Cancer Cell

CellPress
OPEN ACCESS

Review

Artificial intelligence for multimodal data integration in oncology

Jana Lipkova,[1,2,3,4] Richard J. Chen,[1,2,3,4,5] Bowen Chen,[1,2,8] Ming Y. Lu,[1,2,3,4,7] Matteo Barbieri,[1] Daniel Shao,[1,2,6] Anurag J. Vaidya,[1,2,6] Chengkuan Chen,[1,2,3,4] Luoting Zhuang,[1,3] Drew F.K. Williamson,[1,2,3,4] Muhammad Shaban,[1,2,3,4] Tiffany Y. Chen,[1,2,3,4] and Faisal Mahmood[1,2,3,4,9,*]

[1]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[2]Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[3]Cancer Program, Broad Institute of Harvard and MIT, Cambridge, MA, USA
[4]Data Science Program, Dana-Farber Cancer Institute, Boston, MA, USA
[5]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[6]Harvard-MIT Health Sciences and Technology (HST), Cambridge, MA, USA
[7]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA
[8]Department of Computer Science, Harvard University, Cambridge, MA, USA
[9]Harvard Data Science Initiative, Harvard University, Cambridge, MA, USA
*Correspondence: faisalmahmood@bwh.harvard.edu

https://doi.org/10.1016/j.ccell.2022.09.012

SUMMARY

In oncology, the patient state is characterized by a whole spectrum of modalities, ranging from radiology, histology, and genomics to electronic health records. Current artificial intelligence (AI) models operate mainly in the realm of a single modality, neglecting the broader clinical context, which inevitably diminishes their potential. Integration of different data modalities provides opportunities to increase robustness and accuracy of diagnostic and prognostic models, bringing AI closer to clinical practice. AI models are also capable of discovering novel patterns within and across modalities suitable for explaining differences in patient outcomes or treatment resistance. The insights gleaned from such models can guide exploration studies and contribute to the discovery of novel biomarkers and therapeutic targets. To support these advances, here we present a synopsis of AI methods and strategies for multimodal data fusion and association discovery. We outline approaches for AI interpretability and directions for AI-driven exploration through multimodal data interconnections. We examine challenges in clinical adoption and discuss emerging solutions.

INTRODUCTION

Cancer is a highly complex disease involving a cascade of microscopic and macroscopic changes with mechanisms and interactions that are not yet fully understood. Cancer biomarkers provide insights into the state and course of disease in the form of quantitative or qualitative measurements, which consequently guide patient management. Based on their primary use, biomarkers can be diagnostic, prognostic or predictive of response and resistance to treatment. Diagnostic markers stand at the first line of cancer detection and diagnosis, including examples such as prostate-specific antigen (PSA) values, indications in radiologic imaging or neoplastic changes in a tissue biopsy. Examples of predictive markers include microsatellite instability which is commonly used to predict response to immune-checkpoint-inhibitor therapy in colorectal cancer (Marcus et al., 2019), and KRAS mutations used to indicate resistance to anti-EGFR treatment (Van Cutsem et al., 2009). Prognostic markers forecast risks associated with clinical outcomes such as survival, recurrence, or disease progression. Such prognostic markers range from tumor grade and stage to genomic and transcriptomic assays such as Oncotype DX and Prosigna (PAM50), often used to estimate recurrence and survival likelihood (Paik et al., 2004). Despite the vital role of biomarkers, patients with similar profiles can exhibit diverse outcomes, treatment responses (Shergalis et al., 2018), recurrence rates (Roy et al., 2015), or treatment toxicity (Kennedy and Salama, 2020), while the underlying reasons for such dichotomies largely remain unknown. There is a crucial need to identify novel and more-specific biomarkers. Modern cancer centers acquire a cornucopia of data over the course of a patient's diagnosis and treatment trajectory, ranging from radiology, histology, clinical and laboratory tests, to familial and patient histories, with each modality providing additional insights on the patient state. A holistic framework integrating complementary information and clinical context from diverse data sources would enable discovery of new, highly-specific biomarkers, paving the path to the next generation of personalized medicine, as illustrated in Figure 1. An analysis of possible correlation and patterns across diverse data modalities can easily become too complex during subjective analysis, making it an attractive application for AI-methods (Boehm et al., 2022). The capacity of AI models to leverage diverse complementary information from multimodal data and identify predictive features within and across modalities allows for automated and objective exploration and discovery of novel biomarkers. Additionally, AI can identify accessible surrogates for existing, but highly-specialize yet expensive markers, to facilitate the spread of advanced targeted therapies and large-scale population screenings.

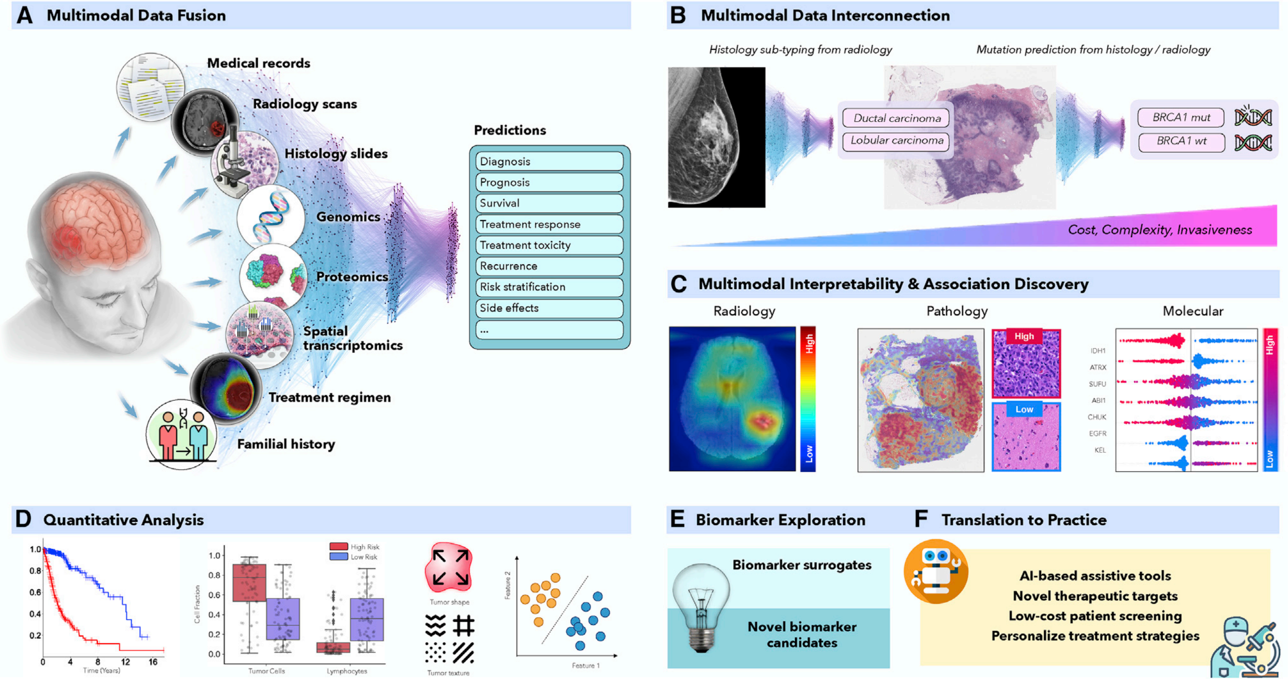

Figure 1. AI-driven multimodal data integration

(A and C–F) (A) AI models can integrate complementary information and clinical context from diverse data sources to provide more accurate outcome predictions. The clinical insights identified by such models can be further elucidated through (C) interpretability methods and (D) quantitative analysis to guide and accelerate the discovery of new biomarkers or therapeutic targets (E and F).

(B) AI can reveal novel multimodal interconnections, such as relations between certain mutations and changes in cellular morphology or associations between radiology findings and histology tumor subtypes or molecular features. Such associations can serve as non-invasive or cost-efficient alternatives to existing biomarkers to support large-scale patient screening (E and F).

Historically, the biomarker discovery process has typically involved the examination of potentially informative qualitative features (such as tissue morphology) or quantitative measurements (such as genomic, transcriptomic alterations) and their association with clinical endpoints. For instance, standardized morphologic assesment pipelines such as the the Nottingham grading system in breast cancer (Rakha et al., 2008) and the Gleason grading in prostate cancers (Epstein et al., 2016) was determined through dedicated examination of thousands of histopathology slides, revealing associations between morphological features and patient outcome. Although the identification of each new biomarker represents a milestone in oncology, this process faces several challenges. Manual assessment is time and resource intensive, often without the possibility of translating observations from one cancer model to another. Morphologic cancer assessment is often qualitative, with substantial interrater variability, which hinders reproducibility and contributes to inconsistent outcomes in clinical trials. Given the large complexity of medical data, current biomarkers are mostly unimodal. However, constraining the biomarkers to a single modality can significantly reduce their clinical potential. For instance, glioma patients with similar genetic or histology profiles can have diverse outcomes caused by macroscopic factors, such as a tumor location preventing full resection and irradiation or disruption of the blood-brain barrier, altering the efficacy of drug delivery (Miller, 2002).

Over the past years, artificial intelligence (AI) and in particular representation learning methods have demonstrated great performance in many clinically relevant tasks inclusing tasks that are often not trivial for human observers (Bera et al., 2019; Lu et al., 2021). AI models are able to integrate complementary information and clinical context from diverse data sources to provide more accurate patient predictions (Figure 1A) (Hosny et al., 2018). The clinical insights identified by successful models can be further elucidated through interpretability methods and quantitative analysis to guide and accelerate the discovery of new biomarkers (Figures 1C and 1D). Similarly, AI models can discover associations across multiple modalities, such as relations between certain mutations and specific changes in cellular morphology (Coudray et al., 2018) or associations between radiology findings and histology-specific tumor subtypes (Ferreira-Junior et al., 2020; Hyun et al., 2019) or molecular features (Yan et al., 2021) (Figure 1B). Such associations can identify accessible or non-invasive alternatives for existing biomarkers to support large-scale population screenings or selection of patients for clinical trials (Figures 1E and 1F). In this review, we summarize AI methods and strategies for multimodal data fusion, outline prospective on AI driven exploration through multimodal associations and interpretability methods, and conclude with directions for AI adoption in precision oncology.

