatment heterogeneity. Especially for cancer therapy, image-aided decision-making is crucial. For instance, the integrated radiomics and pathological features corresponding to a specific breast cancer molecular signature can provide prognosis markers and surrogates to predict patient outcomes, drug responsiveness, and eventually enhance treatment efficacy. Prospective studies using artificial intelligence as the predominant tool for classification and/or prediction tasks, along with omics and imaging data, could certainly facilitate and accelerate research output and accuracy. However, machine learning requires informed human supervision, as results without proper interpretation are not of much value.

Acknowledgments

We thank Melissa Stauffer, PhD, for editing the manuscript.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

35. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563-77.