AI METHODS IN ONCOLOGY

AI methods can be categorized as supervised, weakly supervised, or unsupervised. To highlight the concepts specific to

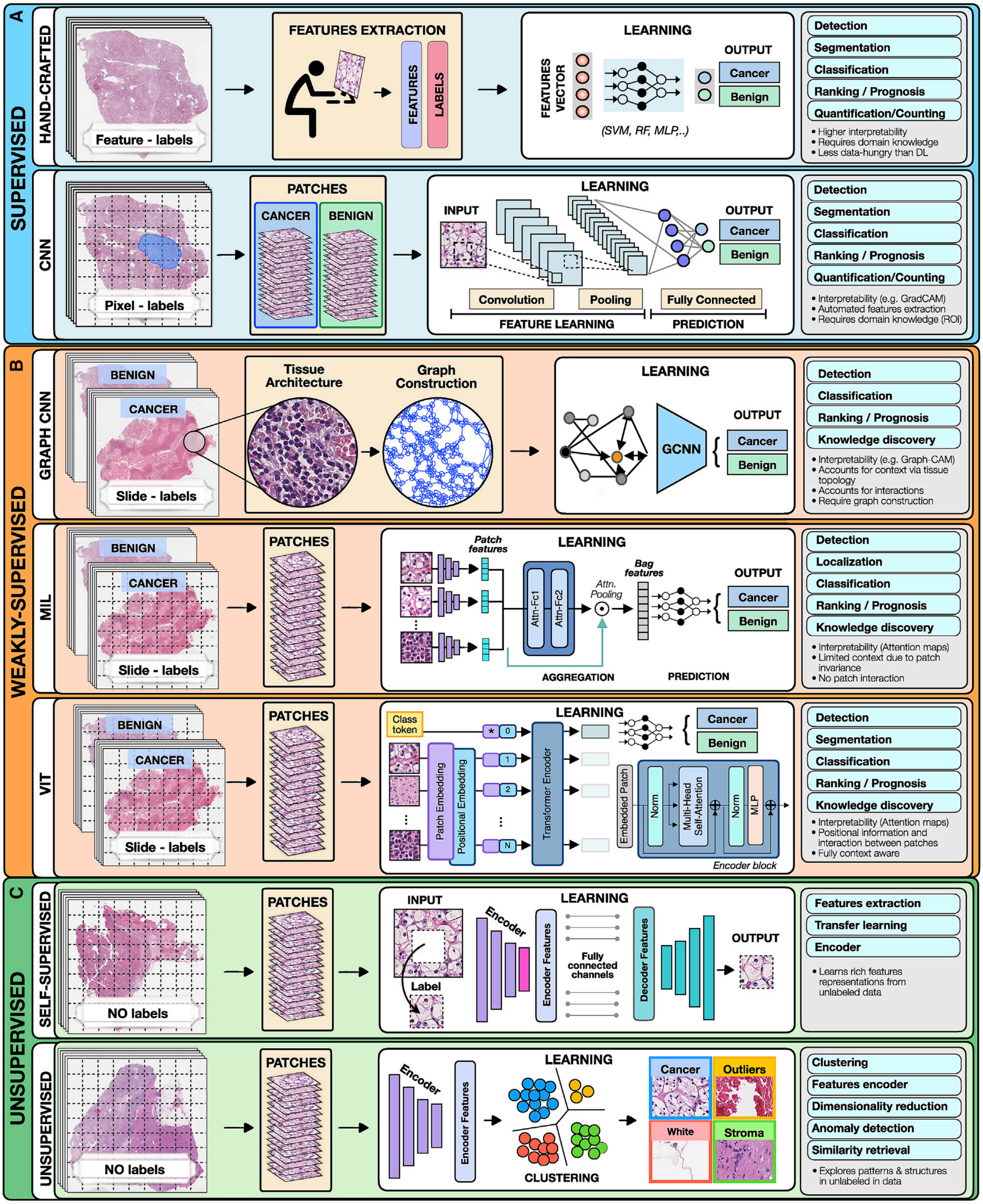

Figure 2. Overview of AI methods
(A) Supervised methods use strong supervision whereby each data point (e.g., feature or image patch) is assigned a label.
(B) Weakly supervised methods allow one to train the model with weak, patient-level labels, avoiding the need for manual annotations.
(C) Unsupervised methods explore patterns, subgroups, and structures in unlabeled data. For comparison, all methods are illustrated on a binary cancer detection task.

each category we present all methods in the framework of computer vision as applied to digital pathology (Figure 2).

Supervised methods

Supervised methods map input data to predefined labels (e.g., cancer/non-cancer) using annotated data points such as digitized slides with pixel-level annotations, or radiology images with patient outcome. Examples of fully supervised methods include hand-crafted and representation learning methods.

Hand-crafted methods

These methods take as input a set of predefined features (e.g., cell shape or size) extracted from the data before the training, not the data themselves. The training is performed with standard machine-learning (ML) models, such as random forest (RF), support-vector machine (SVM), or multilayer perceptron (MLP) (Bertsimas and Wiberg, 2020) (Figure 2). Since the feature extraction is not part of the learning process, the models typically have simpler architecture, lower computation cost, and may require less training data than DL models. An additional benefit is a high level of interpretability, since the predictive features can be related to the data. On the other hand, the feature extraction is time consuming and can translate human bias to the models. A downside is that manual feature extraction or engineering limits the models ability to features already known and understood by humans and prevents the utility and downstream discovery of new relevent features. Moreover, human perception cannot be easily captured by a set of mathematical operators, often leading to simpler features. Since the features are usually tailored to the specific disease, the models cannot be easily translated to other tasks or malignancies. Despite the popularity of DL methods, in many applications the hand-crafted methods are sufficient and preferred due to their simplicity and ability to learn from smaller datasets.

Representation learning methods

Representation learning methods such as deep learning (DL) are capable of learning rich feature representations from the raw data without the need for manual feature engineering. Here we focus on convolutional neural networks (CNNs), the most common DL strategy for image analysis. In CNNs the predictive features are not defined, and the model learnins which concepts and features are useful for explaining relations between inputs and outputs. For instance, in Figure 2, each training whole-slide image (WSI) is manually annotated to outline the tumor region. The WSI is then partitioned into rectangular patches and each patch is assigned with a label, "cancer" or "no-cancer," determined by the tumor annotation. The majority of CNNs have similar architectures, consisting of alternating convolutional, pooling, and non-linear activation layers, followed by a small number of fully connected layers. A convolution layer serves as a feature extractor, while the subsequent pooling layer condenses the features into the most relevant ones. The non-linear activation function allows the model to explore complex relations across features. Fully connected layers then perform the end task, such as classification. The main strength of CNNs is their ability to extract rich feature representations from raw data, resulting in lower preprocessing cost, higher flexibility, and often superior performance over hand-crafted models. The potential limitations come from the model's reliance on pixel-level annotations, which are time intensive and might be affected by interrater variability and human bias. Moreover, predictive regions for many clinical outcomes, such as survival or treatment resistance, may be unknown. CNNs are also often criticized for their lack of interpretability, while we are able to often examine regions used by the model to make predictive determinations, the overall feature representations remain abstract. Despite these limitations, CNNs come with impressive performance, contributing to widespread usage in many clinically relevent applications.

Weakly supervised methods

Weakly supervised learning is a sub-category of supervised learning with batch annotations on large clusters of data essentially representing a scenario where the supervisory signal is weak compared to the amount of noise in the dataset. A common example of the utility of weak supervision is detection of small tumor regions in a biopsy or resection in a large gigapixel whole slide image with labels at the level of the slide or case. Weakly supervised methods allow one to train models with weak, patient-level labels (such as diagnosis or survival), avoiding the need for manual data annotations. The most common weakly supervised methods include graph convolutional networks (GCNs), multiple-instance learning (MIL), and vision transformers (VITs).

Graph convolutional networks

Graphs can be used to explicitly capture structure within data and encode relations between objects making them ideal for analysis of tissue biospy images. A graph is defined by nodes connected by edges. In histology, a node can represent a cell, an image patch, or even a tissue region. Edges encode spatial relations and interactions between nodes (Zhang et al., 2019). The graph, combined with the patient-level labels, is processed by a GCN (Ahmedt-Aristizabal et al., 2021), which can be seen as a generalization of CNNs that operate on unstructured graphs. In GCNs, feature representations of a node are updated by aggregating information from neighboring nodes. The updated representations then serve as input for the final classifier (Figure 2). GCNs can incorporate larger context and spatial tissue structure as compared to a conventional deep models for digital pathology which patch the image into small regions which remain mutually exclusive. This can be beneficial in tasks where the spatial context spans beyond the scope of a single patch (e.g., Gleason score). On the other hand, the interdependence of the nodes in GCNs comes with higher training costs and memory requirements, since the nodes cannot be processed independently.

Multiple-instance learning

MIL is a type of weakly supervised learning where multiple instances of the input are not individually labeled and the supervisory signal is only available collectively for a set of instances commonly reffered to as a bag (Carbonneau et al., 2018; Cheplygina et al., 2019) The label of a bag is assumed positive if there is at least one positive instance in the bag. The goal of the model is to predict the bag label. MIL models comprise three main modules: feature learning or extraction, aggregation, and prediction. The first module is used to embed the images or other higher dimentional data into lower-dimensional embeddings this module can be trained on the fly (Campanella et al., 2019) or a pre-trained encoder from supervised or self-supervised learning can be used to reduce training time and data-efficiency (Lu et al., 2021). The instance-level embeddings are aggregated to create patient-level representations, which serve as input for the final classification module. A commonly used aggrigation stratergy is attention-based pooling, (Ilse et al., 2018), where two fully connected networks are used to learn the relative importance of each instance (Ilse et al., 2018). The patch-level representations, weighted by the corresponding attention score, are summed up to build the patient-level representation. The attention scores can be also be used in understanding the predictive basis of the model (see "multimodal interpretability" for additional details). In large scale medical datasets fine annotations are often not available which makes MIL an ideal approach for training deep models, there are several recent examples in cancer pathology (Campanella et al., 2019; Lu et al., 2021a,b) and genomics (Sidhom et al., 2021).

Vision transformers

VITs (Dosovitskiy et al., 2020; Vaswani et al., 2017) are a type of attention-based learning which allows for the model to be fully context aware. In contrast to MIL, where patches are assumed independent and identically distributed, VITs account for correlation and context among patches. The main components of VITs include positional encoding, self-attention, and multihead self-attention. Positional encoding learns the spatial structure of the image and the relative distances between patches. The self-attention mechanism determines the relevance of each patch while also accounting for the context and contributions from the other patches. Multihead self-attention simultaneously deploys multiple self-attention blocks to account for different types of interactions between the patches and combines them into a single self-attention output. A typical VIT architecture is shown in Figure 2. A WSI is converted into a series of patches, each coupled with positional information. Learnable encoders map each patch and its position into a single embedding vector, referred to as a token. An additional tokens is introduced for the classification task. The class token together with the patch tokens is fed into the transformer encoder to compute multihead self-attention and output the learnable embeddings of patches and the class. The output class token serves as a slide-level representation used for the final classification. The transformer encoder consists of several stacked identical blocks. Each block includes multihead self-attention and MLP, along with layer normalization and residual connections. The positional encoding and multiple self-attention heads allow one to incorporate spatial information, increase the context and robustness (Li et al., 2022; Shamshad et al., 2022) of VIT methods over other methods. On the other hand, VITs tend to be more data hungry (Dosovitskiy et al., 2020), a limitation that the machine learning community is actively working to overcome.

Weakly supervised methods offer several benefits. The liberation from manual annotations reduces the cost of data preprocessing and mitigates the bias and interrater variability. Consequently, the models can be easily applied to large datasets, diverse tasks, and also situations where the predictive regions are unknown. Since the models are free to learn from the entire scan, they can identify predictive features even beyond the regions typically evaluated by pathologists. The great performance demonstrated by weakly supervised methods suggests that many tasks can be addressed without expensive manual annotations or hand-crafted features.

Unsupervised methods

Unsupervised methods explore structures, patterns, and subgroups in data without relying on any labels. These include self-supervised and fully unsupervised strategies.

Self-supervised methods

Self-supervised methods aim to learn rich feature representations from within data by posing the learning problem as a task the ground truth for which is defined within the data. Such encoders are often used to obtain high quality lower dimentional embeddings of complex high dimentional datasets for making downstream tasks more efficient interms of data and training efficiency. For example in pathology images self-supervised methods exploit available unlabeled data to learn high-quality image features and then transfer this knowledge to supervised models. To achieve this, supervised methods such as CNNs are used to solve various pretext tasks (Jing and Tian, 2019) for which the labels are generated automatically from the data. For instance, a patch can be removed from an image and a deep network is trained to predict the missing part of the image from its surroundings, using the actual patch as a label (Figure 2). The patch prediction has no direct clinical relevance, but it guides the model to learn general-purpose features of image characteristics, which can be beneficial for other practical tasks. The early layers of the network are usually capture general image features, while the later layers pick features relevant for the task at hand. The later layers can be excluded, while the early layers serve as feature extractors in for supervised models (i.e., transfer learning).

Unsupervised feature analysis

These methods allow for exploring structure, similarity and common features across data points. For example, using embeddings from a pre-trained encoder one could extract features from a large dataset of diverse patients and cluster said embeddings to find common features across the entire patient cohorts. The most common unsupervised methods include clustering and dimensionality reduction. Clustering methods (Rokach and Maimon, 2005) partition data into subgroups such that the similarities within the subgroup and the separation between subgroups are maximized. Although the output clusters are not task specific, they can reveal different cancer subtypes or patient subgroups. The aim of dimensionality reduction is to obtain low-dimensional representation capturing the main characteristics and correlations in the data.

MULTIMODAL DATA FUSION

The aim of multimodal data fusion is to extract and combine complementary contextual information across different modalities for better decision-making (Zitnik et al., 2019). This is of particular relevance in medicine, where similar findings in one modality may have diverse interpretations in combination with other modalities (Iv et al., 2021). For instance, an *IDH1* mutation status or histology profile alone is insufficient for explaining the variance in patient outcomes, whereas the combination of both has been recently used to redefine the WHO classification of diffuse glioma (Louis et al., 2016). AI offers an automated and objective way to incorporate complementary information and clinical context from diverse data for improved predictions. Multimodal data-driven AI models can also utilize

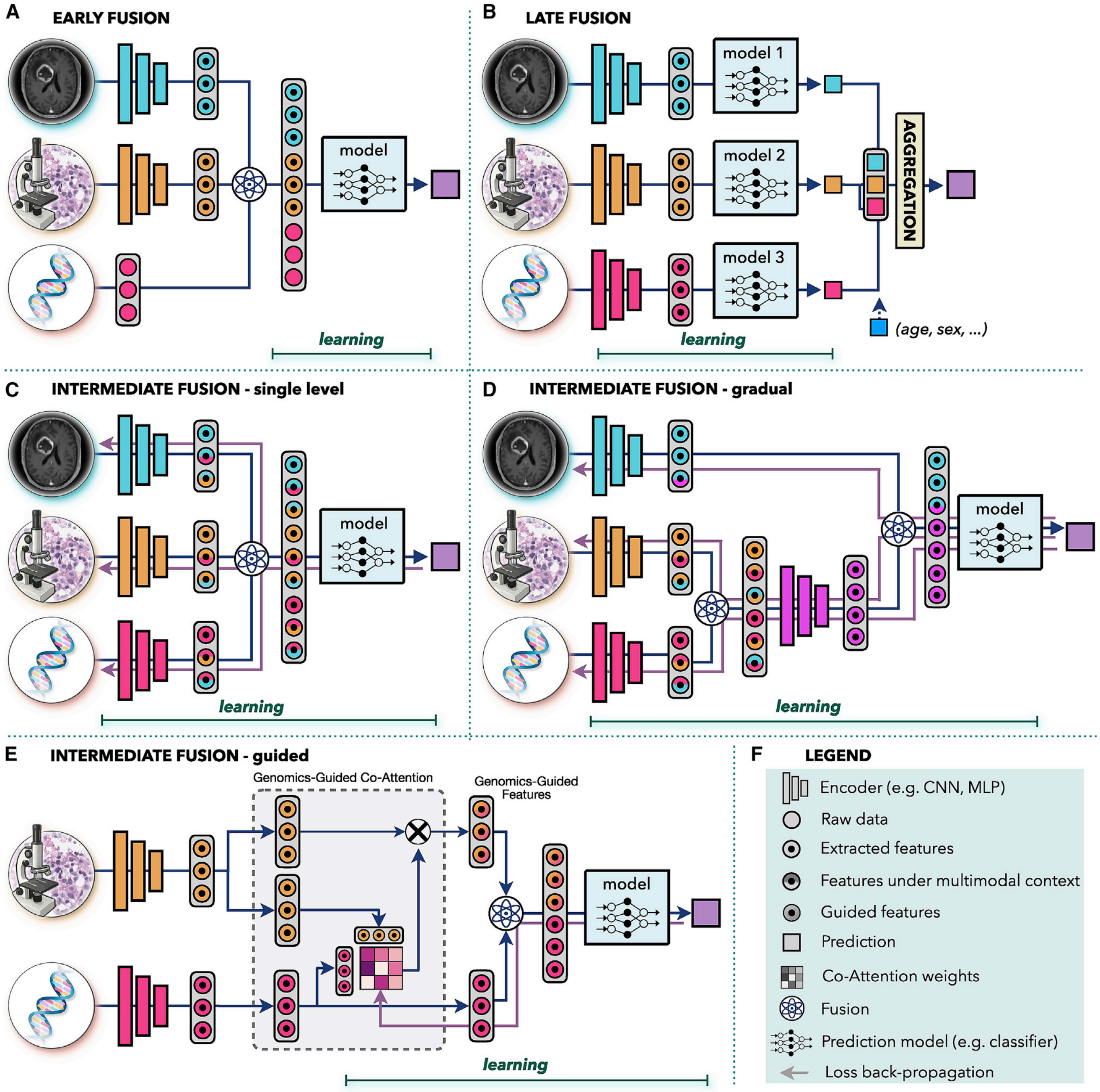

Figure 3. Multimodal data fusion

(A) Early fusion builds a joint representation from raw data or features at the input level, before feeding it to the model.

(B) Late fusion trains a separate model for each modality and aggregates the predictions from individual models at the decision level.

(C–E) In intermediate fusion, the prediction loss is propagated back to the feature extraction layer of each modality to iteratively learn improved feature representations under the multimodal context. The unimodal data can be fused (C) at a single level or (D) gradually in different layers.

(E) Guided fusion allows the model to use information from one modality to guide feature extraction from another modality.

(F) Key for the symbols used.

complementary and supplementary information in modalities; if unimodal data are noisy or incomplete, supplementing redundant information from other modalities can improve the robustness and accuracy of the predictions. AI-driven data fusion strategies (Baltrušaitis et al., 2018) can be divided as early, late, and intermediate (see Figure 3).

Early fusion

Early fusion integrates information from all modalities at the input level before feeding it into a single model. The modalities can be represented as raw data, hand crafted, or deep features. The joint representation is built through operations such as vector concatenation, element-wise sum, element-wise multiplication

(Hadamard product), or bilinear pooling (Kronecker product) (Huang et al., 2020; Ramachandram and Taylor, 2017). In early fusion, only one model is trained, which simplifies the design process. However, it is assumed that the single model is well suited to all modalities. Early fusion requires a certain level of alignment or synchronization between the modalities. Although this is more obvious in other domains, such as synchronization of audio and visual signals in speech recognition, it is also relevant in clinical settings. If the modalities come from significantly different time points, such as pre- and postinterventions, then early fusion might not be an appropriate choice.

Applications of early fusion include integration of similar modalities such as multimodal, multiview ultrasound images for breast cancer detection (Qian et al., 2021) or fusion of structural computed tomography (CT) and/or MRI data with metabolic positron emission tomography (PET) scans for cancer detection (Le et al., 2017), treatment planning (Lipková et al., 2019), or survival prediction (Nie et al., 2019). Other examples include fusion of imaging data with electronic medical records (EMRs), such as integration of dermoscopic images and patient data for skin lesion classification (Yap et al., 2018), or fusion of a cervigram and EMRs for cervical dysplasia diagnosis (Xu et al., 2016). Several studies investigate the correlation between changes in gene expression and tissue morphology, integrating genomics data with histology and/or radiology images for cancer classification (Khosravi et al., 2021), survival (Chen et al., 2020b, 2021c), and treatment response (Feng et al., 2022; Sammut et al., 2022) prediction.

Late fusion

Late fusion, also known as decision-level fusion, trains a separate model for each modality and aggregates the predictions from individual models for the final prediction. The aggregation can be performed by averaging, majority voting, Bayes-based rules (Ramanathan et al., 2022), or learned models such as MLP. Late fusion allows one to use a different model architecture for each modality and does not pose any constraints on data synchronization, making it suitable for systems with large data heterogeneity or modalities from different time points. In cases of missing or incomplete data, late fusion retains the ability to make predictions, since each model is trained separately, and aggregations, such as majority voting, can be applied even if a modality is missing. Similarly, inclusion of a new modality can be performed without the need to retrain the full model. Simple covariates, such as age or gender, are often included through late fusion due to its simplicity (see Figure 3B). If the unimodal data do not complement one another or do not have strong interdependencies, late fusion might be preferable thanks to the simpler architecture and smaller number of parameters compared with other fusion strategies. This is also beneficial in situations with limited data. Furthermore, errors from individual models tend to be uncorrelated, resulting in potentially lower bias and variance in late-fusion predictions. In situations when information density varies significantly across modalities, predictions from shared representations can be heavily influenced by the most dominant modality. In late fusion, the contribution from each modality can be accounted for in a controlled manner by setting equal or diverse weights per modality in the aggregation step.

Examples of late fusion include integration of imaging data with non-imaging inputs, such as fusion of MRI scans and PSA blood tests for prostate cancer diagnosis (Reda et al., 2018), integration of histology scans and patient gender for inferring origin of metastatic tumors (Lu et al., 2021), fusion of genomics and histology profiles for survival prediction (Chen et al., 2021c; Shao et al., 2019), combination of pretreatment MRI or CT scans with EMRs for chemotherapy response prediction (Joo et al., 2021), and survival estimation (Nie et al., 2016).

Intermediate fusion

This is a strategy wherein the loss from the multimodal model propagates back to the feature extraction layer of each modality to iteratively improve feature representations under the multimodal context. For comparison, in early and late fusion, the unimodal embeddings are not affected by the multimodal information. Intermediate fusion can combine individual modalities at different levels of abstractions. Moreover, in systems with three or more modalities the data can be fused either all at once (Figure 3C) or gradually across different levels (Figure 3D). The intermediate single-level fusion is similar to early fusion; however, in early fusion the unimodal embeddings are not affected by the multimodal context. Gradual fusion allows one to combine data from highly correlated channels at the same level, forcing the model to consider the cross-correlations between specific modalities, followed by fusion with less correlated data in later layers. For instance, in Figure 3D, genomics and histology data are fused first, to account for the interplay between mutations and changes in the tissue morphology, while the relation with the macroscopic radiology data is considered in the later layer. Gradual fusion has shown improved performance over single-level fusion in some applications (Joze et al., 2020; Karpathy et al., 2014). Lastly, guided-fusion allows model to use informaiton from one modality to guide feature extraction from another modality. For instance, in Figure 2E, genomics information guides the selection of histology features. The motivation is that different tissue regions might be relevant in the presence of specific mutations. Guided fusion learns co-attention scores that reflect the relevance of different histology features in the presence of specific molecular information. The co-attention scores are learned with the multimodal model, where the genomics feature and the corresponding genomics-guided histology features are combined for the final model predictions.

Examples of intermediate fusion include integration of diverse imaging modalities, such as the fusion of PET and CT scans in lung cancer detection (Kumar et al., 2019), fusion of MRI and ultrasound images in prostate cancer classification (Sedghi et al., 2020), or combination of multimodel MRI scans in glioma segmentation (Havaei et al., 2016). Fusion of diverse multiomics data was used for cancer subtyping (Liang et al., 2014) or survival prediction (Lai et al., 2020). Genomics data have been used in tandem with histology (Vale-Silva and Rohr, 2021) or mammogram (Yala et al., 2019) images for improved survival prediction. Guided fusion of different radiology modalities was used to improve segmentation of liver lesions (Mo et al., 2020) and anomalies in breast tissue (Lei et al., 2020). EMRs were used to guide feature extraction from dermoscopic (Zhou and Luo, 2021) and mammography (Vo et al., 2021) images to improve detection and

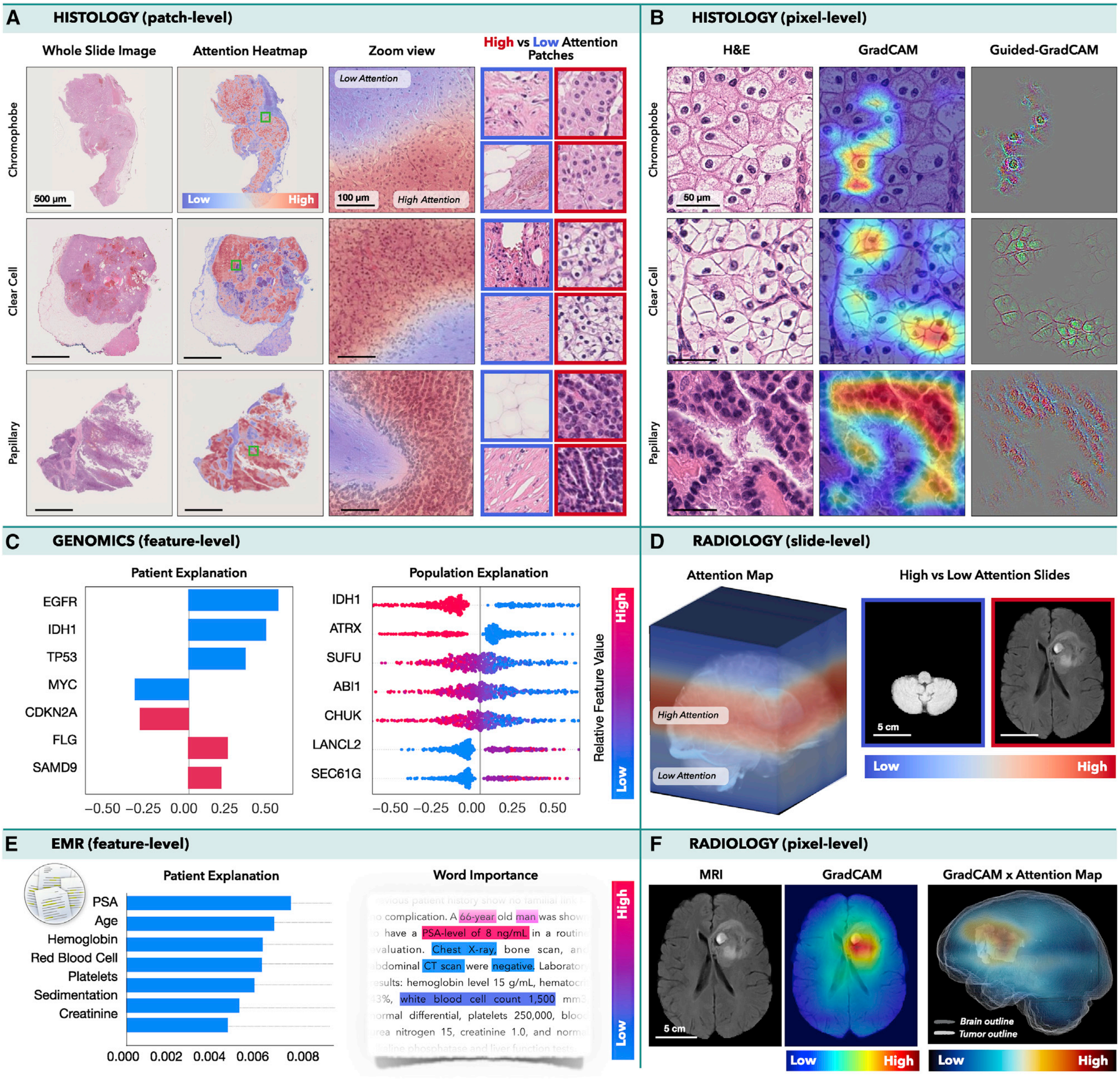

Figure 4. Multimodal interpretability and introspection

(A and B) Histology: an MIL model was trained to classify subtypes of renal cell carcinoma in WSIs, while CNN was trained to perform the same task in image patches. (A) Attention heatmaps and patches with the lowest and highest attention scores. (B) GradCAM attributions for each class.

(C and E) Integrated gradient attributions can be used to analyze (C) genomics or (E) EMRs. The attribution magnitude corresponds to the importance of each feature, and direction indicates feature impact toward low (left) vs. high (right) risk. The color specifies the value of the input features: copy number gain and presence of mutation are shown in red, while blue is used for copy number loss and wild-type status. (E) Attention scores can be used to analyze the importance of words in the medical text.

(D and F) Radiology: an MIL model was trained to predict survival from MRI scans using axial slides as individual instances. (D) Attention heatmaps mapped into the 3D MRI scan and slides with the highest and lowest attention. (F) GradCAM was used to obtain pixel-level interpretability in each MRI slide. A 3D pixel-level interpretability is computed by weighting the slide-level GradCAM maps by the attention score of the respective slide.

classification of lesions. Chen et al. (Chen et al., 2021b) used genomics information to guide selection of histology features for improved survival prediction in multiple cancer types.

There is no conclusive evidence that one fusion type is ultimately better than the others, as each type is heavily data and task specific.

MULTIMODAL INTERPRETABILITY

Interpretability and model introspection is a crucial component of AI development, deployment, and validation. With the ability of AI models to learn abstract feature representations, there is concern that the models might use spurious shortcuts for predictions, instead of learning clinically relevant aspects. Such models might fail to generalize when presented with new data or discriminate against certain populations (Banerjee et al., 2021; Chen et al., 2021a). On the other hand, the models can discover novel and clinical relevant insights. Here we present a brief overview of different methods used for model introspection in oncology (Figure 4), more technical details can be found in a recent review (Arrieta et al., 2020). It is worth indicating that these methods allow us to introspect parts of the data deemed important by the model in making predictive determinations yet the feature representation itself remains abstract.

Histopathology

In histopathology, VITs or MIL can reveal the relative importance of each image patch for the model predictions. Depending on the model architecture attention or probability scores can be mapped to obtain slide-level attention heatmaps as shown in Figure 4A, where an MIL model was trained to classify cancer subtypes in WSIs. Although no manual annotations were used, the model learned to identify morphology specific for each cancer type and to discriminate between normal and malignant tissues. Class activation methods (CAMs), such as GradCAM (Selvaraju et al., 2017) or GradCAM++ (Chattopadhay et al., 2018), allow one to determine the importance of the model inputs (e.g., pixels) by computing how the changes in the inputs affect the model outputs for each prediction class. GradCAM is often used in tandem with the guided-backpropagation method, the so-called guided-GradCAM (Selvaraju et al., 2016), where the guided backpropagation determines the pixel-level importance inside the predictive regions specified by the GradCAM. This is illustrated in Figure 4B, where a CNN was trained to classify cancer subtypes in image patches. For comparison, in the attention methods, the importance of each instance is determined during the training, while the CAM-based methods are model agnostic, i.e., independent of the model training.

Radiology

In radiology, the interpretability methods are similar to those used in histology. The attention scores can reflect the importance of slides in a 3D scan. For instance, in Figure 4D, an MIL model was trained to predict survival in glioma patients (Zhuang et al., 2022). The model considered the 3D MRI scan as a bag, where the axial slides are modeled as individual instances. Even in the absence of manual annotations, the model placed high attention to the slides with tumor, while low attention was assigned to healthy tissue. CAM-based methods can be consequently deployed to localize the predictive regions within individual slides (Figure 4F).

Molecular data

Molecular data can be analyzed by the integrated gradient method (Sundararajan et al., 2017), which computes attribution values indicating how changes in specific inputs affect the model outputs. For the regression tasks, such as survival analysis, the attribution values can reflect the magnitude of the importance as well as the direction of the impact: features with positive attribution increase the predicted output (i.e., higher risk), while features with negative attribution reduce the predictive values (i.e., lower risk). At the patient level, this is visualized as a bar plot, where the y axis corresponds to the specific features (ordered by their absolute attribution value) and the x axis shows the corresponding attribution values. At the population level, the attribution plots depict the distribution of the attribution scores across all subjects. Figure 4C shows the attribution plots for most important genomics features used for survival prediction in glioma patients (Chen et al., 2021c). Other tabular data, such as hand-crafted features or values obtained from EMR, can be interpreted in the same way. EMRs can be also analyzed by natural language processing (NLP) methods, such as transformers, where the attention scores determine the importance of specific words in the text (Figure 4E).

Multimodal models

In multimodal models, the attribution plots can also determine the contribution of each modality toward the model predictions. All previously mentioned methods can be used in multimodal models to explore interpretability within each modality. Moreover, shifts in feature importance under unimodal and multimodal settings can be investigated to analyze the impact of the multimodal context.

The interpretability methods usually come without any accuracy measures, and thus it is important not to overinterpret them. While CAM- or attention-based methods can localize the predictive regions, they cannot specify which features are relevant, i.e., they can explain *where* but not *why*. Moreover, there is no guarantee that all high-attention/attribution regions carry clinical relevance. High scores just mean that the model has considered these regions more important than others.

MULTIMODAL DATA INTERCONNECTION

The aim of multimodal data interconnection is to reveal associations and shared information across modalities. Such associations can provide new insights into cancer biology and guide the discovery of novel biomarkers. Although there are many approaches for data exploration, here we illustrate a few possible directions (Figure 5).

Morphologic associations

Malignant changes often propagate across different scales; oncogenic mutations can affect cell behavior, which in turn reshapes tissue morphology or the tumor microenvironment visible in histology images. Consequently, the microscopic changes might have an impact on tumor metabolic activity and macroscopic appearance detectable by PET or MRI scans. The feasibility of AI methods to identify associations across modalities was first demonstrated by Coudray et al. (Coudray et al., 2018), who showed that certain mutations in lung cancer can be inferred directly from hematoxylin and eosin (H&E)-stained WSIs. Other studies followed shortly, predicting the mutation status from WSIs in liver (Chen et al., 2020a), bladder (Loeffler et al., 2021), colorectal (Jang et al., 2020), and thyroid cancer (Tsou and Wu, 2019), as well as pan-cancer pan-mutation studies attempting to predict any genetic alternation in any tumor type (Fu et al., 2020; Kather et al., 2020). Additional molecular biomarkers, such as gene expression (Anand et al., 2020; Binder et al., 2021; Schmauch et al., 2020), hormone-receptor status (Naik et al., 2020), tumor mutational burden (Jain and Massoud, 2020), and microsatellite instability (Cao et al., 2020; Echle et al.,

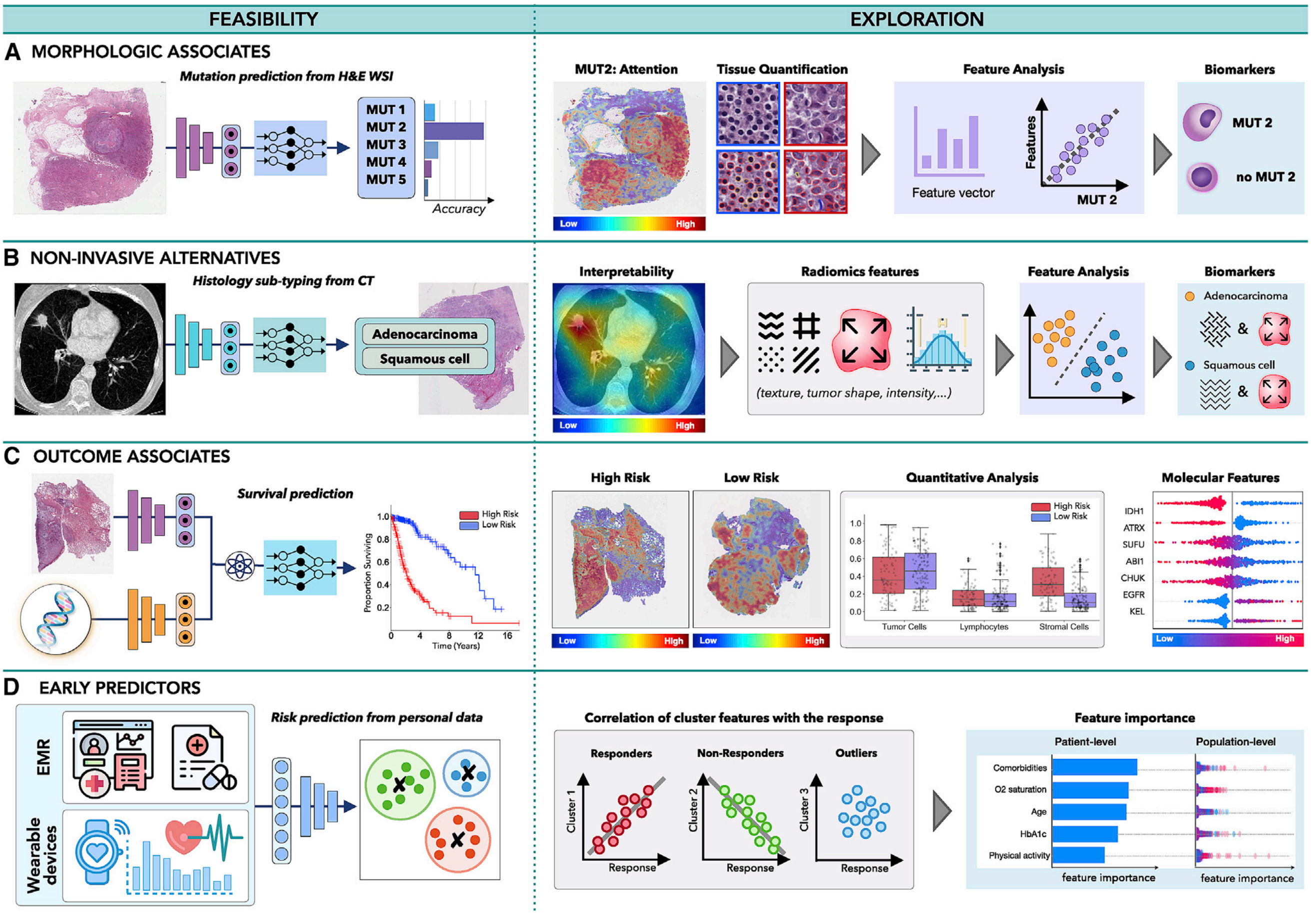

Figure 5. Multimodal data interconnection

(A and B) AI can identify associations across modalities, such as (A) the feasibility of inferring certain mutations from histology or radiology images or (B) the relation between non-invasive and invasive modalities, such as prediction of histology subtype from radiomics features.

(C) The models can uncover associations between clinical data and patient outcome, contributing to the discovery of predictive features within and across modalities.

(D) Information acquired by EMRs or wearable devices can be analyzed to identify risk factors related to cancer onset or uncover patterns related with treatment response or resistance, to support early interventions.

2020), have also been inferred from WSIs (Murchan et al., 2021). In radiology, AI models have predicted *IDH* mutation and *1p/19q* co-deletion status from preoperative brain MRI scans (Bangalore Yogananda et al., 2020; Yogananda et al., 2020) and *BRCA1* and *BRCA2* mutational status from breast mammography (Ha et al., 2017) and MRI (Vasileiou et al., 2020) scans, while *EGFR* and *KRAS* mutations have been detected from CT scans in lung (Wang et al., 2019) and colorectal (He et al., 2020) cancer.

By discovering the presence of morphological associations across modalities, AI models can enhance exploratory studies and reduce the search space for possible biomarker candidates. For instance, in Figure 5A, AI has revealed that one of the studied mutations can be reliably inferred from WSI. Although the predictive features used by the model might be unknown, interpretability methods can provide additional insights. Attention heatmaps can reveal tissue regions relevant for the prediction of the specific mutation. Distinct tissue structures and cell types in the regions with the high- and low-attention scores can be identified, and their properties, such as nucleus shape or volume, can be further extracted and analyzed. Clustering or dimensionality reduction methods can be deployed to examine the promising features, potentially revealing associations between mutation status and distinct morphological features. The identified morphological associates can serve as cost-efficient biomarker surrogates to support screening in low- to middle-income settings or reveal new therapeutic targets.

Non-invasive alternatives

Similarly, AI can discover relationships between non-invasive and invasive modalities. For instance, AI models were used to predict histology subtypes or grades from radiomics features in lung (Sha et al., 2019), brain (Lasocki et al., 2015), liver (Brancato et al., 2022), and other cancers (Blüthgen et al., 2021). The predictive image regions can be further analyzed to identify textures and patterns with possible diagnostic values (see Figure 5B), which in turn can serve as non-invasive surrogates for existing biomarkers.

Outcome associates

Benefits of personalized medicine are often limited by the paucity of biomarkers able to explain dichotomies in patient outcomes.

On the other hand, AI models are demonstrating great performance in predicting clinical outcomes, such as survival (Lai et al., 2020), treatment response (Echle et al., 2020), recurrence (Yamamoto et al., 2019), and radiation toxicity (Men et al., 2019), using unimodal and multimodal (Chen et al., 2020b, 2021c; Joo et al., 2021; Mobadersany et al., 2018) data. These works imply the feasibility of AI models to discover relevant prognostic patterns in data, which might be elucidated by interpretability methods. For instance, in Figure 5C, a model is trained to predict survival from histology and genomics data. Attention heatmaps reveal tissue regions related to low- and high-risk patient groups, while the molecular profiles are analyzed through attribution plots. The predictive tissue regions can be further analyzed by examining tissue morphology, cell subtypes, or other human-interpretable data characteristics. Tumor-infiltrating lymphocytes can be estimated through co-localization of tumor and immune cells to specify immune hot and cold tumors. Attribution of specific modalities as well as shifts in feature importance in unimodal vs. multimodal data can be explored to determine the influence of multimodal contextualization.

Such exploration studies have already provided new clinical insights. For instance, Geessink et al. (Geessink et al., 2019) showed that the tumor-to-stroma ratio can serve as an independent prognosticator in rectal cancer, while the ratio of tumor area to metastatic lymph node regions has prognostic value in gastric cancer (Wang et al., 2021). Other morphological features, such as the arrangement of collagen fibers in breast histology (Li et al., 2021) or spatial tissue organization in colorectal tissue (Qi et al., 2021), have been identified as possible biomarkers for aggressiveness or recurrence.

Early predictors

AI can also explore diverse data acquired prior to patient diagnosis to identify potential predictive risk factors. EMRs provide rich information on patient history, medication, allergies, or immunizations, which might contribute to patient outcome. Such diverse data can be efficiently analyzed by AI models to search for distinct patient subgroups (Figure 5D). Identified subgroups can be correlated with different patient outcomes, while attribution plots can identify the relevance of different factors at the patient and population level. Recently, Placido et al. (Placido et al., 2021) showed the feasibility of AI to identify patients with a higher risk of developing pancreatic cancer by exploration of EMR. Similarly, EMRs were used to predict treatment response (Chu et al., 2020) or length of hospital stay (Alsinglawi et al., 2022). The identified novel predictive risk factors can support large-scale population screenings and early preventive care.

Outside of the hospital setting, smartphones and wearable devices offer another great opportunity for real-time and continuous patient monitoring. Changes in the measured values, such as a decrease in patient step counts, have been shown as robust predictors of worse clinical outcome, and increased risk of hospitalization (Low, 2020). Furthermore, the modern wearable devices are continually expanding their functionality, including measurements of temperature, stress levels, or blood-oxygen saturation or electrocardiograms. These measurements can be analyzed in tandem with clinical data to search for risk factors indicating early stages of increased toxicity or treatment resistance, to allow personalized interventions during the course of treatment. Research on personalized monitoring and nanotechnologies is investigating novel directions, such as the detection of patient measurements in sweat (Xu et al., 2019) or ingestible sensors to monitor medication compliance and drug absorption (Weeks et al., 2018). All these novel devices provide useful insights into the patient state, which could be analyzed in a larger clinical context through AI models.

CHALLENGES AND CLINICAL ADOPTION

The path of AI into clinical practice is still laden with obstacles, many of which are amplified in the presence of multimodal data (Van der Laak et al., 2021). While several recent works discuss challenges, such as fairness and dataset shifts (Banerjee et al., 2021; Chen et al., 2021a; Cirillo et al., 2020; Howard et al., 2021; Mehrabi et al., 2021; Zhang et al., 2018), limited interpretability (Adebayo et al., 2018; Linardatos et al., 2020; Reyes et al., 2020), or regulatory guidelines (Cruz Rivera et al., 2020; Topol, 2020; Wu et al., 2021), here we focus on challenges specific to multimodal learning.

Missing data

The challenge of missing data refers to the absence of part of a modality or the complete unavailability of one or more modalities. The missing data affect both the model training and the deployment, since the majority of existing AI models cannot handle missing information. Moreover, the need to train models with complete multimodal data significantly constrains the size of the training datasets. Many multimodal datasets have large scale data missingness for example in the cancer genome atlas (TCGA) one of the largest publicly available multimodal datasets has significant missing data points. The incomplete modalities still contain valuable information, and the inability to deploy them poses a significant limitation. Below we discuss two strategies for handling missing data.

Synthetic data generation

Given the paucity of medical data in general synthetic data is increasingly being used to train, develop and augment AI models (Chen et al., 2021). If part of an image is corrupted, or if specific mutations are not reported, the missing information can be synthesized from the remaining data. If a whole modality is missing, its synthetic version can be derived from existing similar modalities. For instance, de Haan et al. (de Haan et al., 2021) trained a supervised model for translation of H&E stains into special stains, using the special stains as ground truth labels. The model was trained on pairs of perfectly aligned data obtained through re-staining of the same slides. If paired data are not available, unsupervised methods such as cycle generative adversarial networks (GANs) (Zhu et al., 2017) can be used. While synthetic data can improve the performance of detection and classification methods, they are less suitable for outcome prediction or biomarker exploration, where the predictive features are not well understood and thus there is no guarantee that the synthetic data contain the relevant disease characteristics. Moreover, the algorithms can also hallucinate malignant features into the supposedly normal synthetic images (Cohen et al., 2018), which can further hurt prediction results.

Dropout-based methods

Dropout-based methods aim to make models robust to missing information. For instance, Choi and Lee (Choi and Lee, 2019) proposed the EmbraceNet model, which can handle incomplete or missing data during training and deployment. The EmbraceNet model probabilistically selects partial information from each modality and combines it into a single representation vector, which then serves as an input for the final decision model. When missing or invalid data are encountered, they are not sampled; instead, other more complete modalities are used to compensate for the missing data. The probabilistic data selection also has a regularization effect, similar to the dropout mechanism.

Data alignment

To investigate cancer processes across different scales and modalities, a certain level of data alignment is required. This might include alignment of (1) similar or (2) diverse modalities.

Alignment of similar modalities

This method typically involves different imaging modalities of the same system. This is usually achieved through image registration, which is formulated as an optimization problem minimizing the difference between the modalities.

In radiology, rigid anatomical structures can guide the data alignment. For instance, registration of MRI and PET brain scans is usually achieved with high accuracy, even with simple affine registration, thanks to the rigid skull. The situation is more complex in the presence of motion and deformations, e.g., breathing in lung imaging or changes in the body posture between scanning sessions. Alignment of such data usually requires deformable registrations using natural or manually placed landmarks for guidance. A particularly challenging situation is the registration of scans between interventions, e.g., registration of preoperative and postoperative scans, which exhibit lot of non-trivial changes due to tumor resection, response to treatment, or tissue compression (Haskins et al., 2020).

In histology, each stained slide usually comes from a different tissue cut. Even in consecutive tissue cuts there are substantial differences in the tissue appearance caused by changes in the tissue microenvironment or artifacts such as tissue folding, tearing, or cutting (Taqi et al., 2018), which all complicate data alignment. Robust and automated registration of histology images can be challenging (Borovec et al., 2020), and thus many studies deploy non-algorithmic strategies such as clearing and re-staining of the tissue slides (Hinton et al., 2019). A newly emerging direction is stainless imaging, including approaches such as ultraviolet microscopy (Fereidouni et al., 2017), stimulated Raman histology (Hollon et al., 2020), or colorimetric imaging (Balaur et al., 2021).

Alignment of diverse modalities

This refers to the integration of data from different scales, time points, or measurements. Often an acquisition of one modality results in the destruction of the sample, preventing collection of multiple measurements from the same system. For instance, most omics measurements require tissue disintegration, which inevitably affects the possibility of studying relations between cell appearance and corresponding gene expression. Here, cross-modal autoencoders can be used to enable integration and translation between arbitrary modalities. Cross-modal autoencoders (Dai Yang et al., 2021) build a pair of encoder-decoder networks for each modality, where the encoder maps each modality into a lower-dimensional latent space, while the decoder maps it back into the original space. A discriminative objective function is used to match the different modalities in the common latent space. With the shared latent space in place, one can combine an encoder of one modality with the decoder of another modality to align one modality to another one. Dai Yang et al. (2021) demonstrated translations between single-cell chromatin images and RNA-sequencing data. The feasibility and utility of the cross-modal autoencoders are yet to be tested with large scale clinically relevent datasets. However, if proven potent, they hold great potential to address challenges with alignment and harmonization of data from diverse sources.

Transparency and prospective clinical trials

Given the complexity of representation learning-based modern AI methods and the fact that they use abstract feature representations, it is possible that their mechanisms will not be fully understood in the near future. However, one may argue that many aspects in medicine are not fully understood, either (Kirkpatrick, 2005). Some of the interpretability methods discussed earlier are capable of indicating regions within data used to make prediction determination yet the actual feature representation remains abstract. And thus, rather than dwelling on the full opacity of AI methods, we should advocate for their rigorous validation under randomized clinical trials, same as is done for other medical devices and drugs (Ghassemi et al., 2021) . Prospective trials will allow us to stress test the models under real-world conditions, compare their performance against standard-of-care and current practice, estimate how clinicians will interact with the AI tool, and find the best way in which the models can enhance, rather than disturb, the clinical workflow. In the case of biomarker surrogates discovered by AI methods, regulation paths similar to "me-too" drugs and devices (Aronson and Green, 2020) could be used to ensure comparable levels of performance. Transparency about study design and the data used are necessary to determine the intended use and conditions under which the model performance has been verified and evaluated (Haibe-Kains et al., 2020). Prospective clinical trials are inevitable to truly demonstrate and quantify the added value of AI models, which will in turn increase trust and motivation of practitioners toward the AI tools.

OUTLOOK AND DISCUSSION

AI has the potential to have an impact on the whole landscape of oncology, ranging from prevention to intervention. AI models can explore complex and diverse data to identify factors related to high risks of developing cancer to support large population screenings and preventive care. The models can further reveal associations across modalities to help identify diagnostic or prognostic biomarkers from easily accessible data to improve patient risk stratification or selection for clinical trials. In a similar way, the models can identify non-invasive alternatives to existing biomarkers to minimize invasive procedures. Prognostic models can predict risk factors or adverse treatment outcomes prior to interventions to guide patient management. Information acquired from personal wearable devices or nanotechnologies could be further analyzed by AI models to search for early signs of treatment toxicity or resistance, with other great application yet to come.

As with any great medical advance, there is a need for rigorous validation and examination via clinical studies, prospective trials to verify the promises made by AI models. The role of AI in advancing the field of oncology is not autonomous; rather, it is a partnership between models and human experience that will drive further progress. AI models come with limitations and challenges; however, these should not intimidate but rather inspire us. With increasing incidence rates of cancer, it is our obligation to capitalize on benefits offered by AI methods to accelerate discovery and translation of advances into clinical practice to serve patients and health care providers.

ACKNOWLEDGMENTS

This work was supported in part by the BWH President's Fund, National Institute of General Medical Sciences (NIGMS) R35GM138216 (to F.M.), Google Cloud Research Grant, Nvidia GPU Grant Program, and internal funds from BWH and MGH Pathology. M.Y.L. was additionally supported by the Tau Beta Pi Fellowship and the Siebel Foundation. T.Y.C. was additionally supported by the NIH National Cancer Institute (NCI) Ruth L. Kirschstein National Service Award, T32CA251062. R.J.C. was additionally supported by the National Science Foundation (NSF) graduate fellowship. The content is solely the responsibility of the authors and does not reflect the official views of the NIH, NSF, NCI or NIGMS.

DECLARATION OF INTERESTS

F.M. and R.J.C. are inventors on a patent related to multimodal learning.

REFERENCES

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., and Kim, B. (2018). Sanity checks for saliency maps. In Advances in Neural Information Processing Systems, *31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds. (Curran Associates, Inc).

Ahmedt-Aristizabal, D., Armin, M.A., Denman, S., Fookes, C., and Petersson, L. (2021). A Survey on Graph-Based Deep Learning for Computational Histopathology (Computerized Medical Imaging and Graphics), p. 102027.

Alsinglawi, B., Alshari, O., Alorjani, M., Mubin, O., Alnajjar, F., Novoa, M., and Darwish, O. (2022). An explainable machine learning framework for lung cancer hospital length of stay prediction. Sci. Rep. *12*, 607–610.

Anand, D., Kurian, N.C., Dhage, S., Kumar, N., Rane, S., Gann, P.H., and Sethi, A. (2020). Deep learning to estimate human epidermal growth factor receptor 2 status from hematoxylin and eosin-stained breast tissue images. J. Pathol. Inform. *11*, 19.

Aronson, J.K., and Green, A.R. (2020). Me-too pharmaceutical products: history, definitions, examples, and relevance to drug shortages and essential medicines lists. Br. J. Clin. Pharmacol. *86*, 2114–2122.

Arrieta, A.B., Dıaz-Rodŕıguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garćıa, S., Gil-López, S., Barredo Arrieta, A., Díaz-Rodríguez, N., et al. (2020). Explainable artificial intelligence (xai): concepts, taxonomies, opportunities and challenges toward responsible ai. Inf. Fusion *58*, 82–115.

Balaur, E., O'Toole, S., Spurling, A.J., Mann, G.B., Yeo, B., Harvey, K., Sadatnajafi, C., Hanssen, E., Orian, J., Balaur, E., et al. (2021). Colorimetric histology using plasmonically active microscope slides. Nature *598*, 65–71.

Baltrušaitis, T., Ahuja, C., and Morency, L.-P. (2019). Multimodal machine learning: a survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. *41*, 423–443.

Banerjee, I., Bhimireddy, A.R., Burns, J.L., Celi, L.A., Chen, L.-C., Correa, R., Dullerud, N., Ghassemi, M., Huang, S.-C., Kuo, P.-C., et al. (2021). Reading race: ai recognises patient's racial identity in medical images. Preprint at arXiv, 2107.10356.

Bangalore Yogananda, C.G., Shah, B.R., Vejdani-Jahromi, M., Nalawade, S.S., Murugesan, G.K., Yu, F.F., Bangalore Yogananda, C.G., Shah, B.R., Vejdani-Jahromi, M., Nalawade, S.S., et al. (2020). A novel fully automated mri-based deep-learning method for classification of idh mutation status in brain gliomas. Neuro Oncol. *22*, 402–411.

Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V., and Madabhushi, A. (2019). Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol *16*, 703–715.

Bertsimas, D., and Wiberg, H. (2020). Machine learning in oncology: methods, applications, and challenges. JCO Clin. Cancer Inform. *4*, 885–894.

Binder, A., Bockmayr, M., Hägele, M., Wienert, S., Heim, D., Hellweg, K., Ishii, M., Stenzinger, A., Hocke, A., Denkert, C., et al. (2021). Morphological and molecular breast cancer profiling through explainable machine learning. Nat. Mach. Intell. *3*, 355–366.

Blüthgen, C., Patella, M., Euler, A., Baessler, B., Martini, K., von Spiczak, J., Schneiter, D., Opitz, I., and Frauen- felder, T. (2021). Computed tomography radiomics for the prediction of thymic epithelial tumor histology, tnm stage and myasthenia gravis. PLoS One *16*, e0261401.

Boehm, K.M., Khosravi, P., Vanguri, R., Gao, J., and Shah, S.P. (2022). Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer *22*, 114–126.

Borovec, J., Kybic, J., Arganda-Carreras, I., Sorokin, D.V., Bueno, G., Khvostikov, A.V., Bakas, S., Chang, E.I.C., Heldmann, S., Kartasalo, K., et al. (2020). Anhir: automatic non-rigid histological image registration challenge. IEEE Trans. Med. Imaging *39*, 3042–3052.

Brancato, V., Garbino, N., Salvatore, M., and Cavaliere, C. (2022). Mri-based radiomic features help identify lesions and predict histopathological grade of hepatocellular carcinoma. Diagnostics *12*, 1085.

Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., and Fuchs, T.J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med *25*, 1301–1309.

Cao, R., Yang, F., Ma, S.-C., Liu, L., Zhao, Y., Li, Y., Wu, D.-H., Wang, T., Lu, W.-J., Cai, W.-J., et al. (2020). Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer. Theranostics *10*, 11080–11091.

Carbonneau, M.-A., Cheplygina, V., Granger, E., and Gagnon, G. (2018). Multiple instance learning: a survey of problem characteristics and applications. Pattern Recogn. *77*, 329–353.

Chattopadhay, A., Sarkar, A., Howlader, P., and Balasubramanian, V.N. (2018). Grad-cam++: generalized gradient- based visual explanations for deep convolutional networks. In In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (IEEE), pp. 839–847.

Chen, M., Zhang, B., Topatana, W., Cao, J., Zhu, H., Juengpanich, S., Mao, Q., Yu, H., and Cai, X. (2020a). Classification and mutation prediction based on histopathology h&e images in liver cancer using deep learning. NPJ Precis. Oncol. *4*, 14–17.

Chen, R.J., Lu, M.Y., Chen, T.Y., Williamson, D.F., and Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng *5*, 493–497.

Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., and Mahmood, F. (2020b). Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis (IEEE Transactions on Medical Imaging).

Chen, R.J., Chen, T.Y., Lipkova, J., Wang, J.J., Williamson, D.F., Lu, M.Y., Sahai, S., and Mahmood, F. (2021a). Algorithm fairness in ai for medicine and healthcare. Preprint at arXiv, 2110.00603.

Chen, R.J., Lu, M.Y., Weng, W.-H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., and Mahmood, F. (2021b). Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4025.

Chen, R.J., Lu, M.Y., Williamson, D.F., Chen, T.Y., Lipkova, J., Shaban, M., Shady, M., Williams, M., Joo, B., Noor, Z., et al. (2021c). Pan-cancer integrative histology-genomic analysis via interpretable multimodal deep learning. Preprint at arXiv, 2108.02278.

Cheplygina, V., de Bruijne, M., and Pluim, J.P.W. (2019). Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Med. Image Anal. *54*, 280–296.

Choi, J.-H., and Lee, J.S. (2019). A robust deep learning architecture for multimodal classification. Inf. Fusion *51*, 259–270.

Chu, J., Dong, W., Wang, J., He, K., and Huang, Z. (2020). Treatment effect prediction with adversarial deep learning using electronic health records. BMC Med. Inform. Decis. Mak. *20*, 139–214.

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Re- menteria, M.J., Chadha, A.S., and Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit. Med. *3*, 1–11.

Cohen, J.P., Luck, M., and Honari, S. (2018). Distribution matching losses can hallucinate features in medical image translation. In In International conference on medical image computing and computer-assisted intervention (Springer), pp. 529–536.

Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., and Tsirigos, A. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat. Med. *24*, 1559–1567.

Cruz Rivera, S., Liu, X., Chan, A.-W., Denniston, A.K., and Calvert, M.J.; SPIRIT-AI and CONSORT-AI Working Group; SPIRIT-AI and CONSORT-AI Steering Group; SPIRIT-AI and CONSORT-AI Consensus Group (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the spirit-ai extension. Nat. Med. *26*, 1351–1363.

Dai Yang, K., Belyaeva, A., Venkatachalapathy, S., Damodaran, K., Katcoff, A., Radhakrishnan, A., Shiv- ashankar, G., and Uhler, C. (2021). Multi-domain translation between single-cell imaging and sequencing data using autoencoders. Nat. Commun. *12*, 1–10.

de Haan, K., Zhang, Y., Zuckerman, J.E., Liu, T., Sisk, A.E., Diaz, M.F.P., Jen, K.-Y., Nobori, A., Liou, S., Zhang, S., et al. (2021). Deep learning-based transformation of h&e stained tissues into special stains. Nat. Commun. *12*, 4884–4913.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. Preprint at arXiv, 2010.11929.

Echle, A., Grabsch, H.I., Quirke, P., van den Brandt, P.A., West, N.P., Hutchins, G.G.A., Heij, L.R., Tan, X., Richman, S.D., Krause, J., et al. (2020). Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology *159*, 1406–1416.e11.

Epstein, J.I., Zelefsky, M.J., Sjoberg, D.D., Nelson, J.B., Egevad, L., Magi-Galluzzi, C., Vickers, A.J., Parwani, A.V., Reuter, V.E., Fine, S.W., and Eastham, J.A. (2016). A contemporary prostate cancer grading system: a validated alternative to the Gleason score. Eur Urol *69*, 428–435.

Feng, L., Liu, Z., Li, C., Li, Z., Lou, X., Shao, L., Wang, Y., Huang, Y., Chen, H., Pang, X., et al. (2022). Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradio- therapy in locally advanced rectal cancer: a multicentre observational study. Lancet Digital Health *4*, e8–e17.

Fereidouni, F., Harmany, Z.T., Tian, M., Todd, A., Kintner, J.A., McPherson, J.D., Borowsky, A.D., Bishop, J., Lechpammer, M., Demos, S.G., and Levenson, R. (2017). Microscopy with ultraviolet surface excitation for rapid slide-free histology. Nat. Biomed. Eng. *1*, 957–966.

Ferreira-Junior, J.R., Koenigkam-Santos, M., Magalhães Tenório, A.P., Faleiros, M.C., Garcia Cipriano, F.E., Fabro, A.T., Näppi, J., Yoshida, H., and de Azevedo-Marques, P.M. (2020). Ct-based radiomics for prediction of histologic subtype and metastatic disease in primary malignant lung neoplasms. Int. J. Comput. Assist. Radiol. Surg. *15*, 163–172.

Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., and Gerstung, M. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat. Cancer *1*, 800–810.

Geessink, O.G.F., Baidoshvili, A., Klaase, J.M., Ehteshami Bejnordi, B., Litjens, G.J.S., van Pelt, G.W., Mesker, W.E., Nagtegaal, I.D., Ciompi, F., and van der Laak, J.A.W.M. (2019). Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. Cell. Oncol. *42*, 331–341.

Ghassemi, M., Oakden-Rayner, L., and Beam, A.L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health *3*, e745–e750.

Ha, S.M., Chae, E.Y., Cha, J.H., Kim, H.H., Shin, H.J., and Choi, W.J. (2017). Association of brca mutation types, imaging features, and pathologic findings in patients with breast cancer with brca1 and brca2 mutations. AJR Am. J. Roentgenol. *209*, 920–928.

Haibe-Kains, B., Adam, G.A., Hosny, A., Khodakarami, F., Waldron, L., Wang, B., McIntosh, C., Goldenberg, A., Kundaje, A., Greene, C.S., and Broderick, T. (2020). Transparency and reproducibility in artificial intelligence. Nature *586*, E14–E16.

Haskins, G., Kruger, U., and Yan, P. (2020). Deep learning in medical image registration: a survey. Mach. Vis. Appl. *31*, 8–18.

Havaei, M., Guizard, N., Chapados, N., and Hemis, Y.B. (2016). Hetero-modal image segmentation. In In Inter- national Conference on Medical Image Computing and Computer-Assisted Intervention (Springer), pp. 469–477.

He, K., Liu, X., Li, M., Li, X., Yang, H., and Zhang, H. (2020). Noninvasive kras mutation estimation in colorectal cancer using a deep learning method based on ct imaging. BMC Med. Imaging *20*, 1–9.

Hinton, J.P., Dvorak, K., Roberts, E., French, W.J., Grubbs, J.C., Cress, A.E., Tiwari, H.A., and Nagle, R.B. (2019). A method to reuse archived h&e stained histology slides for a multiplex protein biomarker analysis. Methods Protoc. *2*, 86.

Hollon, T.C., Pandian, B., Adapa, A.R., Urias, E., Save, A.V., Khalsa, S.S.S., Eichberg, D.G., D'Amico, R.S., Farooq, Z.U., Lewis, S., et al. (2020). Near real-time intraoperative brain tumor diagnosis using stimulated raman histology and deep neural networks. Nat. Med. *26*, 52–58.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., and Aerts, H.J. (2018). Artificial intelligence in radiology. Nat Rev Cancer *18*, 500–510.

Howard, F.M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O.I., Kather, J.N., et al. (2021). The impact of site-specific digital histology signatures on deep learning model accuracy and bias. Nat. Commun. *12*, 4423–4513.

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M.P. (2020). Fusion of medical imaging and elec- tronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit. Med. *3*, 136–139.

Hyun, S.H., Ahn, M.S., Koh, Y.W., and Lee, S.J. (2019). A machine-learning approach using pet-based radiomics to predict the histological subtypes of lung cancer. Clin. Nucl. Med. *44*, 956–960.

Ilse, M., Tomczak, J., and Welling, M. (2018). Attention-based deep multiple instance learning. In In International conference on machine learning (PMLR), pp. 2127–2136.

Iv, W.C.S., Kapoor, R., and Ghosh, P. (2021). Multimodal Classification: Current Landscape, Taxonomy and Future Directions (ACM Computing Surveys (CSUR)).

Jain, M.S., and Massoud, T.F. (2020). Predicting tumour mutational burden from histopathological images using multiscale deep learning. Nat. Mach. Intell. *2*, 356–362.

Jang, H.-J., Lee, A., Kang, J., Song, I.H., and Lee, S.H. (2020). Prediction of clinically actionable genetic alterations from colorectal cancer histopathology images using deep learning. World J. Gastroenterol. *26*, 6207–6223.

Jing, L., and Tian, Y. (2019). Self-supervised visual feature learning with deep neural networks: a survey. Preprint at arXiv, 1902.06162.

Joo, S., Ko, E.S., Kwon, S., Jeon, E., Jung, H., Kim, J.-Y., Chung, M.J., and Im, Y.-H. (2021). Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Sci. Rep. *11*, 18800–18808.

Joze, H.R.V., Shaban, A., Iuzzolino, M.L., and Koishida, K. (2020). Mmtm: multimodal transfer module for cnn fusion. In In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13289–13299.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In

In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1725–1732.

Kather, J.N., Heij, L.R., Grabsch, H.I., Loeffler, C., Echle, A., Muti, H.S., Krause, J., Niehues, J.M., Sommer, K.A.J., Bankhead, P., et al. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. Nat. Cancer *1*, 789–799.

Kennedy, L.B., and Salama, A.K.S. (2020). A review of cancer immunotherapy toxicity. CA. A Cancer J. Clin. *70*, 86–104.

Khosravi, P., Lysandrou, M., Eljalby, M., Li, Q., Kazemi, E., Zisimopoulos, P., Sigaras, A., Brendel, M., Barnes, J., Ricketts, C., et al. (2021). A deep learning approach to diagnostic classification of prostate cancer using pathology– radiology fusion. J. Magn. Reson. Imaging *54*, 462–471.

Kirkpatrick, P. (2005). New clues in the acetaminophen mystery. Nat. Rev. Drug Discov. *4*, 883.

Kumar, A., Fulham, M., Feng, D., and Kim, J. (2020). Co-learning feature fusion maps from pet-ct images of lung cancer. IEEE Trans. Med. Imaging *39*, 204–217.

Lai, Y.-H., Chen, W.-N., Hsu, T.-C., Lin, C., Tsao, Y., and Wu, S. (2020). Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. Sci. Rep. *10*, 4679–4711.

Lasocki, A., Tsui, A., Tacey, M.A., Drummond, K.J., Field, K.M., and Gaillard, F. (2015). Mri grading versus histology: pre- dicting survival of world health organization grade ii–iv astrocytomas. AJNR. Am. J. Neuroradiol. *36*, 77–83.

Le, M.H., Chen, J., Wang, L., Wang, Z., Liu, W., Cheng, K.-T.T., and Yang, X. (2017). Automated diagnosis of prostate cancer in multi-parametric mri based on multimodal convolutional neural networks. Phys. Med. Biol. *62*, 6497–6514.

Lei, B., Huang, S., Li, H., Li, R., Bian, C., Chou, Y.-H., Qin, J., Zhou, P., Gong, X., and Cheng, J.-Z. (2020). Self-co- attention neural network for anatomy segmentation in whole breast ultrasound. Med. Image Anal. *64*, 101753.

Li, H., Bera, K., Toro, P., Fu, P., Zhang, Z., Lu, C., Feldman, M., Ganesan, S., Goldstein, L.J., Davidson, N.E., et al. (2021). Collagen fiber orientation disorder from h&e images is prognostic for early stage breast cancer: clinical trial validation. NPJ Breast Cancer *7*, 104–110.

Li, J., Chen, J., Tang, Y., Landman, B.A., and Zhou, S.K. (2022). Transforming medical imaging with transform- ers? a comparative review of key properties, current progresses, and future perspectives. Preprint at arXiv, 2206.01136.

Liang, M., Li, Z., Chen, T., and Zeng, J. (2014). Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. IEEE/ACM Trans. Comput. Biol. Bioinform. *12*, 928–937.

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020). A review of machine learning inter- pretability methods. Entropy *23*, 18.

Lipková, J., Angelikopoulos, P., Wu, S., Alberts, E., Wiestler, B., Diehl, C., Preibisch, C., Pyka, T., Combs, S.E., Hadjidoukas, P., et al. (2019). Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and bayesian inference. IEEE Trans. Med. Imaging *38*, 1875–1884.

Loeffler, C.M.L., Bruechle, N.O., Jung, M., Seillier, L., Rose, M., Laleh, N.G., Knuechel, R., Brinker, T.J., Trautwein, C., Gaisa, N.T., et al. (2021). Artificial Intelligence–Based Detection of Fgfr3 Mutational Status Directly from Routine Histology in Bladder Cancer: A Possible Preselection for Molecular Testing? (European Urology Focus).

Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol. *131*, 803–820.

Low, C.A. (2020). Harnessing consumer smartphone and wearable sensors for clinical cancer research. NPJ Digit. Med. *3*, 140–147.

Lu, M.Y., Chen, T.Y., Williamson, D.F.K., Zhao, M., Shady, M., Lipkova, J., and Mahmood, F. (2021). Ai-based pathology predicts origins for cancers of unknown primary. Nature *594*, 106–110.

Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., and Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng *5*, 555–570.

Marcus, L., Lemery, S.J., Keegan, P., and Pazdur, R. (2019). Fda approval summary: pembrolizumab for the treatment of microsatellite instability-high solid tumors. Clin. Cancer Res. *25*, 3753–3758.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Comput. Surv. *54*, 1–35.

Men, K., Geng, H., Zhong, H., Fan, Y., Lin, A., and Xiao, Y. (2019). A deep learning model for predicting xerostomia due to radiation therapy for head and neck squamous cell carcinoma in the rtog 0522 clinical trial. Int. J. Radiat. Oncol. Biol. Phys. *105*, 440–447.

Miller, G. (2002). Breaking Down Barriers.

Mo, S., Cai, M., Lin, L., Tong, R., Chen, Q., Wang, F., Hu, H., Iwamoto, Y., Han, X.-H., and Chen, Y.-W. (2020). Multi- modal priors guided segmentation of liver lesions in mri using mutual information based graph co-attention networks. In In International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer), pp. 429–438.

Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., Brat, D.J., and Cooper, L.A.D. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. Proc. Natl. Acad. Sci. USA *115*, E2970–E2979.

Murc̦han, P., Ó'Brien, C., O'Connell, S., McNevin, C.S., Baird, A.-M., Sheils, O., Ó Broin, P., and Finn, S.P. (2021). Deep learning of histopathological features for the prediction of tumour molecular genetics. Diagnostics *11*, 1406.

Naik, N., Madani, A., Esteva, A., Keskar, N.S., Press, M.F., Ruderman, D., Agus, D.B., and Socher, R. (2020). Deep learning-enabled breast cancer hormonal receptor status determination from base-level h&e stains. Nat. Commun. *11*, 5727–5728.

Nie, D., Zhang, H., Adeli, E., Liu, L., and Shen, D. (2016). 3d deep learning for multi-modal imaging-guided sur- vival time prediction of brain tumor patients. In In International conference on medical image computing and computer-assisted intervention (Springer), pp. 212–220.

Nie, D., Lu, J., Zhang, H., Adeli, E., Wang, J., Yu, Z., Liu, L., Wang, Q., Wu, J., and Shen, D. (2019). Multi-channel 3d deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. Sci. Rep. *9*, 1103–1114.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. *351*, 2817–2826.

Placido, D., Yuan, B., Hjaltelin, J.X., Haue, A.D., Yuan, C., Kim, J., Umeton, R., Antell, G., Chowdhury, A., Franz, A., et al. (2021). Pancreatic cancer risk predicted from disease trajectories using deep learning. Preprint at bioRxiv.

Qi, L., Ke, J., Yu, Z., Cao, Y., Lai, Y., Chen, Y., Gao, F., and Wang, X. (2021). Identification of prognostic spatial organization features in colorectal cancer microenvironment using deep learning on histopathology images. Med. Omics *2*, 100008.

Qian, X., Pei, J., Zheng, H., Xie, X., Yan, L., Zhang, H., Han, C., Gao, X., Zhang, H., Zheng, W., et al. (2021). Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning. Nat. Biomed. Eng. *5*, 522–532.

Rakha, E.A., El-Sayed, M.E., Lee, A.H., Elston, C.W., Grainge, M.J., Hodi, Z., Blamey, R.W., and Ellis, I.O. (2008). Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. J Clin Oncol *26*, 3153–3158.

Ramachandram, D., and Taylor, G.W. (2017). Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process. Mag. *34*, 96–108.

Ramanathan, T.T., Hossen, M., Sayeed, M., et al. (2022). Näıve bayes based multiple parallel fuzzy reasoning method for medical diagnosis. J. Eng. Sci. Technol. *17*, 0472–0490.

Reda, I., Khalil, A., Elmogy, M., Abou El-Fetouh, A., Shalaby, A., Abou El-Ghar, M., Elmaghraby, A., Ghazal, M., and El-Baz, A. (2018). Deep learning role in early diagnosis of prostate cancer. Technol. Cancer Res. Treat. *17* 15330346 18775530.

Reyes, M., Meier, R., Pereira, S., Silva, C.A., Dahlweid, F.-M., von Tengg-Kobligk, H., Summers, R.M., and Wiest, R. (2020). On the interpretability of artificial intelligence in radiology: challenges and opportunities. Radiol. Artif. Intell. *2*, e190043.

Rokach, L., and Maimon, O. (2005). Clustering methods. In In Data mining and knowledge discovery handbook (Springer), pp. 321–352.

Roy, S., Lahiri, D., Maji, T., and Biswas, J. (2015). Recurrent glioblastoma: where we stand. South Asian J. Cancer *4*, 163–173.

Sammut, S.-J., Crispin-Ortuzar, M., Chin, S.-F., Provenzano, E., Bardwell, H.A., Ma, W., Cope, W., Dariush, A., Dawson, S.-J., Abraham, J.E., et al. (2022). Multi-omic machine learning predictor of breast cancer therapy response. Nature *601*, 623–629.

Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maillé, P., Calderaro, J., Kamoun, A., Sefta, M., Toldo, S., Zaslavskiy, M., et al. (2020). A deep learning model to predict rna-seq expression of tumours from whole slide images. Nat. Commun. *11*, 3877–3915.

Sedghi, A., Mehrtash, A., Jamzad, A., Amalou, A., Wells, W.M., Kapur, T., Kwak, J.T., Turkbey, B., Choyke, P., Pinto, P., et al. (2020). Improving detection of prostate cancer foci via information fusion of mri and temporal enhanced ultrasound. Int. J. Comput. Assist. Radiol. Surg. *15*, 1215–1223.

Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Gradcam, D.B. (2016). Why did you say that?. Preprint at arXiv, 1611.07450.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: visual explanations from deep networks via gradient-based localization. In In Proceedings of the IEEE international conference on computer vision, pp. 618–626.

Sha, X., Gong, G., Qiu, Q., Duan, J., Li, D., and Yin, Y. (2019). Identifying pathological subtypes of non-small-cell lung cancer by using the radiomic features of 18f-fluorodeoxyglucose positron emission computed tomography. Transl. Cancer Res. *8*, 1741–1749.

Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., and Fu, H. (2022). Transformers in medical imaging: a survey. Preprint at arXiv, 2201.09873.

Shao, W., Han, Z., Cheng, J., Cheng, L., Wang, T., Sun, L., Lu, Z., Zhang, J., Zhang, D., and Huang, K. (2019). Integrative analysis of pathological images and multi-dimensional genomic data for early-stage cancer prognosis. IEEE Trans. Med. Imaging *39*, 99–110.

Shergalis, A., Bankhead, A., Luesakul, U., Muangsin, N., and Neamati, N. (2018). Current challenges and opportunities in treating glioblastoma. Pharmacol. Rev. *70*, 412–445.

Sidhom, J.W., Larman, H.B., Pardoll, D.M., and Baras, A.S. (2021). DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. Nat Commun *12*, 1–12.

Sundararajan, M., Taly, A., and Yan, Q. (2017). Axiomatic attribution for deep networks. In In International conference on machine learning (PMLR), pp. 3319–3328.

Taqi, S.A., Sami, S.A., Sami, L.B., and Zaki, S.A. (2018). A review of artifacts in histopathology. J. Oral Maxillofac. Pathol. *22*, 279.

Topol, E.J. (2020). Welcoming new guidelines for ai clinical research. Nat. Med. *26*, 1318–1320.

Tsou, P., and Wu, C.-J. (2019). Mapping driver mutations to histopathological subtypes in papillary thyroid carcinoma: applying a deep convolutional neural network. J. Clin. Med. *8*, 1675.

Vale-Silva, L.A., and Rohr, K. (2021). Long-term cancer survival prediction using multimodal deep learning. Sci. Rep. *11*, 13505–13512.

Van Cutsem, E., Köhne, C.H., Hitre, E., Zaluski, J., Chang Chien, C.-R., Makhson, A., D'Haens, G., Pintér, T., Lim, R., Bodoky, G., et al. (2009). Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. N. Engl. J. Med. Overseas. Ed. *360*, 1408–1417.

Van der Laak, J., Litjens, G., and Ciompi, F. (2021). Deep learning in histopathology: the path to the clinic. Nature medicine *27*, 775–784.

Vasileiou, G., Costa, M.J., Long, C., Wetzler, I.R., Hoyer, J., Kraus, C., Popp, B., Emons, J., Wunderle, M., Wenkel, E., et al. (2020). Breast mri texture analysis for prediction of brca-associated genetic risk. BMC Med. Imaging *20*, 86.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Atten- tion is all you need. Adv. Neural Inf. Process. Syst. *30*.

Vo, H.Q., Yuan, P., He, T., Wong, S.T., and Nguyen, H.V. (2021). Multimodal Breast Lesion Classification Using Cross-Attention Deep Networks. In In 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) (IEEE), pp. 1–4.

Wang, S., Shi, J., Ye, Z., Dong, D., Yu, D., Zhou, M., Liu, Y., Gevaert, O., Wang, K., Zhu, Y., et al. (2019). Predicting egfr mutation status in lung adenocarcinoma on computed tomography image using deep learning. Eur. Respir. J. *53*, 1800986.

Wang, X., Chen, Y., Gao, Y., Zhang, H., Guan, Z., Dong, Z., Zheng, Y., Jiang, J., Yang, H., Wang, L., et al. (2021). Predict- ing gastric cancer outcome from resected lymph node histopathology images using deep learning. Nat. Commun. *12*, 1637–1713.

Weeks, W.A., Dua, A., Hutchison, J., Joshi, R., Li, R., Szejer, J., and Azevedo, R.G. (2018). A low-power, low-cost in- gestible and wearable sensing platform to measure medication adherence and physiological signals. In In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), pp. 5549–5553.

Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D.E., and Zou, J. (2021). How medical ai devices are evaluated: limitations and recommendations from an analysis of fda approvals. Nat. Med. *27*, 582–584.

Xu, S., Jayaraman, A., and Rogers, J.A. (2019). Skin Sensors Are the Future of Health Care.

Xu, T., Zhang, H., Huang, X., Zhang, S., and Metaxas, D.N. (2016). Multimodal Deep Learning for Cervical Dysplasia Diagnosis. In In International conference on medical image computing and computer-assisted intervention (Springer), pp. 115–123.

Yala, A., Lehman, C., Schuster, T., Portnoi, T., and Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. Radiology *292*, 60–66.

Yamamoto, Y., Tsuzuki, T., Akatsuka, J., Ueki, M., Morikawa, H., Numata, Y., Takahara, T., Tsuyuki, T., Tsut- sumi, K., Nakazawa, R., et al. (2019). Automated acquisition of explainable knowledge from unannotated histopathology images. Nat. Commun. *10*, 5642–5649.

Yan, J., Zhang, B., Zhang, S., Cheng, J., Liu, X., Wang, W., Dong, Y., Zhang, L., Mo, X., Chen, Q., et al. (2021). Quantitative mri-based radiomics for noninvasively predicting molecular subtypes and survival in glioma patients. NPJ Precis. Oncol. *5*, 72–79.

Yap, J., Yolland, W., and Tschandl, P. (2018). Multimodal skin lesion classification using deep learning. Exp. Dermatol. *27*, 1261–1267.

Yogananda, C.G.B., Shah, B.R., Yu, F.F., Pinho, M.C., Nalawade, S.S., Murugesan, G.K., Wagner, B.C., Mickey, B., Patel, T.R., Fei, B., et al. (2020). A novel fully automated mri-based deep-learning method for classification of 1p/19q co-deletion status in brain gliomas. Neurooncol. Adv. *2* (*Supplement 4*), iv42–iv48.

Zhang, B.H., Lemoine, B., and Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. In In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 335–340.

Zhang, S., Tong, H., Xu, J., and Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. Comput. Soc. Netw. *6*, 11.

Zhou, L., and Luo, Y. (2021). Deep Features Fusion with Mutual Attention Transformer for Skin Lesion Diagnosis. In In 2021 IEEE International Conference on Image Processing (ICIP) (IEEE), pp. 3797–3801.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In In Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.

Zhuang, L., Lipkova, J., Chen, R., and Mahmood, F. (2022). Deep learning-based integration of histology, radiology, and genomics for improved survival prediction in glioma patients. In In Medical Imaging 2022: Digital and Computational Pathology, *12039* (SPIE), p. 120390Z.

Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., and Hoffman, M.M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. Inf Fusion *50*, 71–91.